

# Prevalence of Diseases and Statistical Power of the Japan Nurses' Health Study

Toshiharu FUJITA<sup>1\*</sup>, Kunihiro HAYASHI<sup>2</sup>, Kota KATANODA<sup>3</sup>,  
Yasuhiro MATSUMURA<sup>4</sup>, Jung Su LEE<sup>5</sup>, Hirofumi TAKAGI<sup>6</sup>,  
Shosuke SUZUKI<sup>7</sup>, Hideki MIZUNUMA<sup>8</sup> and Takeshi ASO<sup>9</sup>

<sup>1</sup> Department of Data Science, The Institute of Statistical Mathematics, 4–6–7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan

<sup>2</sup> Department of Basic Allied Medicine, School of Health Sciences, Gunma University, 3–39–22 Showa-Machi, Maebashi, Gunma 371-8511, Japan

<sup>3</sup> Center for Cancer Control & Information Services, National Cancer Center, 5–1–1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

<sup>4</sup> Information Center, National Institute of Health and Nutrition, 1–23–1 Toyama, Shinjuku-ku, 162-8636 Tokyo, Japan

<sup>5</sup> Department of Health Promotion Science, Graduate School of Medicine, The University of Tokyo, 7–3–1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

<sup>6</sup> Department of International Health and Nursing, School of Nursing, Faculty of Medicine, Toho University, 4–16–20 Omori-nishi, Ota-ku, Tokyo 143-0015, Japan

<sup>7</sup> Gunma Occupational Health Promotion Center, Gunma Medical Center 2F, 1–7–4 Chiyoda-Machi, Maebashi, Gunma 371-0022, Japan

<sup>8</sup> Department of Obstetrics and Gynecology, Hirosaki University School of Medicine, 5 Zaifucho, Hirosaki, Aomori 036-8562, Japan

<sup>9</sup> Department of Comprehensive Reproductive Medicine, Tokyo Medical and Dental University, 1–5–45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan

*Received June 2, 2007 and accepted September 11, 2007*

**Abstract:** The Japan Nurses' Health Study (JNHS) is a long-term, large-scale cohort study investigating the effects of various lifestyle factors and healthcare habits on the health of Japanese women. Based on currently limited statistical data regarding the incidence of disease among Japanese women, our initial sample size was tentatively set at 50,000 during the design phase. The actual number of women who agreed to participate in follow-up surveys was approximately 18,000. Taking into account the actual sample size and new information on disease frequency obtained during the baseline component, we established the prevalence of past diagnoses of target diseases, predicted their incidence, and calculated the statistical power for JNHS follow-up surveys. For all diseases except ovarian cancer, the prevalence of a past diagnosis increased markedly with age, and incidence rates could be predicted based on the degree of increase in prevalence between two adjacent 5-yr age groups. The predicted incidence rate for uterine myoma, hypercholesterolemia, and hypertension was  $\geq 3.0$  (per 1,000 women, per year), while the rate of thyroid disease, hepatitis, gallstone disease, and benign breast tumor was predicted to be  $\geq 1.0$ . For these diseases, the statistical power to detect risk factors with a relative risk of 1.5 or more within ten years, was 70% or higher.

**Key words:** Cohort Study, Women's Health, Nurses, Incidence rate, Statistical power

\*To whom correspondence should be addressed.

## Introduction

Many large-scale epidemiological studies of women's health have been conducted in various parts of the world, such as the Nurses' Health Study (NHS)<sup>1, 2)</sup>, the Iowa Women's Health Study<sup>3)</sup>, the Women's Health Initiative Study (WHI)<sup>4)</sup>, the Women's Health Study<sup>5)</sup>, the Million Women Study (MWS)<sup>6)</sup>, the Australian Longitudinal Study on Women's Health<sup>7)</sup>, and the Shanghai Women's Health Study<sup>8)</sup>. Nevertheless, large-scale research of this nature has been comparatively rare in Japan. Among those investigations previously conducted, the NHS incorporates a unique occupational cohort and observes lifestyle and health events by biennial questionnaires. Although the NHS has greatly contributed to women's health worldwide by reporting the findings on oral contraceptives, postmenopausal use of hormones, smoking, food and shift work<sup>1)</sup>, the findings cannot necessarily be generalized for Japanese women. Japanese women may have a different lifestyle, working environment, and healthcare practice from women in western countries, and as such, it is important to consider cultural and biological interactions during women's health studies<sup>9)</sup>.

The Japan Nurses' Health Study (JNHS), was designed to be comparable with the NHS, and is a large-scale epidemiological cohort study investigating women's lifestyle habits and health, that is planned to follow female nurses aged 25 yr and over for at least ten years<sup>10)</sup>. The objectives of the JNHS are to monitor the occurrence of various diseases prospectively, including female-specific diseases, and to assess the effects of various lifestyle factors and healthcare habits on the health of Japanese women. The JNHS consists mainly of a cross-sectional baseline survey and longitudinal follow-up surveys. The baseline survey was completed in 2007, during which responses were received from 49,914 women from 47 prefectures all over Japan. Among the respondents, approximately 18,000 women agreed to participate in the follow-up surveys.

Calculating the sample size of a study population is indispensable for clinical trials. Even in an observational epidemiological study, it is essential that the study protocol describes sample size and adequate statistical power. For example, the Women's Health Initiative (WHI) study described its statistical power calculations for an observational study as well as for clinical trials conducted by the WHI<sup>4)</sup>. If an association between a lifestyle factor and a disease truly exists, the possibility to detect statistical significance of the association greatly depends on the strength of the association and incidence rates of this disease in the

population. Nevertheless, we were unable to find any reliable statistical data on the incidence rate of diseases among Japanese women, except for population-based cancer registries<sup>12)</sup>. The initial sample size of the JNHS baseline survey was tentatively set at 50,000 based on the population-based cancer registries, i.e. the incidence rates of breast cancer and endometrial cancer in the Japanese general female population at the study design phase. Therefore, it is important to examine the statistical power for long-term studies, based on more reliable information of the JNHS population which may be obtained after initiation of the study. For example, in a paper describing the design of the Million Women's Study<sup>6)</sup>, it was reported that the statistical power was updated during the study entry period. These revisions were based on information obtained from new studies indicating that mortality in women taking hormonal agents is likely to be 20% lower than that of the general female population.

In the JNHS, it was considered essential to examine the statistical power of follow-up surveys, as the actual sample size of participants, who consented to the follow-up, was approximately 18,000. The present study investigated the prevalence of a past diagnosis of various diseases among 5-yr age groups based on the JNHS cross-sectional baseline survey. We also predicted the incidence rates of these diseases, and examined the statistical power to detect risk factors for diseases in the JNHS.

## Materials and Methods

The present study analyzed data obtained from a total of 44,139 women who participated in the JNHS cross-sectional baseline survey in 2001–2004. These women were divided into 5-yr age groups as follows: 30–34 yr (*n*=10,471); 35–39 yr (*n*=9,466); 40–44 yr (*n*=9,060); 45–49 yr (*n*=7,572); 50–54 yr (*n*=5,542); and 55–59 yr (*n*=2,018). The study design of the JNHS and characteristics of the respondents at the baseline survey have been described in detail elsewhere<sup>10)</sup>. The following diseases were studied: hypertension, myocardial infarction, angina pectoris, subarachnoid hemorrhage, cerebral hemorrhage, cerebral infarction, transient ischemic attack, vein thrombosis/pulmonary embolism, artery thrombosis of lower limbs, diabetes mellitus, thyroid disease, hypercholesterolemia, gallstone disease, hepatitis, endometriosis, uterine myoma, uterine cervical cancer, endometrial cancer, ovarian cancer, benign breast tumor, breast cancer, stomach cancer, colorectal cancer, and osteoporosis.

The prevalence of a past diagnosis, which refers to the

proportion of women who had been diagnosed at the time of the baseline survey, was calculated for each of these diseases. The Cochran-Armitage test for trend was used to examine whether the prevalence of past diagnosis increased monotonously with age among the six 5-yr age groups. Analyses were performed on Statistical Analysis System Software (Ver 8, SAS Institute, Cary, NC), and the level of significance was set at a two-sided 5%. When the prevalence of past diagnosis of a disease increased monotonously with age, it was assumed that women developed the disease between adjacent age groups, and the incidence rate of the disease was predicted based on the degree of increase between the two adjacent age groups. For example, for one 5-yr age group ( $i$ ) and the next 5-yr age group ( $i+1$ ), the incidence rate of a rare disease between the two groups was calculated using the following formula:

$$I_{i \rightarrow i+1} = \frac{(P_{i+1} - P_i)}{(AGE_{i+1} - AGE_i)}$$

where,  $N_i$  and  $N_{i+1}$  represent the numbers of subjects,  $D_i$  and  $D_{i+1}$  are the numbers of individuals diagnosed with the disease,  $P_i$  and  $P_{i+1}$  are the prevalence of past diagnosis ( $P_i = D_i/N_i$ ), and  $AGE_i$  and  $AGE_{i+1}$  are the average ages.

- $i =$
- 1: 30–34 yr
  - 2: 35–39 yr
  - 3: 40–44 yr
  - 4: 45–49 yr
  - 5: 50–54 yr
  - 6: 55–59 yr

As shown by the following formula, the inverse of the variance in each incidence rate between two adjacent age groups (e.g., between 30–34-yr-olds and 35–39-yr-olds) was weighed in order to calculate the overall incidence rate from 30–59 yr of age,  $I$ , for each disease.

$$I = \frac{\sum_{i=1}^3 I_{i \rightarrow i+1} \times w_i}{\sum_{i=1}^3 w_i}, \quad w_i = \frac{(AGE_{i+1} - AGE_i)^2}{\left\{ \frac{P_i(1-P_i)}{n_i} + \frac{P_{i+1}(1-P_{i+1})}{n_{i+1}} \right\}}$$

We also examined how many years of follow-up would be needed in the JNHS in order to detect the statistical significance of an association. When observing the occurrence of a disease by following 18,000 participants while setting the level of significance (two-sided  $\alpha$ ) at 5% and the proportion of exposure to a specific factor among the survey participants (exposed group) at 20%, statistical power ( $1-\beta$ ) was calculated with the minimum relative risks of 1.5 and 2.5 to be detected at 2, 4, 6 and 10 yr after the

baseline survey, according to the following formula<sup>11)</sup>:

$$z_{1-\beta} = \frac{\sqrt{\frac{n}{1+K}} \times |p_0(1-R)| - z_{1-\alpha/2} \times \sqrt{(1+1/K)U(1-U)}}{\sqrt{p_0R(1-Rp_0) + p_0(1-p_0)/K}}$$

where,  $Z_{1-\alpha/2}$  and  $Z_{1-\beta}$  refer to the unit normal deviates corresponding to level of significance ( $\alpha$ ), and statistical power ( $1-\beta$ );  $n$  is the sample size (18,000);  $R$  is the minimum relative risk to be detected;  $K$  is the ratio of the sample size of the unexposed to that of exposed groups (4, because the proportion of participants who had ever used female hormone agents (oral contraceptives and/or postmenopausal use of hormones) was about 20% at the baseline survey);  $U$  is the predicted overall incidence rate of each disease among the surveyed population; and  $p_0$  is the cumulative incidence rate of each disease up to “ $y$ ” years among the unexposed group.

$$p_0 = 1 - \exp \left\{ -I \times \frac{(K+I)}{(K+R)} \times y \right\}$$

$$U = \frac{Kp_0 + p_0R}{K+1}$$

## Results

### *Prevalence of a past diagnosis and the predicted incidence rate of each disease*

The prevalence of a past diagnosis of uterine myoma, hypercholesterolemia, endometriosis and hypertension exceeded 5% (Table 1). Because the weight of young participants was relatively high in the baseline survey, the prevalence of past diagnosis of gynecological diseases that are often seen in women aged 30–34 yr was high. As the results of the Cochran-Armitage test for trend, the prevalence of past diagnosis of all diseases increased monotonously according to age, excluding ovarian cancer. This tendency was particularly marked for hypercholesterolemia, uterine myoma, diabetes, gallstone disease, hepatitis, osteoporosis, angina pectoris, and hypertension. Figure 1 shows the prevalence of past diagnosis for the three most frequent diseases among 55–59-yr-old women (i.e., hypercholesterolemia, hypertension, and uterine myoma) for each 5-yr age group. The prevalence of past diagnosis increased markedly with age, thus suggesting disease occurrence. Based on the degree of increase in the prevalence of past diagnosis between two adjacent 5-yr age groups, overall incidence rates were predicted from 30 to 59 yr of age. The predicted overall incidence rate was  $\geq 3.0$  (per 1,000 women per year) for uterine myoma, hypercholesterolemia, and hypertension, and  $\geq 1.0$  for thyroid disease, hepatitis, gallstone disease, and benign breast tumor (Table 2). Among all age groups, the

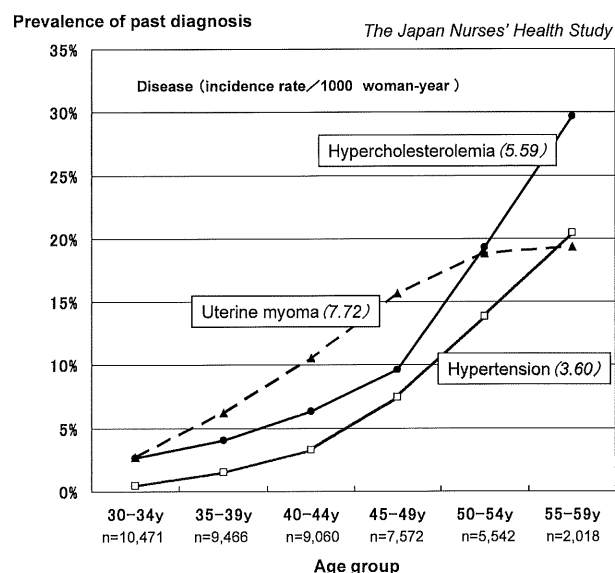
**Table 1.** Prevalence of a past diagnosis of each disease by age group during the baseline survey

	Prevalence of past diagnosis							Cochran-Armitage test	
	Total n=44,129	30–34 yr n=10,471	35–39 yr n=9,466	40–44 yr n=9,060	45–49 yr n=7,572	50–54 yr n=5,542	55–59 yr n=2,018	$\chi^2$ value	p value
Hypertension	5.06%	0.48%	1.50%	3.26%	7.45%	13.86%	20.47%	222.27	<0.001
Myocardial infarction	0.06%	0.00%	0.02%	0.03%	0.09%	0.20%	0.25%	33.31	<0.001
Angina pectoris	0.52%	0.07%	0.18%	0.26%	0.66%	1.44%	2.58%	260.38	<0.001
Subarachnoid hemorrhage	0.10%	0.03%	0.05%	0.09%	0.11%	0.20%	0.50%	29.93	<0.001
Cerebral hemorrhage	0.06%	0.03%	0.03%	0.07%	0.08%	0.13%	0.10%	6.97	0.008
Cerebral infarction	0.20%	0.07%	0.03%	0.14%	0.22%	0.54%	0.84%	77.08	<0.001
Transient ischemic attack	0.45%	0.18%	0.24%	0.39%	0.58%	0.70%	1.88%	91.41	<0.001
Vein thrombosis/pulmonary embolism	0.28%	0.13%	0.20%	0.36%	0.37%	0.40%	0.40%	15.31	<0.001
Artery thrombosis of lower limbs	0.17%	0.10%	0.10%	0.13%	0.28%	0.29%	0.30%	16.12	<0.001
Diabetes mellitus	1.17%	0.17%	0.38%	0.86%	1.68%	2.87%	4.91%	475.86	<0.001
Thyroid disease	4.22%	2.32%	3.32%	4.69%	5.53%	5.94%	6.54%	209.55	<0.001
Hypercholesterolemia	8.23%	2.66%	4.05%	6.36%	9.60%	19.29%	29.68%	2358.86	<0.001
Gallstone disease	2.08%	0.46%	1.24%	2.25%	3.08%	3.88%	5.00%	378.50	<0.001
Hepatitis	3.40%	1.62%	2.22%	3.29%	4.38%	6.17%	7.28%	374.66	<0.001
Endometriosis	5.29%	4.18%	4.85%	5.70%	6.50%	5.65%	5.70%	37.35	<0.001
Uterine myoma	10.08%	2.67%	6.24%	10.58%	15.65%	18.82%	19.33%	1672.37	<0.001
Uterine cervical cancer	0.91%	0.59%	0.92%	1.05%	1.11%	0.96%	0.94%	8.68	0.003
Endometrial cancer	0.16%	0.03%	0.12%	0.11%	0.25%	0.16%	0.84%	41.52	<0.001
Ovarian cancer	0.20%	0.11%	0.22%	0.20%	0.34%	0.16%	0.20%	3.03	0.082
Benign breast tumor	3.48%	2.23%	3.12%	3.63%	4.70%	4.19%	4.36%	81.51	<0.001
Breast cancer	0.77%	0.16%	0.32%	0.68%	1.23%	1.68%	2.23%	206.00	<0.001
Stomach cancer	0.39%	0.06%	0.17%	0.43%	0.54%	0.96%	0.89%	99.18	<0.001
Colorectal cancer	0.22%	0.00%	0.05%	0.12%	0.36%	0.69%	0.79%	116.30	<0.001
Osteoporosis	0.93%	0.32%	0.36%	0.65%	0.82%	2.24%	4.81%	335.94	<0.001

predicted incidence rate of gynecological diseases such as uterine myoma, endometriosis, uterine cervical cancer, and benign breast tumor increased from the 30's to the early 40's, while the incidence of so-called 'lifestyle-related' diseases such as hypertension, angina pectoris, cerebral infarction, diabetes, hypercholesterolemia, and osteoporosis increased from the late 40's to the 50's.

#### *Calculation of the statistical power to detect a risk factor for a particular disease*

Table 3 shows the statistical power for diseases listed in decreasing order of the predicted incidence rate in relation to the follow-up period. When following 18,000 women aged 30–59 yr, the statistical power for uterine myoma (7.72/1,000 women per year) after 2 yr of follow-up was 81% with a relative risk of 1.5 or more. With regard to other diseases, the statistical power exceeded 70% after 4 yr for hypercholesterolemia and hypertension, after 6 yr for thyroid disease, after 8 yr for hepatitis and gallstone disease, and after 10 yr for benign breast tumor. For diseases where the statistical power did not reach 70% after 10 yr of follow-up

**Fig. 1.** Prevalence of a past-diagnosis in the female nurse population.

**Table 2. Predicted incidence rate of each disease by age group**

	Overall incidence rate (95% confidence interval)	Predicted incidence rate <sup>a</sup>				
		30–34 yr	35–39 yr	40–44 yr	45–49 yr	50–54 yr
		35–39 yr	40–44 yr	45–49 yr	50–54 yr	55–59 yr
Hypertension	3.60 (3.17, 4.04)	2.06	3.54	8.39	13.31	13.33
Myocardial infarction	0.05 (0.00, 0.10)	0.04	0.02	0.12	0.22	0.10
Angina pectoris	0.35 (0.20, 0.50)	0.23	0.17	0.79	1.63	2.29
Subarachnoid hemorrhage	0.07 (0.00, 0.15)	0.05	0.07	0.03	0.19	0.60
Cerebral hemorrhage	0.03 (0.00, 0.10)	0.01	0.07	0.03	0.10	–0.05
Cerebral infarction	0.08 (0.00, 0.17)	–0.07	0.23	0.16	0.66	0.61
Transient ischemic attack	0.27 (0.09, 0.44)	0.12	0.29	0.39	0.25	2.38
Vein thrombosis/pulmonary embolism	0.14 (0.00, 0.29)	0.14	0.33	0.01	0.06	0.00
Artery thrombosis of lower limbs	0.07 (0.00, 0.18)	0.00	0.08	0.29	0.02	0.02
Diabetes mellitus	0.82 (0.59, 1.05)	0.42	0.97	1.63	2.48	4.11
Thyroid disease	1.96 (1.38, 2.54)	2.01	2.77	1.69	0.84	1.22
Hypercholesterolemia	5.59 (4.90, 6.27)	2.78	4.65	6.49	20.12	20.97
Gallstone disease	1.72 (1.35, 2.09)	1.57	2.05	1.65	1.67	2.27
Hepatitis	1.91 (1.41, 2.41)	1.20	2.16	2.19	3.71	2.25
Endometriosis	0.93 (0.28, 1.59)	1.34	1.70	1.60	–1.77	0.10
Uterine myoma	7.72 (6.92, 8.53)	7.19	8.74	10.13	6.58	1.02
Uterine cervical cancer	0.25 (0.00, 0.54)	0.66	0.26	0.12	–0.32	–0.03
Endometrial cancer	0.12 (0.02, 0.22)	0.18	–0.01	0.28	–0.18	1.37
Ovarian cancer	0.07 (0.00, 0.20)	0.24	–0.05	0.29	–0.38	0.07
Benign breast tumor	1.18 (0.65, 1.72)	1.78	1.04	2.14	–1.07	0.35
Breast cancer	0.56 (0.35, 0.77)	0.31	0.74	1.09	0.93	1.11
Stomach cancer	0.31 (0.17, 0.46)	0.23	0.53	0.22	0.86	–0.13
Colorectal cancer	0.15 (0.07, 0.23)	0.11	0.14	0.47	0.68	0.22
Osteoporosis	0.50 (0.28, 0.72)	0.07	0.59	0.34	2.95	5.18

<sup>a</sup>: per 1,000 women per year.

A negative value for the incidence rate means that the prevalence of a past diagnosis of one 5-yr age group was higher than that of the next 5-yr age group.

with a relative risk of 1.5 or more, the statistical power was recalculated with a relative risk of 2.5 or more. The statistical power exceeded 70% after 4 yr for endometriosis, diabetes, and breast cancer, after 6 yr for osteoporosis; after 8 yr for angina pectoris and stomach cancer, and after 10 yr for transient ischemic attack and uterine cervical cancer.

## Discussion

The present study identified marked increases with age in the prevalence of past diagnosis for all diseases, excluding ovarian cancer. We proposed a simple method for predicting an incidence rate based on the degree of age-related increase in the prevalence of past diagnosis. The incidence rate of each disease was also predicted. Nevertheless, there are several assumptions and limitations associated with the prediction of incidence rates. Firstly, we assumed the lack of a cohort effect on disease occurrence. In other words, we assumed that the disease incidence rate among 5-yr age

groups was the same for all cohorts. While it is possible that changes in lifestyle habits bring about systemic changes in disease occurrence among cohorts, this point was not taken into account. For example, it is uncertain whether an observed increase in the prevalence of hepatitis is affected by aging, such as the duration of occupational hazard exposure, or by a cohort effect. The incidence rates of gynecological diseases such as endometriosis, benign breast tumor, ovarian cancer and endometrial cancer between 45–49-yr-old and 50–54 yr-old were negative in value (Table 2). This might be influenced by a cohort effect related to reproductive health.

Secondly, data from women who died due to diseases were not included in the JNHS. The prevalence of past diagnosis of ovarian cancer was highest among women aged 45–49 yr, and did not indicate a monotonous statistical increase (Table 1). This might be influenced by a higher fatality rate of ovarian cancer compared with that of uterine cervical cancer or endometrial cancer. Thirdly, most participants were working. They had past diagnoses of diseases but were

**Table 3. Calculated statistical power for each disease in the Japan Nurses' Health Study**

Minimum relative risk to be detected = 1.5

	Overall incidence rate (per 1,000 women, per year)	Statistical power (%)				
		Years after the baseline survey				
		2 yr	4 yr	6 yr	8 yr	10 yr
Uterine myoma	7.72	81	98	100	100	100
Hypercholesterolemia	5.59	68	92	99	100	100
Hypertension	3.60	51	78	92	97	99
Thyroid disease	1.96	32	54	70	82	89
Hepatitis	1.91	31	53	69	81	88
Gallstone disease	1.72	29	49	65	77	85
Benign breast tumor	1.18	22	37	50	61	71
Endometriosis	0.93	19	31	42	52	61
Diabetes mellitus	0.82	17	28	38	47	56
Breast cancer	0.56	14	21	29	35	42
Osteoporosis	0.50	13	20	26	32	38
Angina pectoris	0.35	11	16	20	25	29
Stomach cancer	0.31	10	15	19	23	27
Transient ischemic attack	0.27	9	13	17	21	24
Uterine cervical cancer	0.25	9	13	16	20	23
Colorectal cancer	0.15	7	10	12	14	16
Vein thrombosis/pulmonary embolism	0.14	7	10	12	14	16
Endometrial cancer	0.12	7	9	11	13	14
Cerebral infarction	0.08	6	7	9	10	11
Subarachnoid hemorrhage	0.07	6	7	8	10	11
Artery thrombosis of lower limbs	0.07	6	7	8	10	11
Ovarian cancer	0.07	6	7	8	10	11
Myocardial infarction	0.05	6	6	7	8	9
Cerebral hemorrhage	0.03	5	6	7	7	7

Alpha = 0.05 (two-tailed), total sample size = 18,000, unexposed-exposed ratio = 4:1.

Minimum relative risk to be detected = 2.5

	Overall incidence rate (per 1,000 women, per year)	Statistical power (%)				
		Years after the baseline survey				
		2 yr	4 yr	6 yr	8 yr	10 yr
Endometriosis	0.93	63	88	97	99	100
Diabetes mellitus	0.82	58	84	95	98	100
Breast cancer	0.56	45	71	85	93	97
Osteoporosis	0.50	42	66	81	90	95
Angina pectoris	0.35	32	53	68	79	87
Stomach cancer	0.31	30	49	64	75	83
Transient ischemic attack	0.27	27	44	57	69	77
Uterine cervical cancer	0.25	26	42	56	67	75
Colorectal cancer	0.15	19	29	39	47	55
Vein thrombosis/pulmonary embolism	0.14	18	28	38	46	54
Endometrial cancer	0.12	16	25	33	41	48
Cerebral infarction	0.08	13	19	24	29	34
Subarachnoid hemorrhage	0.07	12	18	23	28	33
Artery thrombosis of lower limbs	0.07	12	18	23	28	33
Ovarian cancer	0.07	12	18	23	28	33
Myocardial infarction	0.05	11	15	19	22	26
Cerebral hemorrhage	0.03	9	12	15	17	19

Alpha = 0.05 (two-tailed), total sample size = 18,000, unexposed-exposed ratio = 4:1.

able to work, thus they had diseases with relatively favorable prognoses. It is possible that severe outcomes were excluded, which could have resulted in an underestimate of the prevalence of past diagnoses. Nevertheless, the prediction of an incidence rate might not be greatly influenced, because this underestimate may share similar tendencies between 5-yr age groups, and that non-severe outcomes occur more frequently than severe ones.

Few reliable statistical data are available regarding the incidence rate of diseases among Japanese women, except for an estimated incidence rate of cancer based on population-based cancer registries<sup>12)</sup>. According to these cancer registries, the incidence rate of breast cancer is 0.170 (per 1,000 women per year) among women aged 30–34 yr, 0.419 among women aged 35–39 yr, 0.810 among women aged 40–44 yr, 1.264 among women aged 45–49 yr, 1.122 among women aged 50–54 yr, and 1.012 among women aged 55–59 yr. These incidence rates are consistent with the predicted incidence rates of breast cancer for corresponding age groups in Table 2. With regard to uterine cervical cancer, uterine cancer, ovarian cancer, stomach cancer, and colorectal cancer, there were marked variations in each incidence rate among the age groups due to sample size limitations. However, the predicted overall incidence rate from ages 30 to 59 was mostly comparable with that of the cancer registries. We believe that the predicted incidence rate can be a reference for the incidence rate of each disease investigated in our study population of the JNHS.

Based on the predicted overall incidence rate, the statistical power for each disease was examined. For diseases with high incidence rates, such as uterine myoma, hypercholesterolemia, and hypertension, the statistical power was able to assess an increase of 1.5 or more in relative risk for factors with an exposure prevalence of 20% within 4 yr. With regard to diseases with an overall incidence rate of  $\geq 1.2$  (per 1,000 women per year), such as thyroid disease, hepatitis, and gallstone disease, it was possible to investigate an increase of about 1.5 or more in relative risk within 10 yr. Furthermore, with an increase of about 2.5 or more in relative risk, it was possible to investigate diseases with an overall incidence rate of  $\geq 0.25$ , such as endometriosis, diabetes mellitus, osteoporosis, stomach cancer, and uterine cervical cancer, within 10 yr.

The incidence rate of diseases was predicted in the present study when the age composition of the survey population was the same as that of the baseline survey. However, as the follow-up surveys continue, the age distribution of the survey population will gradually shift towards higher ages. Therefore, the overall incidence rate of the entire population

group will be higher, even if it is assumed that the incidence rate for each age group does not change. It is expected that the JNHS will have greater statistical power to detect a risk factor for each disease. On the other hand, a loss of follow-up reduces statistical power. However, the U.S. Nurses' Health Study had maintained a high follow-up rate of  $>90\%$  during more than 10 yr<sup>1)</sup>. Maintenance strategies for our JNHS cohort are described in detail elsewhere<sup>10)</sup>.

## Conclusion

Overall, we conclude that even with the current sample size, it is possible to achieve the stated goal of the JNHS; to clarify gynecological health problems in Japan and establish epidemiological evidence of health care at various life stages among Japanese women.

## Acknowledgements

We would like to express our sincere appreciation of the participating nurses, the Japanese Nursing Association, and the Japan Menopause Society for their cooperation. This research was supported in part by a Grant-in-Aid for Scientific Research (B: 19390180) from the Japan Society for the Promotion of Science, and by the grants from the Japan Menopause Society.

## References

- 1) Colditz GA, Manson JE, Hankinson SE (1997) The Nurses' Health Study—20-year contribution to the understanding of health among women. *J Women's Health* **6**, 49–62.
- 2) Solomon CG, Willett WC, Carey VJ, Rich-Edwards J, Hunter DJ, Colditz GA, Stampfer MJ, Speizer FE, Spiegelman D, Manson JE (1997) A prospective study of pregravid determinants of gestation diabetes mellitus. *JAMA* **278**, 1078–83.
- 3) Bostick RM, Potter JD, Sellers TA, McKenzie DR, Kushi LH, Folsom AR (1993) Relation of calcium, vitamin D, and dairy food intake to incidence of colon cancer among older women. The Iowa Women's Health Study. *Am J Epidemiol* **137**, 1302–17.
- 4) The Women's Health Initiative Study Group (1998) Design of the Women's Health Initiative Clinical Trial and Observational Study. *Controlled Clinical Trials* **19**, 61–109.
- 5) Rexrode KM, Lee IM, Cook NR, Hennekens CH, Buring JE (2000) Baseline characteristics of participants in the women's health study. *J Women's Health Gend Based Med* **9**, 19–27.
- 6) The Million Women Study Collaborative Group (1999) The Million Women Study: design and characteristics of the study population. *Breast Cancer Research* **1**, 73–80.
- 7) Brown WJ, Bryson L, Byles JE, Dobson AJ, Lee C, Mishra

- G, Schofield M (1998) Women's health Australia: recruitment for a national longitudinal cohort study. *Women Health* **28**, 23–40.
- 8) Villegas R, Shu XO, Li H, Yang G, Matthews CE, Leitzmann M, Li Q, Cai H, Gao YT, Zheng W (2006) Physical activity and the incidence of type 2 diabetes in the Shanghai women's health study. *Int J Epidemiol* **35**, 1553–62.
- 9) Crawford SL (2007) The roles of biologic and nonbiologic factors in cultural differences in vasomotor symptoms measured by surveys. *Menopause* **14**, 725–33.
- 10) Hayashi K, Mizunuma H, Fujita T, Suzuki S, Imazeki S, Katanoda K, Matsumura Y, Kubota T, Aso T (2007) Design of the Japan Nurses' Health Study: a prospective occupational cohort study of women's health in Japan. *Ind Health* **45**, 679–686.
- 11) Schlesselman JJ (1974) Sample size requirements in cohort and case-control studies of disease. *Am J Epidemiol* **99**, 381–4.
- 12) Marugame T, Kamo K, Katanoda K, Ajiki W, Sobue T (2006) Cancer incidence and incidence rates in Japan in 2000: Estimate based on data from 11 population-based cancer registries. *Jpn J Clin Oncol* **36**, 668–75.