

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## Validity of self-reported diagnoses of gynecological and breast cancers in the Japan Nurses' Health Study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-045491
Article Type:	Original research
Date Submitted by the Author:	04-Oct-2020
Complete List of Authors:	Takamatsu, Kiyoshi; Tokyo Dental College Ichikawa General Hospital, Obstetrics and Gynecology Ideno, Yuki; Gunma University, Center for Mathematics and Data Science Kikuchi, Mami; Gunma University Hospital, Center of Regional Medical Research and Education Yasui, Toshiyuki; University of Tokushima Graduate School of Biomedical Sciences, Department of Reproductive and Menopausal Medicine Maruoka, Naho; Gunma University Graduate School of Health Sciences, Department of International/Community Health Laboratory Sciences Nagai, Kazue; Gunma University Graduate School of Health Sciences, Gunma University, Department of International/Community Health Laboratory Sciences Hayashi, Kunihiko; Gunma University Graduate School of Health Sciences, Department of International/Community Health Laboratory Sciences
Keywords:	EPIDEMIOLOGY, Breast tumours < ONCOLOGY, Gynaecological oncology < ONCOLOGY

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Validity of self-reported diagnoses of gynecological and breast cancers in the Japan Nurses' Health Study

Kiyoshi Takamatsu, MD, PhD, <sup>1</sup> Yuki Ideno, PhD, <sup>2</sup> Mami Kikuchi, MD, <sup>3</sup> Toshiyuki Yasui, MD, PhD, <sup>4</sup> Naho Maruoka, BS, <sup>5</sup> Kazue Nagai, PhD, <sup>5</sup> Kunihiko Hayashi, PhD <sup>5</sup>

### Author affiliations

<sup>1</sup>Department of Obstetrics and Gynecology, Tokyo Dental College Ichikawa General Hospital, Chiba, Japan

<sup>2</sup>Gunma University Center for Mathematics and Data Science, Gunma, Japan

<sup>3</sup>Center of Regional Medical Research and Education, Gunma University Hospital, Maebashi, Gunma, Japan

<sup>4</sup>Department of Reproductive and Menopausal Medicine, Institute of Biomedical Science, Tokushima University Graduate School, Tokushima, Japan

<sup>5</sup>Department of International/Community Health Laboratory Sciences, Graduate School of Health Sciences, Gunma University, Gunma, 371-8514, Japan

Correspondence to Dr Kiyoshi Takamatsu, Department of Obstetrics and Gynecology, Tokyo Dental College Ichikawa General Hospital, Sugano 5-11-13, Ichikawa-city, Chiba 272-0151, JAPAN; ktakamatsu@tdc.ac.jp

## ABSTRACT

**Background** The validity of self-reported diagnoses of gynecological and breast cancers is controversial. We investigated it in a nationwide prospective cohort study of nursing professionals: the Japan Nurses' Health Study (JNHS)

**Methods** Data were reviewed for 15,717 subjects. The mean age at baseline was  $41.6 \pm 8.3$  years (median: 41), and the mean follow-up period was  $10.5 \pm 3.8$  years (median: 12). Participants are regularly mailed a follow-up questionnaire once every 2 years. Respondents who self-reported a positive cancer diagnosis were sent an additional confirmation questionnaire and contacted the diagnosing facility to confirm the diagnosis based on medical records. A review panel of experts verified the disease status. Regular follow-up, confirmation questionnaires, and expert review were validated for their positive predictive value (PPV) and negative predictive value (NPV).

**Results** New incidences were verified in 37, 47, 26, and 300 cervical, endometrial, ovarian, and breast cancer cases, respectively. The estimated incidence rates were 22.0, 25.4, 13.8 and 160.4 per 100,000 person-years. These were comparable to those of national data from regional cancer registries in Japan. For regular follow-up, the corresponding PPVs for cervical, endometrial, ovarian, and breast cancer were 16.9%, 54.2%, 45.1%, and 81.4%, and the NPVs were 100%, 99.9%, 99.9%, and 99.9%.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

respectively. Adding the confirmation questionnaire improved the PPVs to 31.5%, 88.9%, 76.7%, and 99.9%; the NPVs were uniformly 99.9%. Expert review yielded PPVs and NPVs that were all ~100%.

**Conclusion** Gynecological cancer cannot be accurately assessed by self-reporting alone.

Additionally, the external validity of cancer incidence in this cohort was confirmed.

### **Strengths and limitations of this study**

- ▶ This study investigated the validity of self-reporting of gynecologic and breast cancers in a large, nationwide prospective cohort study of nursing professionals, the Japan Nurses' Health Study (JNHS).
- ▶ Participants of JNHS cohort, which was composed entirely of female nursing professionals, are likely to answer the cancer history more accurately than general population.
- ▶ Periodic questionnaires, meticulous review of subjects' medical records and death-certificate surveys were employed to establish self-report validity, circumventing the limitations presented by Japan's lack of complete national cancer registries.
- ▶ Not all answer for confirmation questionnaire was sent back.
- ▶ There was relatively small number of young skew of the participants' ages in this cohort.

## INTRODUCTION

Self-reporting is frequently used to assess disease status in cohort research. The methodology's cost-effectiveness and feasibility make it an attractive approach in countries without comprehensive national disease registries such as Japan. However, the unreliability of self-reported information is problematic and can introduce errors into epidemiological investigations of risk factors, especially for new cancer incidences in a cohort. Self-reporting appears to accurately reflect diabetes status and surgical history of hysterectomies<sup>1,2</sup>; however, body weight is often under-reported<sup>3</sup>. Regarding patients' cancer history, healthcare providers must consider that an affirmative response on a questionnaire is not equivalent to a definitive medical diagnosis because patients may remember incorrectly. Ideally, their answers should be corroborated against their medical records, but these typically cannot be acquired for an entire cohort. Additionally, validity can depend on background factors, such as ethnicity and cohort-specific characteristics, which further complicates interpreting self-report data.

The Japan Nurses' Health Study (JNHS) is a nationwide prospective cohort study of over 15,000 female nurses, which began in 2001 to ascertain how women's health is affected by lifestyle factors, healthcare practices, and physical status over their lifetime<sup>4</sup>. Here, we investigated the validity of self-reported diagnoses of three gynecological

1  
2  
3  
4  
5  
6 cancers (i.e., cervical, endometrial, and ovarian) and breast cancer in our cohort. Also, we  
7  
8  
9 checked the external validity of our cohort by confirming the cancer incidence.  
10  
11  
12  
13  
14

## 15 **METHODS**

### 16 17 18 **Subjects**

19  
20  
21 The JNHS is an ongoing prospective cohort study investigating the association between  
22  
23 lifestyle, health care practices and women's health in Japan. Detailed information on its  
24  
25 design, population, protocol, and sample-size calculations were published previously <sup>4,5</sup>.  
26  
27  
28 Briefly, the baseline survey was conducted from 2001–2007, with planned follow-up for  
29  
30 30 years. In total, 15,019 women agreed to follow-up, signing and returning the informed-  
31  
32 consent form with the completed survey. At the time of the baseline survey, the study  
33  
34 population consisted of female licensed nursing professionals, including registered nurses,  
35  
36 licensed practical nurses, public health nurses, and midwives, aged  $\geq 25$  years, and  
37  
38 residing in Japan. Follow-up is currently ongoing; subjects are regularly mailed a self-  
39  
40 administered questionnaire once every 2 years to complete and return by post.  
41  
42  
43  
44  
45  
46  
47  
48  
49

50  
51 Before initiating the JNHS, the feasibility of its research strategy and the validity  
52  
53 of its questionnaires were investigated and confirmed in a pilot cohort study started in  
54  
55 1999 (the Gunma Nurses' Health Study; GNHS, n=698) <sup>6,7</sup>. We combined the JNHS and  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6 GNHS datasets in the present work as JNHS cohort (n=15,717). Table 1 shows the  
7  
8 number of subjects in each age group. Women had a mean age at baseline of 41.6 (8.3)  
9  
10 years (mean (SD) ; median: 41 years) and a mean follow-up of 10.5 (3.8) years (median:  
11  
12  
13  
14  
15  
16 12 years).

17  
18 The JNHS Coordination and Data Center is located in the Epidemiological  
19  
20 Research Office of the School of Health Sciences at Gunma University. This study was  
21  
22 performed under the Declaration of Helsinki, the Guidelines for Good Epidemiology  
23  
24 Practices <sup>8</sup>, and the Japanese Ethical Guidelines for Epidemiological Research <sup>9</sup>. The  
25  
26 GNHS study protocol was approved by the institutional review board of Gunma  
27  
28 University, Japan (approval # 3, 1999), and the JNHS study protocol was approved by  
29  
30 the institutional review board of Gunma University, Japan (approval #101, 2001) and the  
31  
32 ethics review board of Japan's National Institute of Public Health, Japan (approval #  
33  
34 03007, 2003).

### 47 48 **Patient and Public Involvement statement**

49  
50  
51 This research was done without involving participants in defining the research question,  
52  
53 outcome measures or study design. Participants were recruited with the study  
54  
55 information to nursing society. They were not invited to comment on the design and to  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 interpret the results, and were not invited to contribute to the writing or editing of the  
7  
8  
9 manuscript. The results will be reported to participants in the JNHS newsletter, and also  
10  
11  
12 be posted on the Website of JNHS.  
13  
14  
15  
16  
17

### 18 **Data collection and corroboration**

19  
20  
21 In the baseline and regular biennial follow-up questionnaires, women were asked, “Have  
22  
23 you ever been diagnosed with breast cancer (cervical cancer, endometrial cancer, or  
24  
25 ovarian cancer) by a medical doctor?”, and if so, what was their age at first diagnosis. We  
26  
27 identified and isolated those women who self-reported new incidences of one of the  
28  
29 cancers of interest in the regular follow-up by July 2017.  
30  
31  
32  
33  
34  
35

36 To corroborate the self-reported positive cases, an additional confirmation  
37  
38 questionnaire was sent to those women who affirmed a new cancer diagnosis in the  
39  
40 regular follow-up. Subjects were again asked the same question as above and to provide  
41  
42 details about their date of/age at diagnosis, method of detection, tumor stage, and  
43  
44 treatment history. We also asked for permission to access their medical records; if they  
45  
46 consented, we reviewed the records to obtain accurate clinical information on their  
47  
48 condition. For gynecological cancers, the data collected included date of diagnosis,  
49  
50 clinical stage, histological type, treatments, and concomitant cancer(s). For breast cancer,  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 the data included date of diagnosis, tumor site, invasivity, Tumor-Node-Metastasis  
7  
8  
9 classification (Union for International Cancer Control, 7<sup>th</sup> ed.)<sup>10</sup>, diagnostic method(s),  
10  
11  
12 tumor size, mammography category, surgical procedure, histological classification, and  
13  
14  
15 pathological classification (i.e., regional lymph node involvement (pN) and hormone  
16  
17  
18 receptor positivity for estrogen receptor (ER), progesterone receptor (PgR), and human  
19  
20  
21 epidermal growth factor receptor type2 (HER2)). This clinical information was  
22  
23  
24 furnished to an expert review panel comprising specialists on gynecological and breast  
25  
26  
27 cancers to verify each self-reported positive diagnosis.  
28  
29

30  
31 In Japan, the clinical reporting of gynecologic cancers follows the Japan Society of  
32  
33  
34 Obstetrics and Gynecology (JSOG) staging system, which is based on the internationally  
35  
36  
37 recognized surgical staging system published by the International Federation of  
38  
39  
40 Gynecology and Obstetrics (FIGO). When the FIGO criteria were updated during the  
41  
42  
43 study period in 2011<sup>11</sup>, the JSOG system was revised in tandem to remove Stage 0 lesions  
44  
45  
46 from the corresponding definitions, i.e., cervical carcinoma *in situ* (CIS) and atypical  
47  
48  
49 endometrial hyperplasia from cervical and endometrial cancer, respectively. Therefore,  
50  
51  
52 stage 0 cancers were not considered positive in our primary analysis, and all medical  
53  
54  
55 records were double-checked for patients who self-reported a new incidence of  
56  
57  
58 gynecological cancer before 2011. These borderline cases were excluded.  
59  
60

1  
2  
3  
4  
5  
6  
7 If a subject was reported as deceased or inexplicably failed to complete any recent  
8  
9 study activities, we established a cause of death by checking it against death certificate-  
10  
11 related information in Japan's National Vital Statistics database.  
12  
13  
14  
15  
16  
17

## 18 **Validation**

19  
20  
21 Regular follow-up, confirmation questionnaires, and expert review were validated for  
22  
23 their positive predictive value (PPV) and negative predictive value (NPV) for new  
24  
25 incidences of each cancer.  
26  
27  
28  
29

30 For the first two sources, the validation sample included all members of the study  
31  
32 cohort (n=15,717) who reported no past history of the cancer in question at baseline. The  
33  
34 PPV of the regular follow-up was calculated as the number of verified positive cases of  
35  
36 the cancer, i.e., cases whose self-reported positive diagnosis was verified by medical-  
37  
38 record review or cause-of-death investigation, divided by all cases of self-reported new  
39  
40 incidences of the cancer in the regular follow-up. The NPV was calculated as the number  
41  
42 of suspected negative cases, divided by all members of the validation sample who self-  
43  
44 reported no new cancer incidence in the regular follow-up. Here, the suspected negative  
45  
46 cases consisted of all members of the validation sample for the cancer in question minus  
47  
48 A) cases who self-reported new incidences in the regular follow-up and B) positive cases  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 whose status was established only by death certificate (DCO).  
7  
8

9 The PPV of the combined regular follow-up and confirmation questionnaire was  
10 calculated as the number of verified positive cases of the cancer divided by all cases who  
11 corroborated their positive diagnosis on the confirmation questionnaire. The NPV was  
12 calculated as the number of suspected negative cases divided by all members of the  
13 validation sample except those who self-reported their positive diagnosis on the  
14 confirmation questionnaire. Here, the suspected negative cases consisted of all members  
15 of the validation sample minus A) cases who self-reported their positive diagnosis on the  
16 confirmation questionnaire, B) cases ruled positive by DCO, C) cases ruled positive by  
17 cause-of-death investigation, and D) contradictory cases (i.e., women confirmed by  
18 expert review but self-reported a negative status on the confirmation questionnaire or left  
19 the field blank).  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

42 The expert review panel's judgments were also validated for comparison. In this  
43 analysis, the validation sample consisted of all participants who A) returned the  
44 confirmation questionnaire, B) permitted the research team to contact their diagnosing  
45 facility, and C) their provider agreed to respond to the team's inquiry. The PPV was  
46 calculated as the number of cases verified as positive by the diagnosing facility, divided  
47 by the number of cases ruled positive by the expert review panel. The NPV was calculated  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 as the number of cases verified as negative by the diagnosing facility, divided by the  
7  
8  
9 number of cases ruled negative by the panel.  
10

11  
12 After fixing the cancer cases, the incidence rate of each cancer was estimated  
13  
14 from the observed events and person-time at risk for 10 years of observation. Because  
15  
16 of the total numbers and events of the patients during the observation period, the 30–60  
17  
18 year-old age group was used. We calculated the 95% confidence intervals of the  
19  
20 incidence rates based on the exact Poisson confidence interval in accordance with  
21  
22 known methods <sup>12</sup>.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

## 33 **RESULTS**

### 34 **Verified cases of each cancer type**

35  
36 The flow diagram illustrating the validation process of each cancer is listed in the Web  
37  
38 Appendices. The numbers of new cases of self-reported cancers in the regular follow-up  
39  
40 (and incidences in the respective validation sample) were cervical cancer: 219 (1.4%),  
41  
42 endometrial cancer: 83 (0.5%), ovarian cancer: 51 (0.3%), and breast cancer: 365 (2.3%).  
43  
44  
45  
46  
47  
48  
49 New incidence was verified by expert review in 37, 45, 23, and 297 of these cases,  
50  
51 respectively. Some subjects sent the confirmation questionnaire corroborating their  
52  
53 positive diagnosis but were ruled negative by the expert panel (72.1%, 11.1%, 30.3%, and  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 1.0%, respectively), while 37.6%, 33.8%, 25.6%, and 8.3% of subjects, respectively,  
7  
8  
9 responded with negative diagnosis on the confirmation questionnaire.  
10

11  
12 For all observed cases of mortality, cause of death was established as being cervical  
13  
14 cancer (n=4, DCO=0), endometrial cancer (n=7, DCO=2), ovarian cancer (n=3, DCO=3),  
15  
16 or breast cancer (n=16, DCO=3). New incidences of the four cancers were verified in 37,  
17  
18  
19  
20  
21  
22 47, 26, and 300 cases, respectively.  
23

24  
25 In the JNHS cohort, the estimated incidence rates for patients aged 30–60 years  
26  
27 were 22.0/100,000; 25.4/100,000; 13.8/100,000; and 160.4/100,000 person-years for  
28  
29  
30 cervical, endometrial, ovarian, and breast cancer, respectively (Table 2). Considering the  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
lack of heterogeneity between this cohort and Japanese women overall, the incidence rates  
for each age group were compared with the national data from regional cancer registries  
in the 2015 statistics published by Japan's National Cancer Center <sup>13</sup> (Figure 1). For all  
four cancers, the cohort data did not deviate from the national data.

### **Self-reported PPV/NPV for each cancer**

Table 3 summarizes the PPVs and NPVs for the regular follow-up, regular follow-up plus  
confirmation questionnaire, and expert review for the new incidence of each cancer.

Expert review achieved 100% accuracy for each cancer except cervical (PPV:

1  
2  
3  
4  
5  
6 92.3%) because of a single false-positive case, which the participant's provider clarified  
7  
8  
9 to be a different condition.  
10

11  
12 Self-reporting achieved NPVs near 100% for all cancers for both the regular  
13  
14 follow-up and the regular follow-up plus confirmation questionnaire. However, the  
15  
16 corresponding PPVs tended to be somewhat lower and variable across cancers. The PPVs  
17  
18 were worse for gynecological cancers than for breast cancer (breast > endometrial >  
19  
20 ovarian > cervical, in descending order) for both follow-up sources. The PPV for uterine  
21  
22 cancer, which included cervical and endometrial cancers, was 27.2%.  
23  
24  
25  
26  
27  
28  
29

30 The regular follow-up plus confirmation questionnaire achieved higher PPVs in all  
31  
32 cases than did regular follow-up alone; however, while it achieved 99.0% accuracy for  
33  
34 breast cancer, the estimates were lower for endometrial (88.9%) and ovarian (76.7%)  
35  
36 cancer and poor for cervical cancer (31.5%).  
37  
38  
39  
40  
41

42 Considering the changes to the official JSOG clinical staging system during the  
43  
44 survey period, we calculated a similar summary for PPVs and NPVs, adding cases of  
45  
46 cervical CIS, atypical endometrial hyperplasia, and borderline ovarian tumors (Table 4).  
47  
48  
49  
50

51 The resulting PPVs were uniformly higher when all three cancers were included  
52  
53 than when they were excluded. For endometrial and ovarian cancer, the improvements  
54  
55 ranged from 3.3%–6.7%, but for cervical cancer, their inclusion almost doubled the  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6 predictive value for both the regular follow-up and regular follow-up plus confirmation  
7  
8  
9 questionnaire, at +20.1% and +37.9%, respectively.  
10  
11  
12  
13  
14  
15  
16  
17

## 18 **DISCUSSION**

21 In the JNHS cohort, self-reporting in regular follow-up achieved a PPV of 81.4% for  
22  
23 breast cancer but performed poorer for gynecological cancers, especially uterine cancers  
24  
25 (PPV: 27.2%) and cervical cancer alone (PPV: 16.9%). Our PPVs were higher than the  
26  
27 corresponding values reported by the Japan Public Health Center (JPHC) Study, a  
28  
29 population-based prospective cohort study (all cancers in women: 54.2%, breast: 58.4%,  
30  
31 uterine: 21.7%)<sup>14</sup>. The validity of self-reporting is associated with individual  
32  
33 characteristics<sup>15</sup>, and our cohort consisted entirely of nursing professionals. While  
34  
35 evidence suggests that educational level has a negligible association with validity<sup>16</sup>, we  
36  
37 partially attribute the high self-reporting accuracy to the uniformly high level of medical  
38  
39 education and deeper knowledge of cancer in our cohort than in the general population.  
40  
41 Other studies support this argument<sup>17</sup>. However, sizeable percentages of nurses who  
42  
43 affirmed new incidences of cancer in the regular follow-up gave the opposite response on  
44  
45 the confirmation questionnaire (gynecological cancer: 25.6%–37.6%, breast cancer:  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 8.3%). Similarly, considerable percentages of respondents to the confirmation  
7  
8  
9  
10 questionnaire were verified not to have cancer (gynecological cancer: 41.2%–81.2%,  
11  
12 breast cancer: 9.2%). Many who corroborated their self-reported positive diagnosis were  
13  
14 eventually ruled negative by expert review, especially for cervical cancer (72.1%),  
15  
16 followed by ovarian (30.3%), endometrial (11.1%) and breast (1.0%). In summary, self-  
17  
18 reporting alone apparently fails to capture the real cancer incidence, even for this cohort  
19  
20 of nursing professionals with uniformly high medical knowledge. Additional inquiries to  
21  
22 confirm the details are needed.  
23  
24  
25  
26  
27  
28  
29

30  
31 Compared to PPVs of self-report validity in other prospective cohort study datasets  
32  
33 <sup>16, 18, 19</sup>, our PPVs were comparable to the literature values for breast cancer but lower  
34  
35 than these values for uterine cancers. Many studies have shown that self-reporting of  
36  
37 breast cancer has high PPVs <sup>10, 16, 19</sup>. Some evidence has linked higher educational levels  
38  
39 with a greater risk of breast cancer <sup>20</sup>, which may also be true for our cohort. Additionally,  
40  
41 breast cancer diagnoses included ductal carcinoma *in situ*, which may have led to less  
42  
43 confusion than with gynecological cancers that excluded stage 0 cases and borderline  
44  
45 tumors.  
46  
47  
48  
49  
50  
51  
52  
53

54  
55 Studies outside of Japan have also found self-reporting to yield lower PPVs for  
56  
57 uterine cancers than for other cancers <sup>18, 21</sup>, for several possible reasons. One is inaccurate  
58  
59  
60

1  
2  
3  
4  
5  
6 memory of precancerous cervical lesions, which are rarely addressed immediately by  
7  
8  
9 surgical intervention. Additionally, age and sex may have some association; for example,  
10  
11  
12 participants >50 years old in a Native American cohort were more likely to report  
13  
14  
15 incorrectly<sup>22</sup>. Further, a study from Australia found that self-reported breast cancer had  
16  
17  
18 lower PPVs in women aged 70–75 years<sup>23</sup>. Disease-specific considerations may also be  
19  
20  
21 relevant. One study noted that many cases of women’s cancers, especially cervical cancer,  
22  
23  
24 are not recorded in cancer registries<sup>22</sup>, while another estimated false-negative rates of  
25  
26  
27 43.8%, 28.6%, and 20.8% for self-reports of uterine, ovarian, and breast cancers,  
28  
29  
30 respectively<sup>24</sup>. Differences in incidence must also be considered. Because gynecological  
31  
32  
33 cancers are >5 times less prevalent than breast cancer, a difference of one case would  
34  
35  
36 produce a proportionally larger change in PPV.  
37  
38

39  
40 One problem specific to Japan regarding the self-reporting of women’s cancers is  
41  
42 how the results of cytological screening tests are reported for cervical and endometrial  
43  
44  
45 cancers. Today, Pap smear results are recorded using the Bethesda system, the standard  
46  
47  
48 international format, but these results previously followed a class-based system. Class II  
49  
50  
51 status, which shows within the normal range is sometimes confused with stage II cervical  
52  
53  
54 cancer. Similarly, atypical endometrial hyperplasia was previously classified as stage 0  
55  
56  
57 endometrial cancer, which may be confused with non-atypical endometrial hyperplasia.  
58  
59  
60

1  
2  
3  
4  
5  
6  
7 We suspect that another reason the self-report validity in our cohort was so poor  
8  
9 for certain cancers was that subjects were recalling their past medical history during the  
10 regular follow-up, rather than the new incidence as intended. Additionally, ambiguous  
11 language in the questionnaire, such as “dysplasia” or “precancerous lesions”, may have  
12 reduced the self-report validity, as evidenced by the higher PPVs for borderline forms,  
13 such as cervical CIS, atypical endometrial hyperplasia, and borderline ovarian tumors  
14 included in the analysis. Among the three borderline forms, classifying cervical CIS as  
15 cervical cancer led to a greater increase in PPVs than did other cancers. Manjer et al. also  
16 found that self-reporting of malignant cervical cancer was less sensitive when the  
17 definition included cervical CIS<sup>25</sup>. These considerations suggest that compared with  
18 other cancers, diagnoses of cervical cancer and precancerous lesions have a greater risk  
19 of being inaccurately communicated or negatively interpreted by patients.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

42 One of this study’s strengths was our meticulous review of subjects’ medical  
43 records and death-certificate surveys to establish self-report validity, circumventing the  
44 limitations presented by Japan’s lack of complete national cancer registries. Additionally,  
45 we believe that our data better reflect the general Japanese population than did past  
46 findings for other regional cohorts because the nationwide scope of the JNHS minimizes  
47 the geographical variation. Moreover, our cohort was relatively homogenous in terms of  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 sex and occupation, consisting entirely of female nursing professionals.  
7  
8

9 The study also had some limitations. Cohort-specific characteristics may limit the  
10 generalizability of our findings, especially the relatively young skew of the participants'  
11 ages. However, when converted to incidence rates, our rates seem most consistent with  
12 the 2015 statistics published by Japan's National Cancer Center <sup>13</sup>. Additionally, self-  
13 reported diagnoses could not be verified in some cases. Our expert panel made their  
14 judgments based on the specific language nurses used in the questionnaire to describe  
15 their treatments such as "hysterectomy" and "chemotherapy", but the panel still  
16 encountered cases that were difficult to definitively verify. However, we established a  
17 conclusive diagnosis based on all available information such as postmortem exam  
18 findings and supplemental details from primary-care providers. No indeterminate cases  
19 were found among those lacking medical records for verification.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

42 The JNHS database covers all of Japan. The confirmation of external validity of  
43 these cancer incidence in this cohort will allow the further investigation of risk factors for  
44 different cancers such as menopausal hormone therapy and lifestyle factors and their  
45 associations, with unaffected by information bias. We plan to continue our work by  
46 analyzing the respective contributions of different risk factors among confirmed cases of  
47 gynecological and breast cancer, as verified above.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 **Acknowledgments** The authors appreciate the cooperation of the Japanese nurses who  
7  
8  
9 participate in the JNHS and the GNHS. We also thank Ms. Satomi Shimizu at the JNHS  
10  
11  
12 Data Center for her help with data management.  
13  
14  
15

16  
17  
18 **Contributors** KT was the panel member and wrote the initial draft of the paper to which  
19  
20  
21 all authors contributed. YI, NM, and KN collected data and analyzed. MK and TY were  
22  
23  
24 panel members and revised manuscript. KH designed the study, raised funding, and  
25  
26  
27 directed its implementation including quality assurance and control.  
28  
29  
30

31  
32  
33 **Funding** This work was partly supported by a grant from the Japan Society for the  
34  
35  
36 Promotion of Science (JSPS KAKENHI Grant Number: 18H04069).  
37  
38  
39

40  
41  
42 **Competing interests** None declared.  
43  
44  
45

46  
47  
48 **Patient consent for publication** Not required.  
49  
50  
51

52  
53  
54 **Data sharing statement** The data are not publically available due to data transfer  
55  
56  
57 agreements. No additional data is available.  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Provenance and peer review** Not commissioned; externally peer reviewed.

For peer review only

**REFERENCES**

- 1 Jackson JM, DeFor TA, Crain AL, et al. Validity of diabetes self-reports in the Women's Health Initiative. *Menopause* 2014;21(8):861-868.
- 2 Gentry-Maharaj A, Taylor H, Kalsi J, et al. Validity of self-reported hysterectomy: a prospective cohort study within the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *BMJ Open* 2014;4(3):e004421.
- 3 Skeie G, Mode N, Henningsen M, et al. Validity of self-reported body mass index among middle-aged participants in the Norwegian Women and Cancer study. *Clin Epidemiol* 2015;7:313-323.
- 4 Hayashi K, Mizunuma H, Fujita T, et al. Design of the Japan Nurses' Health Study: A prospective occupational cohort study of women's health in Japan. *Ind Health* 2007;45(5):679-686.
- 5 Fujita T, Hayashi K, Katanoda K, et al. Prevalence of diseases and statistical power of the Japan Nurses' Health Study. *Ind Health* 2007;45(5):687-694.
- 6 Maeno T, Ohta A, Hayashi K, et al. Impact of reproductive experience on women's smoking behaviour in Japanese nurses. *Public Health* 2005;119(9):816-824.



- 1  
2  
3  
4  
5  
6  
7 7 Kato C, Shimada J, Hayashi K. Sleepiness during shift work in Japanese nurses: A  
8  
9 comparison study using JESS, SSS, and actigraphy. *Sleep Biol Rhythm*  
10  
11  
12 2012;10(2):109-117.  
13  
14
- 15 8 Andrews EA, Avorn J, Bortnichak EA, et al. ; ISPE. Guidelines for Good  
16  
17 Epidemiology Practices for drug, device, and vaccine research in the United States.  
18  
19 *Pharmacoepidemiol Drug Saf* 1996;5(5):333-338.  
20  
21  
22
- 23 9 Ministry of Education, Culture, Sports, Science and Technology and Ministry of  
24  
25 Health, Labour and Welfare. Ethical Guidelines for Epidemiological Research.  
26  
27  
28 <https://www.niph.go.jp/wadai/ekigakurinri/guidelines.pdf>. Accessed August 24,  
29  
30  
31 2020.  
32  
33  
34
- 35 10 International Union Against Cancer. TNM Classification of Malignant Tumours 7<sup>th</sup>  
36  
37 ed. Eds. Sobin LH, Gospodarowicz MK and Wittekind Ch, Wiley, New York, 2009.  
38  
39  
40
- 41 11 FIGO committee on gynecologic oncology. Revised FIGO staging for carcinoma of  
42  
43 the vulva, cervix, and endometrium. *Int J Gynecol Obstet* 2009;105(2):103-104.  
44  
45  
46  
47
- 48 12 Ulm K. A simple method to calculate the confidence interval of a standardized  
49  
50 mortality ratio (SMR). *Am J Epidemiol* 1990;131(2):373-375.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- 13 Cancer Registry and Statistics. Cancer Information Service, National Cancer Center, Japan (Monitoring of Cancer Incidence in Japan (MCIJ)).
- [https://ganjoho.jp/reg\\_stat/statistics/stat/summary.html](https://ganjoho.jp/reg_stat/statistics/stat/summary.html). Accessed August 24, 2020.
- 14 Inoue M, Sawada N, Shimazu T, et al. Japan Public Health Center-based Prospective Study Group. Validity of self-reported cancer among a Japanese population: recent results from a population-based prospective study in Japan (JPHC Study). *Cancer Epidemiol* 2011;35(3):250-253.
- 15 Jackson CA, Mishra GD, Tooth L, et al. Moderate agreement between self-reported stroke and hospital-recorded stroke in two cohorts of Australian women: a validation study. *BMC Med Res Methodol* 2015;15:7.
- 16 Bergmann MM, Calle EE, Mervis CA, et al. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol* 1998;147(6):556-562.
- 17 Colditz GA, Manson JE, Hankinson SE. The Nurses' Health Study: 20-year contribution to the understanding of health among women. *J Womens Health* 1997;6(1):49-62.

- 1  
2  
3  
4  
5  
6  
7 18 Navarro C, Chirlaque MD, Tormo MJ, et al. Validity of self reported diagnoses of  
8  
9 cancer in a major Spanish prospective cohort study. *J Epidemiol Community Health*  
10  
11 2006;60(7):593-599.  
12  
13  
14  
15  
16 19 Loh V, Harding J, Koshkina V, et al. The validity of self-reported cancer in an  
17  
18 Australian population study. *Aust NZ J Public Health* 2014;38(1):35-38.  
19  
20  
21  
22 20 Goldberg M, Calderon-Margalit R, et al. Socioeconomic disparities in breast cancer  
23  
24 incidence and survival among parous women: findings from a population-based  
25  
26 cohort, 1964-2008. *BMC Cancer* 2015;15:921.  
27  
28  
29  
30  
31 21 Cho S, Shin A, Song D, et al. Validity of self-reported cancer history in the health  
32  
33 examinees (HEXA) study: A comparison of self-report and cancer registry records.  
34  
35  
36  
37 *Cancer Epidemiol* 2017;50(Pt A):16-21.  
38  
39  
40  
41 22 Nash SH, Day G, Hiratsuka VY, et al. Agreement between self-reported and central  
42  
43 cancer registry-recorded prevalence of cancer in the Alaska EARTH study. *Int J*  
44  
45  
46 *Circumpolar Health* 2019;78(1):1571383.  
47  
48  
49  
50 23 Stavrou E, Vajdic CM, Loxton D, et al. The validity of self-reported cancer  
51  
52 diagnoses and factors associated with accurate reporting in a cohort of older  
53  
54 Australian women. *Cancer Epidemiol* 2011;35(6):e75-80.  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4  
5  
6  
7 24 Desai MM, Bruce ML, Desai RA, et al. Validity of self-reported cancer history: a  
8  
9 comparison of health interview data and cancer registry records. *Am J Epidemiol*  
10  
11 2001;153(3):299-306.  
12  
13  
14  
15 25 Manjer J, Merlo J, Berglund G. Validity of self-reported information on cancer:  
16  
17 determinants of under- and over-reporting. *Eur J Epidemiol* 2004;19(3):239-247.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7 **Figure and table legends**  
8  
9

10  
11  
12 **Figure 1** Estimated incidence rates for each age group in the JNHS cohort and the  
13 national data from regional cancer registries.  
14  
15

16  
17  
18 (error bars shows the 95% confidence intervals)  
19

20  
21 A. cervical cancer  
22

23  
24 B. endometrial cancer  
25

26  
27 C. ovarian cancer  
28

29  
30 D. breast cancer  
31  
32  
33  
34  
35

36 **Table 1** Numbers and percentages of subjects in each age group at baseline in the JNHS  
37 cohort  
38  
39  
40  
41  
42  
43  
44

45 **Table 2** Estimated incidence rate of each cancer in patients aged 30–60 years in the  
46 JNHS cohort  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 **Table 3** PPVs/NPVs for regular follow-up, regular follow-up plus confirmation

7  
8  
9 questionnaire, and expert review for new incidences of gynecological and breast cancers  
10  
11  
12 in the JNHS cohort  
13  
14  
15  
16  
17

18 **Table 4** Corresponding PPVs/NPVs including those of cervical CIS, atypical endometrial  
19  
20 hyperplasia, and borderline ovarian tumor in the JNHS cohort  
21  
22  
23  
24  
25  
26  
27

## 28 **Appendix legends**

29  
30  
31  
32

33 **Appendix Figure 1** Flow diagram illustrating the validation process for self-reported  
34  
35 cervical cancer (BL: baseline, CIS: carcinoma in situ, DCO: death certificate only)  
36  
37  
38  
39  
40  
41

42 **Appendix Figure 2** Flow diagram illustrating the validation process for self-reported  
43  
44 endometrial cancer (BL: baseline, aEmH: atypical endometrial hyperplasia, DCO: death certificate  
45  
46 only)  
47  
48  
49  
50  
51

52  
53  
54 **Appendix Figure 3** Flow diagram illustrating the validation process for self-reported  
55  
56 ovarian cancer (BL: baseline, DCO: death certificate only)  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9 **Appendix Figure 4** Flow diagram illustrating the validation process for self-reported  
10  
11  
12 breast cancer (BL: baseline, CIS: carcinoma in situ, DCO: death certificate only)  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

**Table 1** Numbers and percentages of subjects in each age group at baseline in the JNHS cohort

Age (years)	n	(%)
< 30	692	(4.4)
30 – 34	2,955	(18.8)
35 – 39	3,176	(20.2)
40 – 44	3,133	(19.9)
45 – 49	2,767	(17.6)
50 – 54	2,012	(12.8)
55 – 59	797	(5.1)
60 – 64	143	(0.9)
≥ 65	42	(0.3)

**Table 2** Estimated incidence rate of each cancer in patients aged 30 to 60 years in the JNHS cohort

	Cancer cases	Person-years	Incidence rate (per 100,000 person-years)	Lower limit of 95% confidence interval	Upper limit of 95% confidence interval
Cervical cancer	29	131,658.50	22.0	14.8	31.6
Endometrial cancer	32	126041.0	25.4	17.4	35.8
Ovarian cancer	18	130662.5	13.8	8.2	21.8
Breast cancer	210	130960.5	160.4	139.4	183.6



**Table 3** PPVs/NPVs of regular follow-up, regular follow-up plus confirmation questionnaire, and expert review for new incidences of gynecological and breast cancers in the JNHS cohort

	Positive history self-reported at baseline	Validation sample	Positive diagnosis self-reported in regular follow-up	Positive status established by cause-of-death investigation (incl. DCO cases)	Regular follow-up		Regular follow-up plus confirmation questionnaire		Expert review	
					PPV	NPV	PPV	NPV	PPV	NPV
Cervical cancer	167	15,550	219	2 (0)	16.9%	100.0%	31.5%	99.9%	92.3%	100.0%
Endometrial cancer	31	15,686	83	7 (2)	54.2%	99.9%	88.9%	99.9%	100.0%	100.0%
Ovarian cancer	37	15,680	51	3 (3)	45.1%	99.9%	76.1%	99.9%	100.0%	100.0%
Breast cancer	138	15,579	365	5 (3)	81.4%	99.9%	99.0%	99.9%	100.0%	100.0%

**Table 4** Corresponding PPVs/NPVs including cervical CIS, atypical endometrial hyperplasia, and borderline ovarian tumors in the JNHS cohort

	Positive history self-reported at baseline	Validation sample	Positive diagnosis self-reported in regular follow-up	Positive status established by cause-of-death investigation (incl. DCO cases)	Regular follow-up		Regular follow-up plus confirmation questionnaire	
					PPV	NPV	PPV	NPV
Cervical cancer	167	15,550	219	2 (0)	37.0%	100.0%	69.4%	99.9%
Endometrial cancer	31	15,686	83	7 (2)	57.8%	99.9%	95.6%	99.9%
Ovarian cancer	37	15,680	51	3 (3)	49.0%	99.9%	80.0%	99.9%

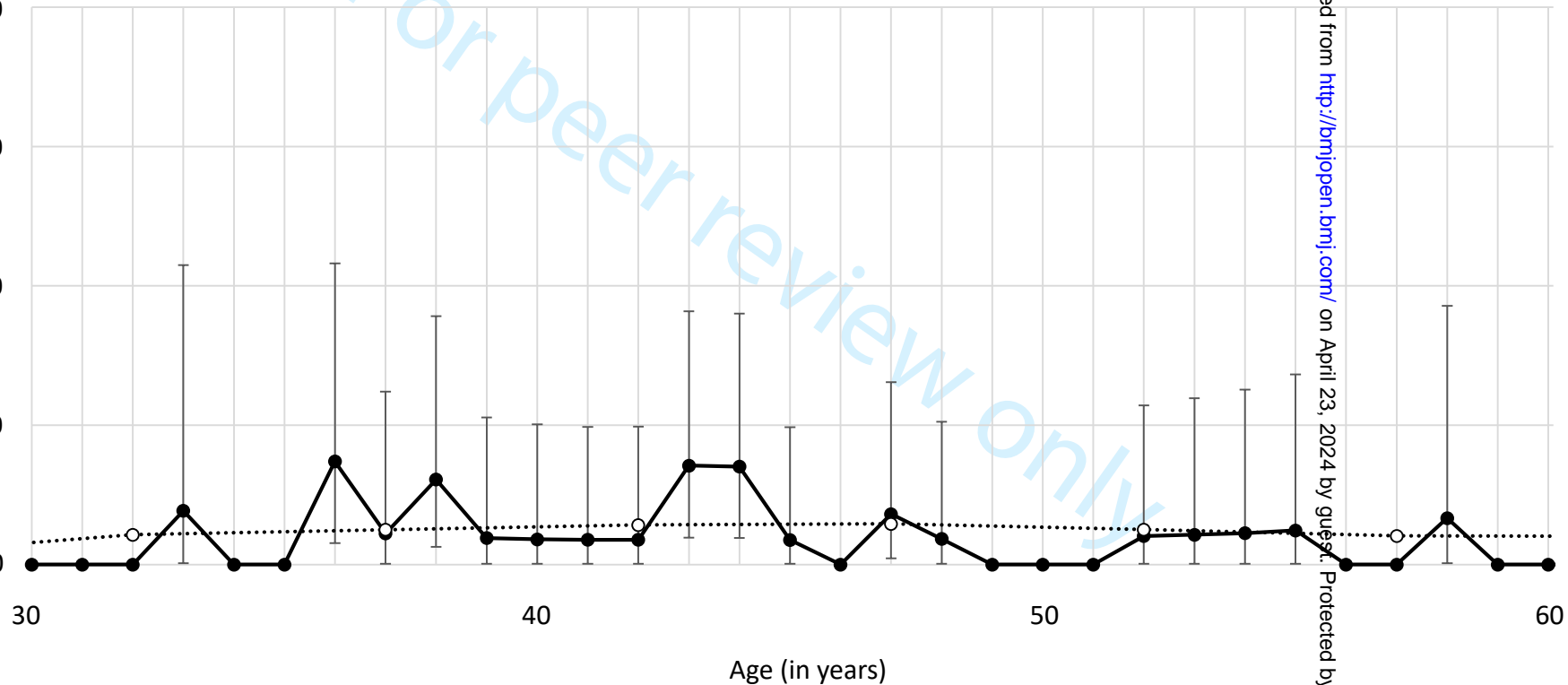
**Figure 1** Estimated incidence rates for each age group in the JNHS cohort and the national data from regional cancer registries (error bars shows 95% confidence intervals)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

A. Cervical cancer

Incidence rates  
(per 100,000 person-years)

● JNHS cohort  
○ National data from regional cancer registries

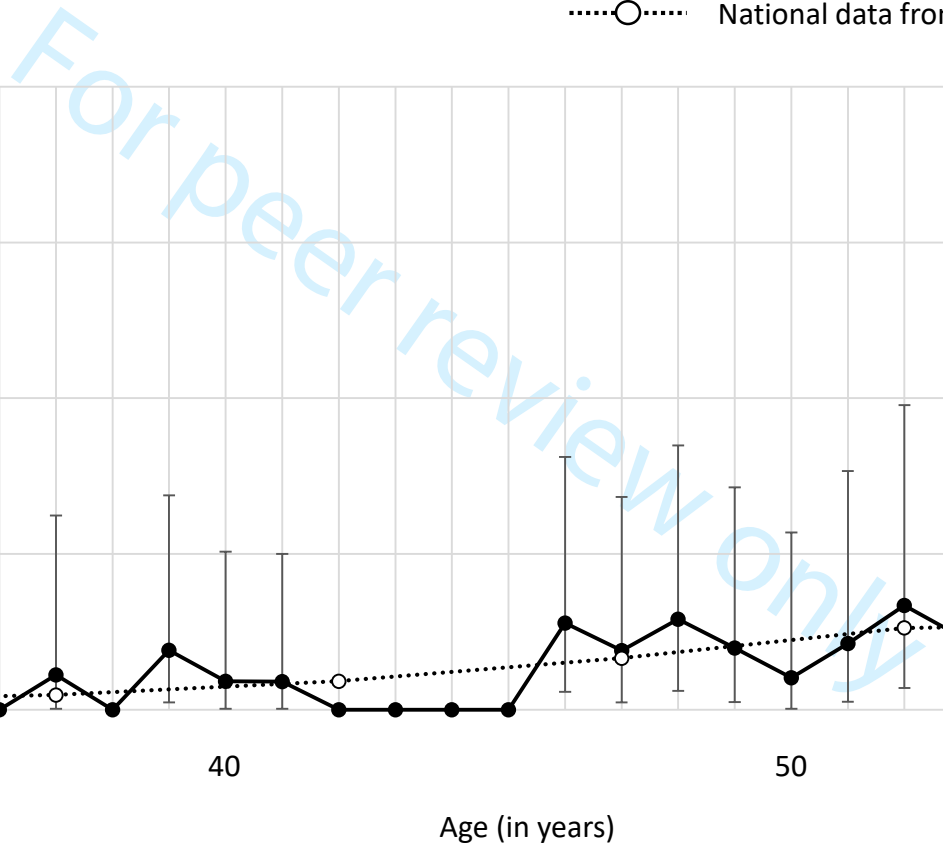


2020-05-14 19:11:49 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

1  
2 **B. Endometrial cancer**  
3  
4

5  
6  
7 Incidence rates  
8 (per 100,000 person-years)  
9

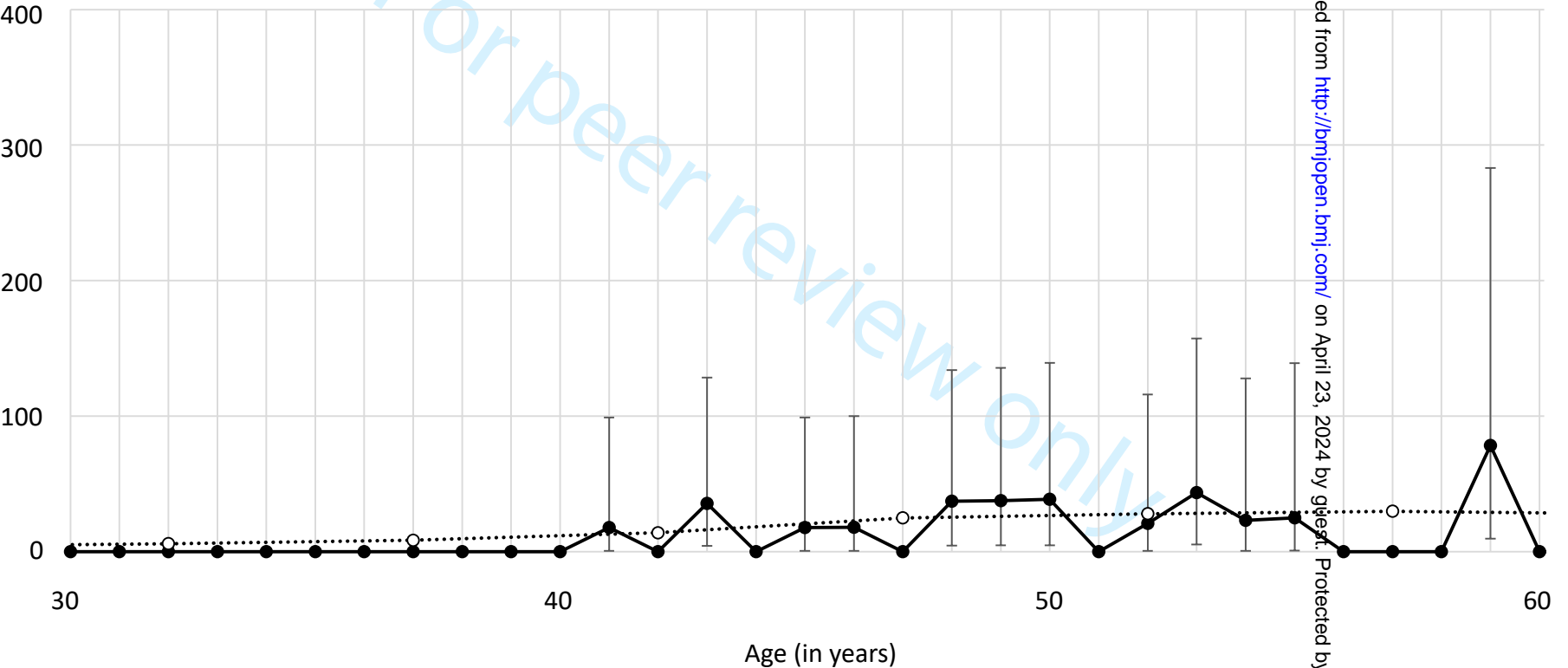
● JNHS cohort  
○ National data from regional cancer registries



C. Ovarian cancer

Incidence rates  
(per 100,000 person-years)

● JNHS cohort  
○ National data from regional cancer registries

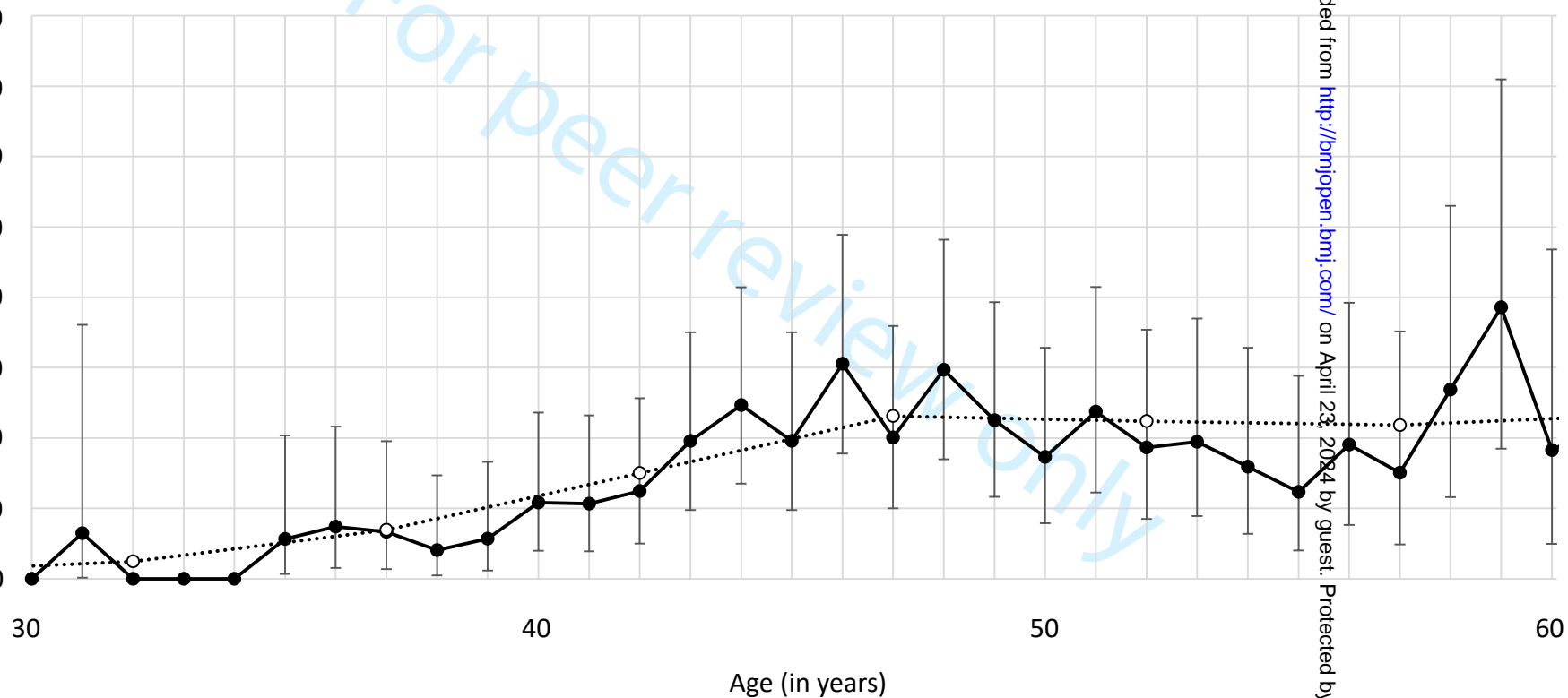


20-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

1  
2 D. Breast cancer  
3  
4  
5

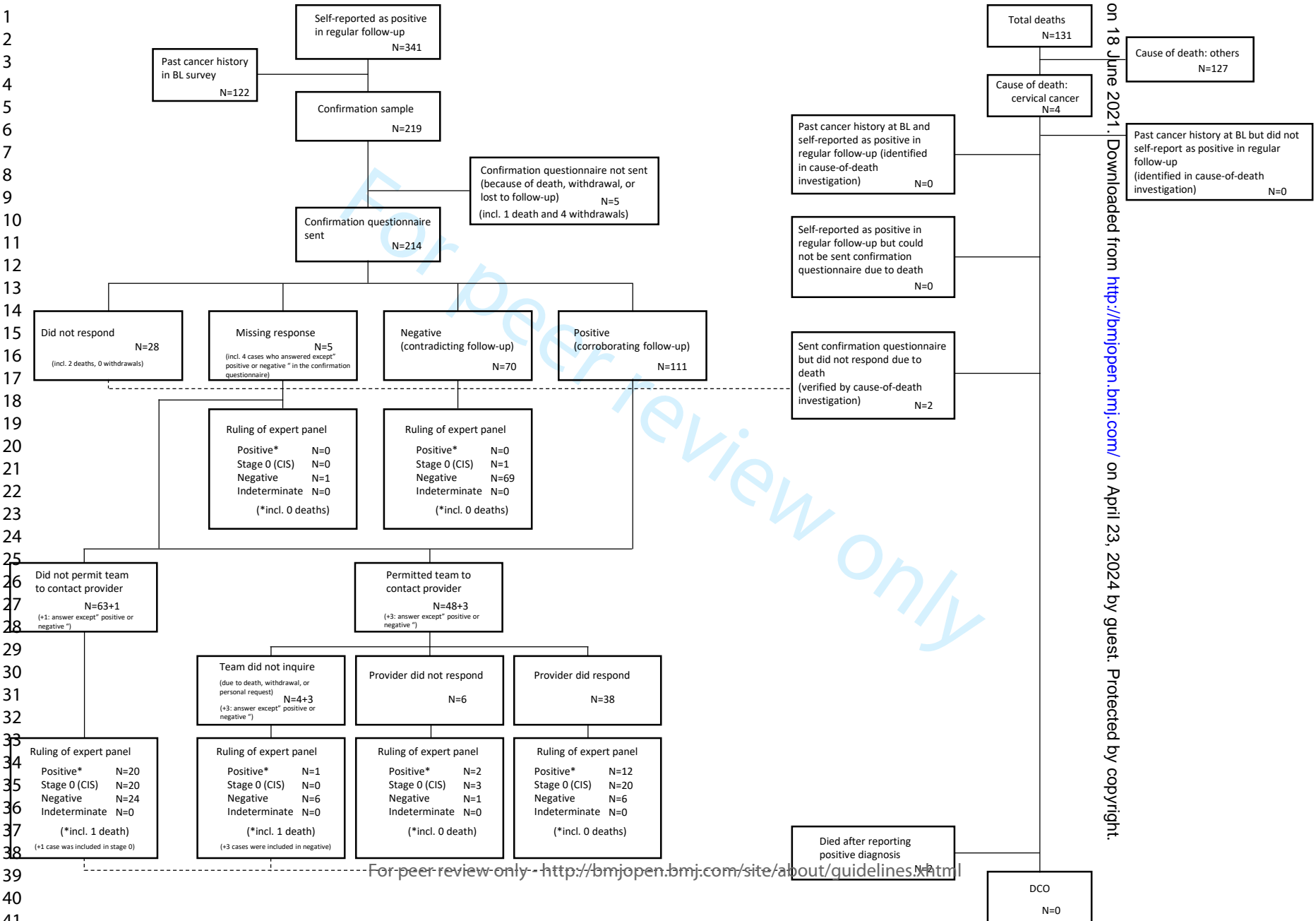
6 Incidence rates  
7 (per 100,000 person-years)  
8  
9

● JNHS cohort  
○ National data from regional cancer registries

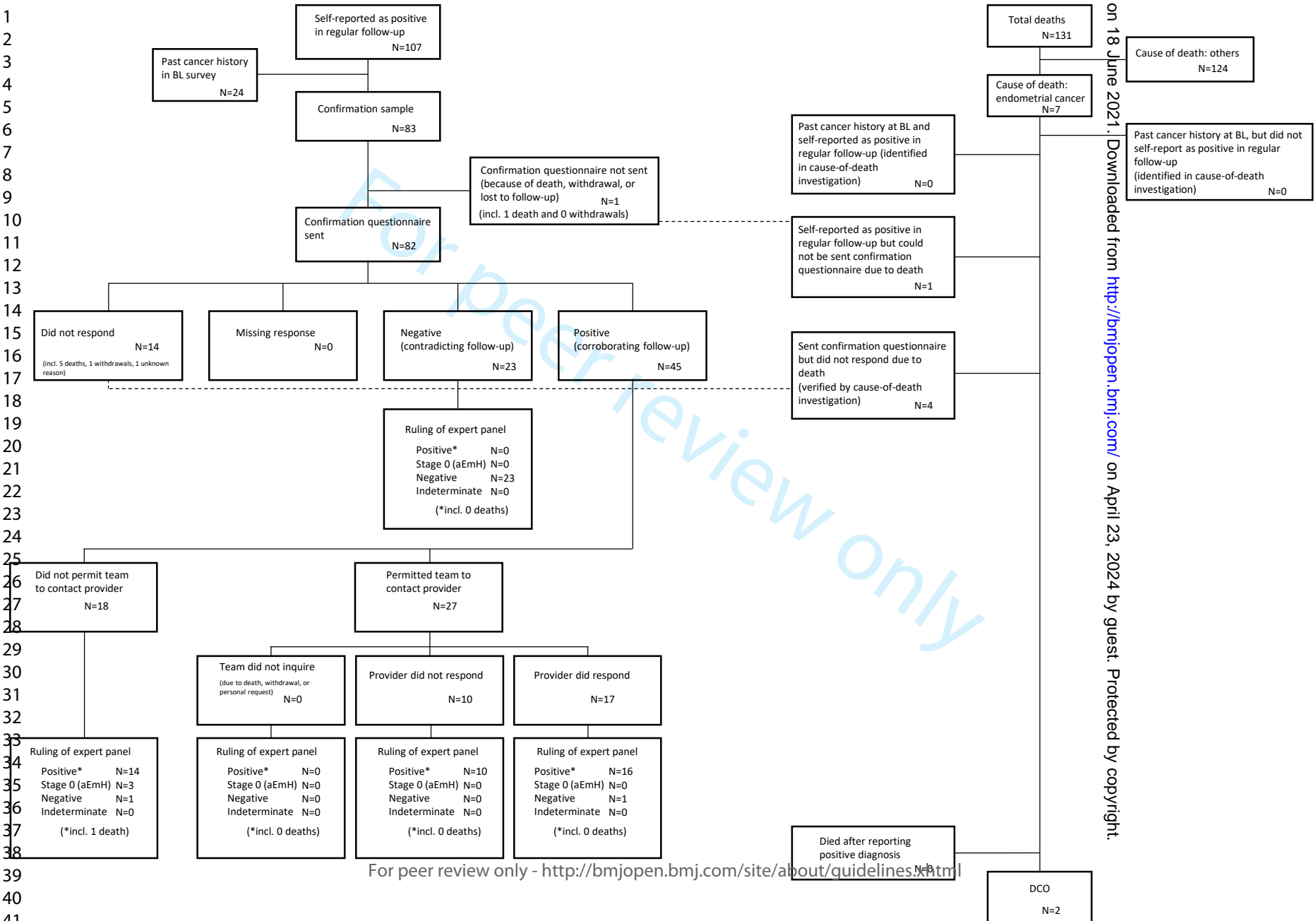


**Appendix Figure 1** Flow diagram illustrating the validation process for self-reported cervical cancer (BL: baseline, CIS: carcinoma in situ, DCO: death certificate only)

20-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

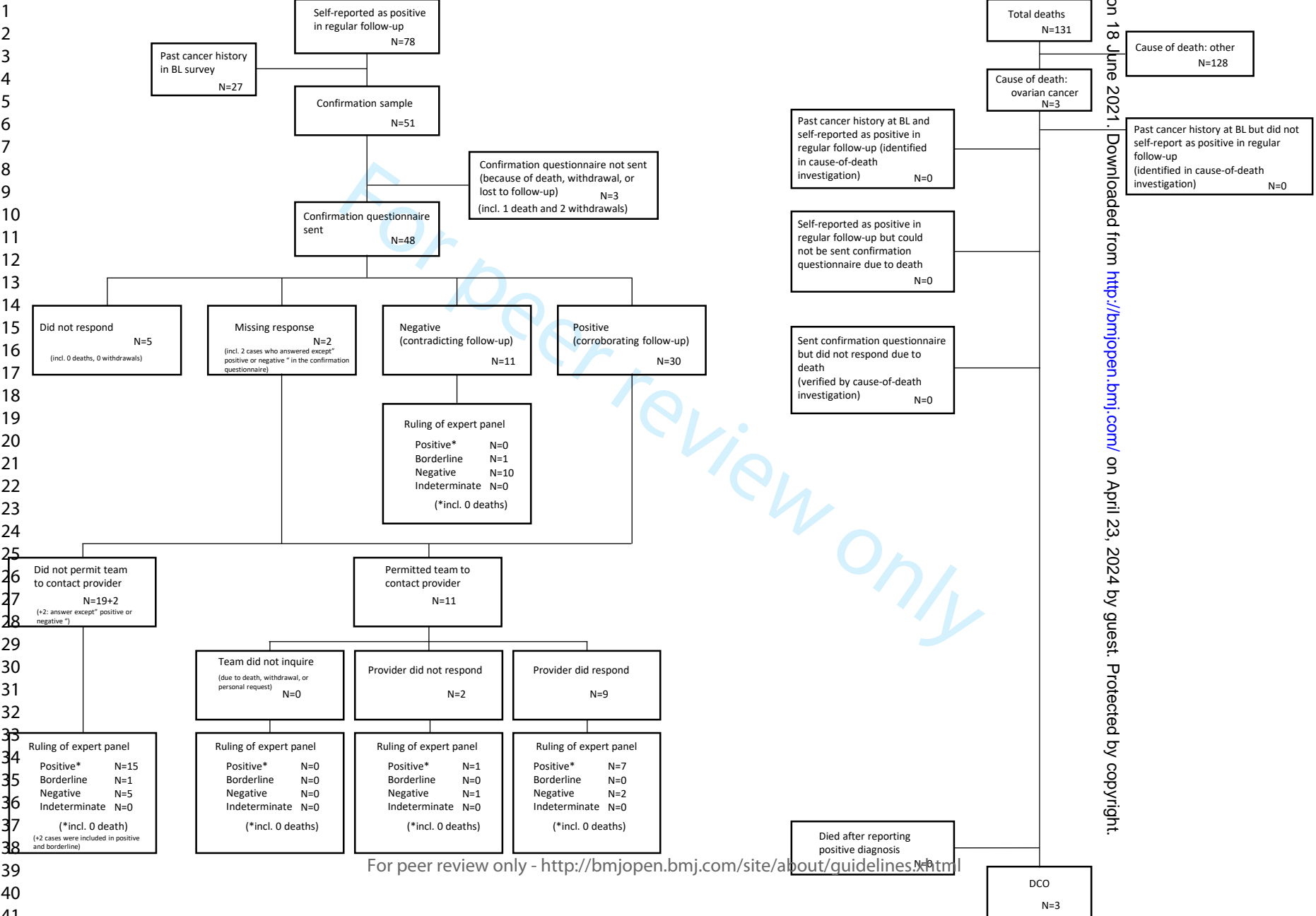


Appendix Figure 2 Flow diagram illustrating the validation process for self-reported endometrial cancer (BL: baseline, aEmH: atypical endometrial hyperplasia, DCO: death certificate only)



20-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

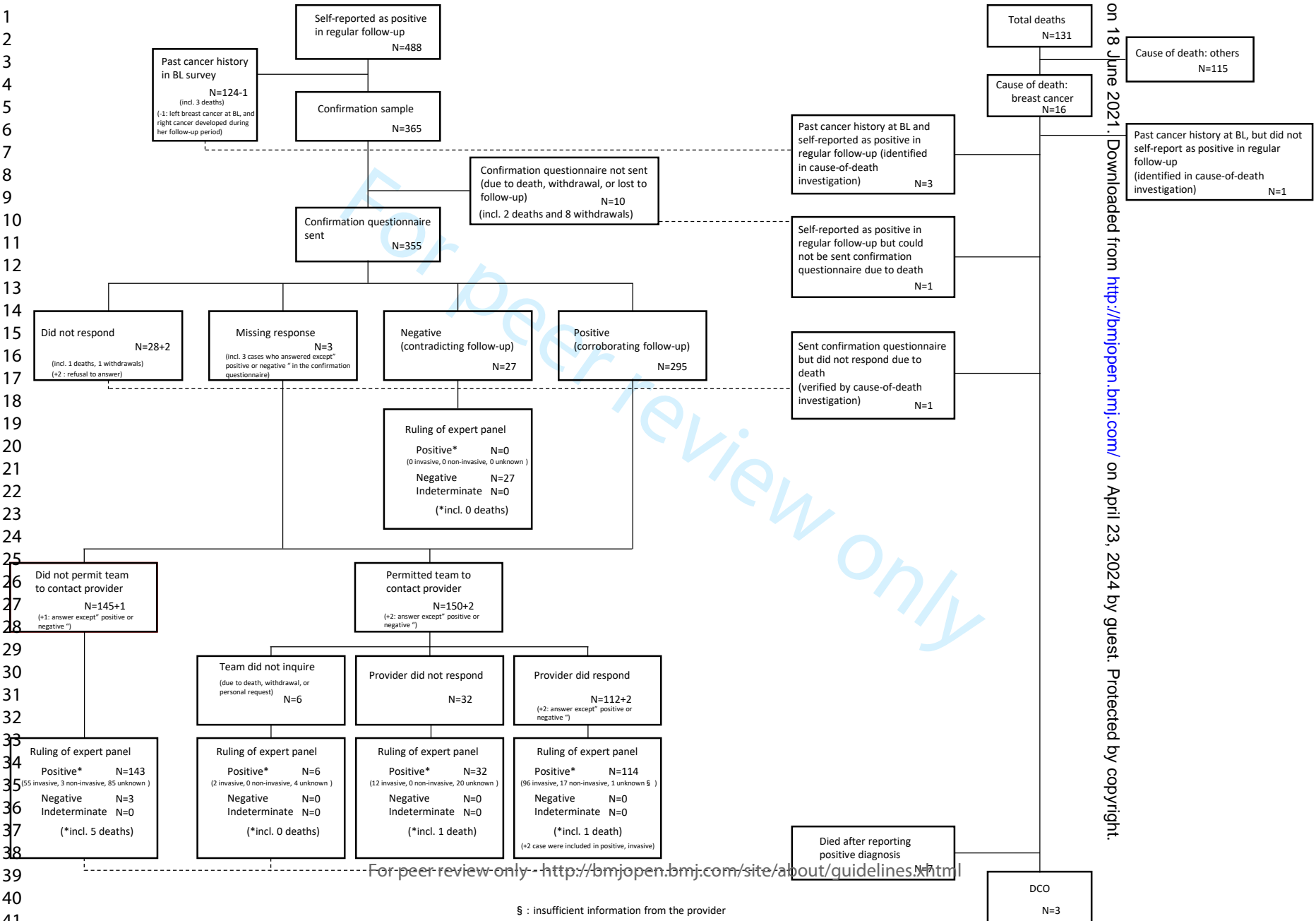
**Appendix Figure 3** Flow diagram illustrating the validation process for self-reported ovarian cancer (BL: baseline, DCO: death certificate only)



20-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.



Appendix Figure 4 Flow diagram illustrating the validation process for self-reported breast cancer (BL: baseline, CIS: carcinoma in situ, DCO: death certificate only)



20-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.html>

§ : insufficient information from the provider

# BMJ Open

## Validity of self-reported diagnoses of gynecological and breast cancers in a prospective cohort study: the Japan Nurses' Health Study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-045491.R1
Article Type:	Original research
Date Submitted by the Author:	04-Mar-2021
Complete List of Authors:	Takamatsu, Kiyoshi; Tokyo Dental College Ichikawa General Hospital, Obstetrics and Gynecology Ideno, Yuki; Gunma University, Center for Mathematics and Data Science Kikuchi, Mami; Gunma University Hospital, Center of Regional Medical Research and Education Yasui, Toshiyuki; University of Tokushima Graduate School of Biomedical Sciences, Department of Reproductive and Menopausal Medicine Maruoka, Naho; Gunma University Graduate School of Health Sciences, Department of International/Community Health Laboratory Sciences Nagai, Kazue; Gunma University Graduate School of Health Sciences, Gunma University, Department of International/Community Health Laboratory Sciences Hayashi, Kunihiko; Gunma University Graduate School of Health Sciences, Department of International/Community Health Laboratory Sciences
<b>Primary Subject Heading</b>:	Epidemiology
Secondary Subject Heading:	Obstetrics and gynaecology
Keywords:	EPIDEMIOLOGY, Breast tumours < ONCOLOGY, Gynaecological oncology < ONCOLOGY

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4  
5  
6 **Validity of self-reported diagnoses of gynecological and breast cancers in a**  
7 **prospective cohort study: the Japan Nurses' Health Study**  
8  
9  
10

11  
12  
13 Kiyoshi Takamatsu, MD, PhD, <sup>1</sup> Yuki Ideno, PhD, <sup>2</sup> Mami Kikuchi, MD, <sup>3</sup> Toshiyuki  
14 Yasui, MD, PhD, <sup>4</sup> Naho Maruoka, BS, <sup>5</sup> Kazue Nagai, PhD, <sup>5</sup> Kunihiko Hayashi, PhD <sup>5</sup>  
15  
16

17  
18  
19 **Author affiliations**

20  
21 <sup>1</sup>Department of Obstetrics and Gynecology, Tokyo Dental College Ichikawa General  
22 Hospital, Chiba, Japan

23  
24 <sup>2</sup>Gunma University Center for Mathematics and Data Science, Gunma, Japan

25  
26 <sup>3</sup>Center of Regional Medical Research and Education, Gunma University Hospital,  
27 Maebashi, Gunma, Japan

28  
29 <sup>4</sup>Department of Reproductive and Menopausal Medicine, Institute of Biomedical Science,  
30 Tokushima University Graduate School, Tokushima, Japan

31  
32 <sup>5</sup>Department of International/Community Health Laboratory Sciences, Graduate School  
33 of Health Sciences, Gunma University, Gunma, 371-8514, Japan  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51

52 Correspondence to Dr Kiyoshi Takamatsu, Department of Obstetrics and Gynecology,  
53 Tokyo Dental College Ichikawa General Hospital, Sugano 5-11-13, Ichikawa-city, Chiba  
54 272-0151, JAPAN; ktakamatsu@tdc.ac.jp  
55  
56  
57  
58  
59  
60

## ABSTRACT

**Objectives** To validate the self-reported diagnoses of gynecological and breast cancers in a nationwide prospective cohort study of nursing professionals: the Japan Nurses' Health Study (JNHS)

**Design and setting** Retrospective analysis of the Japan Nurses' Health Study (JNHS)

**Participants and measures** Data were reviewed for 15,717 subjects. The mean age at baseline was  $41.6 \pm 8.3$  years (median: 41), and the mean follow-up period was  $10.5 \pm 3.8$  years (median: 12). Participants are regularly mailed a follow-up questionnaire once every 2 years. Respondents who self-reported a positive cancer diagnosis were sent an additional confirmation questionnaire and contacted the diagnosing facility to confirm the diagnosis based on medical records. A review panel of experts verified the disease status. Regular follow-up, confirmation questionnaires, and expert review were validated for their positive predictive value (PPV) and negative predictive value (NPV).

**Results** New incidences were verified in 37, 47, 26, and 300 cervical, endometrial, ovarian, and breast cancer cases, respectively. The estimated incidence rates were 22.0, 25.4, 13.8 and 160.4 per 100,000 person-years. These were comparable to those of national data from regional cancer registries in Japan. For regular follow-up, the corresponding PPVs for cervical, endometrial, ovarian, and breast cancer were 16.9%,

1  
2  
3  
4  
5  
6 54.2%, 45.1%, and 81.4%, and the NPVs were 100%, 99.9%, 99.9%, and 99.9%,  
7  
8  
9 respectively. Adding the confirmation questionnaire improved the PPVs to 31.5%, 88.9%,  
10  
11  
12 76.7%, and 99.9%; the NPVs were uniformly 99.9%. Expert review yielded PPVs and  
13  
14  
15 NPVs that were all ~100%.

16  
17  
18 **Conclusions** Gynecological cancer cannot be accurately assessed by self-reporting alone.  
19  
20  
21 Additionally, the external validity of cancer incidence in this cohort was confirmed.  
22  
23

### 24 25 26 27 **Strengths and limitations of this study**

28  
29  
30 ▶ This study investigated the validity of self-reporting of gynecologic and breast cancers  
31  
32  
33 in a large, nationwide prospective cohort study of nursing professionals, the Japan Nurses'  
34  
35  
36 Health Study (JNHS).

37  
38  
39 ▶ Participants of JNHS cohort, which was composed entirely of female nursing  
40  
41  
42 professionals, are likely to answer the cancer history more accurately than general  
43  
44  
45 population.

46  
47  
48 ▶ Periodic questionnaires, meticulous review of subjects' medical records and death-  
49  
50  
51 certificate surveys were employed to establish self-report validity, circumventing the  
52  
53  
54 limitations presented by Japan's lack of complete national cancer registries.

55  
56  
57 ▶ Not all answers for confirmation questionnaire were obtained.  
58  
59  
60

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- ▶ There was relatively small number of young participants in this cohort.

## INTRODUCTION

Self-reporting is frequently used to assess disease status in cohort research. The methodology's cost-effectiveness and feasibility make it an attractive approach in countries without comprehensive national disease registries such as Japan. However, the unreliability of self-reported information is problematic and can introduce errors into epidemiological investigations of risk factors, especially for new cancer incidences in a cohort. Self-reporting appears to accurately reflect diabetes status and surgical history of hysterectomies<sup>1,2</sup>; however, body weight is often under-reported<sup>3</sup>. Regarding patients' cancer history, healthcare providers must consider that an affirmative response on a questionnaire is not equivalent to a definitive medical diagnosis because patients may remember incorrectly. Ideally, their answers should be corroborated against their medical records, but these typically cannot be acquired for an entire cohort. Additionally, validity can depend on background factors, such as ethnicity and cohort-specific characteristics, which further complicates interpreting self-report data. In this sense, validation of self-reported diagnoses of gynecological and breast cancers is not clear in Japan.

The Japan Nurses' Health Study (JNHS) is a nationwide prospective cohort study

1  
2  
3  
4  
5  
6 of over 15,000 female nurses, which began in 2001 to ascertain how women's health is  
7  
8  
9 affected by lifestyle factors, healthcare practices, and physical status over their lifetime <sup>4</sup>.

10  
11  
12 Here, we investigated the validity of self-reported diagnoses of three gynecological  
13  
14  
15 cancers (i.e., cervical, endometrial, and ovarian) and breast cancer in our cohort. Also, we  
16  
17  
18 checked the external validity of our cohort by confirming the cancer incidence.  
19  
20  
21  
22  
23

## 24 **METHODS**

### 25 **Subjects**

26  
27  
28  
29  
30 The JNHS is an ongoing prospective cohort study investigating the association between  
31  
32  
33 lifestyle, health care practices and women's health in Japan. Detailed information on its  
34  
35  
36 design, population, protocol, and sample-size calculations were published previously <sup>4,5</sup>.  
37  
38  
39 Briefly, the baseline survey was conducted from 2001–2007, with planned follow-up for  
40  
41  
42 30 years. In total, 15,019 women agreed to follow-up, signing and returning the informed-  
43  
44  
45 consent form with the completed survey. At the time of the baseline survey, the study  
46  
47  
48 population consisted of female licensed nursing professionals, including registered nurses,  
49  
50  
51 licensed practical nurses, public health nurses, and midwives, aged  $\geq 25$  years, and  
52  
53  
54 residing in Japan. Follow-up is currently ongoing; subjects are regularly mailed a self-  
55  
56  
57 administered questionnaire once every 2 years to complete and return by post.  
58  
59  
60



1  
2  
3  
4  
5  
6  
7 Before initiating the JNHS, the feasibility of its research strategy and the validity  
8  
9 of its questionnaires were investigated and confirmed in a pilot cohort study started in  
10  
11 1999 (the Gunma Nurses' Health Study; GNHS, n=698)<sup>6,7</sup>. We combined the JNHS and  
12  
13 GNHS datasets in the present work as JNHS cohort (n=15,717). Table 1 shows the  
14  
15 number of subjects in each age group. Women had a mean age at baseline of 41.6 (8.3)  
16  
17 years (mean (SD) ; median: 41 years) and a mean follow-up of 10.5 (3.8) years (median:  
18  
19 12 years).  
20  
21  
22  
23  
24  
25

26  
27 The JNHS Coordination and Data Center is located in the Epidemiological  
28  
29 Research Office of the School of Health Sciences at Gunma University. This study was  
30  
31 performed under the Declaration of Helsinki, the Guidelines for Good Epidemiology  
32  
33 Practices<sup>8</sup>, and the Japanese Ethical Guidelines for Epidemiological Research<sup>9</sup>. The  
34  
35 GNHS study protocol was approved by the institutional review board of Gunma  
36  
37 University, Japan (approval # 3, 1999), and the JNHS study protocol was approved by  
38  
39 the institutional review board of Gunma University, Japan (approval #101, 2001) and the  
40  
41 ethics review board of Japan's National Institute of Public Health, Japan (approval #  
42  
43 03007, 2003).  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56

## 57 **Patient and Public Involvement statement**

58  
59  
60

1  
2  
3  
4  
5  
6 This research was done without involving participants in defining the research question,  
7  
8  
9 outcome measures or study design. Participants were recruited with the study  
10  
11  
12 information to nursing society. They were not invited to comment on the design and to  
13  
14  
15 interpret the results, and were not invited to contribute to the writing or editing of the  
16  
17  
18 manuscript. The results will be reported to participants in the JNHS newsletter, and also  
19  
20  
21 be posted on the Website of JNHS.  
22  
23  
24  
25  
26

### 27 **Data collection and corroboration**

28  
29  
30 In the baseline and regular biennial follow-up questionnaires, women were asked, “Have  
31  
32  
33 you ever been diagnosed with breast cancer (cervical cancer, endometrial cancer, or  
34  
35  
36 ovarian cancer) by a medical doctor?”, and if so, what was their age at first diagnosis. We  
37  
38  
39 identified and isolated those women who self-reported new incidences of one of the  
40  
41  
42 cancers of interest in the regular follow-up by July 2017.  
43  
44

45 To corroborate the self-reported positive cases, an additional confirmation  
46  
47  
48 questionnaire was sent to those women who affirmed a new cancer diagnosis in the  
49  
50  
51 regular follow-up. Subjects were again asked the same question as above and to provide  
52  
53  
54 details about their date of/age at diagnosis, method of detection, tumor stage, and  
55  
56  
57 treatment history. We also asked for permission to access their medical records; if they  
58  
59  
60

1  
2  
3  
4  
5  
6 consented, we reviewed the records to obtain accurate clinical information on their  
7  
8  
9 condition. For gynecological cancers, the data collected included date of diagnosis,  
10  
11  
12 clinical stage, histological type, treatments, and concomitant cancer(s). For breast cancer,  
13  
14  
15 the data included date of diagnosis, tumor site, invasivity, Tumor-Node-Metastasis  
16  
17  
18 classification (Union for International Cancer Control, 7<sup>th</sup> ed.)<sup>10</sup>, diagnostic method(s),  
19  
20  
21 tumor size, mammography category, surgical procedure, histological classification, and  
22  
23  
24 pathological classification (i.e., regional lymph node involvement (pN) and hormone  
25  
26  
27 receptor positivity for estrogen receptor (ER), progesterone receptor (PgR), and human  
28  
29  
30 epidermal growth factor receptor type2 (HER2)). This clinical information was  
31  
32  
33 furnished to an expert review panel comprising specialists on gynecological and breast  
34  
35  
36 cancers to verify each self-reported positive diagnosis.  
37  
38

39  
40 In Japan, the clinical reporting of gynecologic cancers follows the Japan Society of  
41  
42 Obstetrics and Gynecology (JSOG) staging system, which is based on the internationally  
43  
44  
45 recognized surgical staging system published by the International Federation of  
46  
47  
48 Gynecology and Obstetrics (FIGO). When the FIGO criteria were updated during the  
49  
50  
51 study period in 2011<sup>11</sup>, the JSOG system was revised in tandem to remove Stage 0 lesions  
52  
53  
54 from the corresponding definitions, i.e., cervical carcinoma *in situ* (CIS) and atypical  
55  
56  
57 endometrial hyperplasia from cervical and endometrial cancer, respectively. Therefore,  
58  
59  
60

1  
2  
3  
4  
5  
6 stage 0 cancers were not considered positive in our primary analysis, and all medical  
7  
8 records were double-checked for patients who self-reported a new incidence of  
9  
10 gynecological cancer before 2011. These borderline cases were excluded.  
11  
12  
13  
14

15 If a subject was reported as deceased or inexplicably failed to complete any recent  
16  
17 study activities, we established a cause of death by checking it against death certificate-  
18  
19 related information in Japan's National Vital Statistics database.  
20  
21  
22  
23  
24  
25  
26  
27

## 28 **Validation**

29  
30 Regular follow-up, confirmation questionnaires, and expert review were validated for  
31  
32 their positive predictive value (PPV) and negative predictive value (NPV) for new  
33  
34 incidences of each cancer.  
35  
36  
37  
38

39 For the first two sources, the validation sample included all members of the study  
40  
41 cohort (n=15,717) who reported no past history of the cancer in question at baseline. The  
42  
43 PPV of the regular follow-up was calculated as the number of verified positive cases of  
44  
45 the cancer, i.e., cases whose self-reported positive diagnosis was verified by medical-  
46  
47 record review or cause-of-death investigation, divided by all cases of self-reported new  
48  
49 incidences of the cancer in the regular follow-up. The NPV was calculated as the number  
50  
51 of suspected negative cases, divided by all members of the validation sample who self-  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 reported no new cancer incidence in the regular follow-up. Here, the suspected negative  
7  
8  
9 cases consisted of all members of the validation sample for the cancer in question minus  
10  
11  
12 A) cases who self-reported new incidences in the regular follow-up and B) positive cases  
13  
14  
15 whose status was established only by death certificate (DCO).  
16  
17

18  
19 The PPV of the combined regular follow-up and confirmation questionnaire was  
20  
21 calculated as the number of verified positive cases of the cancer divided by all cases who  
22  
23 corroborated their positive diagnosis on the confirmation questionnaire. The NPV was  
24  
25 calculated as the number of suspected negative cases divided by all members of the  
26  
27 validation sample except those who self-reported their positive diagnosis on the  
28  
29 confirmation questionnaire. Here, the suspected negative cases consisted of all members  
30  
31 of the validation sample minus A) cases who self-reported their positive diagnosis on the  
32  
33 confirmation questionnaire, B) cases ruled positive by DCO, C) cases ruled positive by  
34  
35 cause-of-death investigation, and D) contradictory cases (i.e., women confirmed by  
36  
37 expert review but self-reported a negative status on the confirmation questionnaire or left  
38  
39 the field blank).  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50  
51 The expert review panel's judgments were also validated for comparison. In this  
52  
53 analysis, the validation sample consisted of all participants who A) returned the  
54  
55 confirmation questionnaire, B) permitted the research team to contact their diagnosing  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 facility, and C) their provider agreed to respond to the team's inquiry. The PPV was  
7  
8  
9 calculated as the number of cases verified as positive by the diagnosing facility, divided  
10  
11  
12 by the number of cases ruled positive by the expert review panel. The NPV was calculated  
13  
14  
15 as the number of cases verified as negative by the diagnosing facility, divided by the  
16  
17  
18 number of cases ruled negative by the panel.  
19

20  
21 After fixing the cancer cases, the incidence rate of each cancer was estimated  
22  
23 from the observed events and person-time at risk for 10 years of observation. Because  
24  
25 of the numbers of participants aged age <30 and  $\geq 60$  years were small, the 30–60  
26  
27 year-old age group was used. We calculated the 95% confidence intervals of the  
28  
29 incidence rates based on the exact Poisson confidence interval in accordance with  
30  
31 known methods <sup>12</sup>.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

## 43 RESULTS

### 44 45 **Verified cases of each cancer type**

46  
47  
48 The flow diagram illustrating the validation process of each cancer is listed in the Web  
49  
50 Appendices (Appendix 1-4). The numbers of new cases of self-reported cancers in the  
51  
52 regular follow-up (and incidences in the respective validation sample) were cervical  
53  
54 cancer: 219 (1.4%), endometrial cancer: 83 (0.5%), ovarian cancer: 51 (0.3%), and breast  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 cancer: 365 (2.3%). New incidence was verified by expert review in 37, 45, 23, and 297  
7  
8  
9 of these cases, respectively. Some subjects sent the confirmation questionnaire  
10  
11 corroborating their positive diagnosis but were ruled negative by the expert panel (72.1%,  
12  
13 11.1%, 30.3%, and 1.0%, respectively), while 37.6%, 33.8%, 25.6%, and 8.3% of subjects,  
14  
15  
16  
17  
18 respectively, responded with negative diagnosis on the confirmation questionnaire.  
19  
20

21 For all observed cases of mortality, cause of death was established as being cervical  
22  
23 cancer (n=4, DCO=0), endometrial cancer (n=7, DCO=2), ovarian cancer (n=3, DCO=3),  
24  
25 or breast cancer (n=16, DCO=3). New incidences of the four cancers were verified in 37,  
26  
27  
28  
29  
30  
31  
32 47, 26, and 300 cases, respectively.

33 In the JNHS cohort, the estimated incidence rates for patients aged 30–60 years  
34  
35 were 22.0/100,000; 25.4/100,000; 13.8/100,000; and 160.4/100,000 person-years for  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60 lack of heterogeneity between this cohort and Japanese women overall, the incidence rates  
for each age group were compared with the national data from regional cancer registries  
in the 2015 statistics published by Japan's National Cancer Center<sup>13</sup> (Figure 1-4). For all  
four cancers, the cohort data did not deviate from the national data.

### Self-reported PPV/NPV for each cancer

1  
2  
3  
4  
5  
6 Table 3 summarizes the PPVs and NPVs for the regular follow-up, regular follow-up plus  
7  
8 confirmation questionnaire, and expert review for the new incidence of each cancer.  
9

10  
11  
12 Expert review achieved 100% accuracy for each cancer except cervical (PPV:  
13  
14 92.3%) because of a single false-positive case, which the participant's provider clarified  
15  
16 to be a different condition.  
17  
18  
19

20  
21 Self-reporting achieved NPVs near 100% for all cancers for both the regular  
22  
23 follow-up and the regular follow-up plus confirmation questionnaire. However, the  
24  
25 corresponding PPVs tended to be somewhat lower and variable across cancers. The PPVs  
26  
27 were worse for gynecological cancers than for breast cancer (breast > endometrial >  
28  
29 ovarian > cervical, in descending order) for both follow-up sources. The PPV for uterine  
30  
31 cancer, which included cervical and endometrial cancers, was 27.2%.  
32  
33  
34  
35  
36  
37  
38

39 The regular follow-up plus confirmation questionnaire achieved higher PPVs in all  
40  
41 cases than did regular follow-up alone; however, while it achieved 99.0% accuracy for  
42  
43 breast cancer, the estimates were lower for endometrial (88.9%) and ovarian (76.7%)  
44  
45 cancer and poor for cervical cancer (31.5%).  
46  
47  
48  
49

50  
51 Considering the changes to the official JSOG clinical staging system during the  
52  
53 survey period, we calculated a similar summary for PPVs and NPVs, adding cases of  
54  
55 cervical CIS, atypical endometrial hyperplasia, and borderline ovarian tumors (Table 4).  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6 The resulting PPVs were uniformly higher when all three cancers were included  
7  
8  
9 than when they were excluded. For endometrial and ovarian cancer, the improvements  
10  
11 ranged from 3.3%–6.7%, but for cervical cancer, their inclusion almost doubled the  
12  
13 predictive value for both the regular follow-up and regular follow-up plus confirmation  
14  
15  
16  
17  
18  
19 questionnaire, at +20.1% and +37.9%, respectively.  
20  
21  
22  
23  
24  
25  
26

## 27 **DISCUSSION**

28  
29  
30 In the JNHS cohort, self-reporting in regular follow-up achieved a PPV of 81.4% for  
31  
32 breast cancer but performed poorer for gynecological cancers, especially uterine cancers  
33  
34 (PPV: 27.2%) and cervical cancer alone (PPV: 16.9%). Our PPVs were higher than the  
35  
36  
37 corresponding values reported by the Japan Public Health Center (JPHC) Study, a  
38  
39  
40 population-based prospective cohort study (all cancers in women: 54.2%, breast: 58.4%,  
41  
42  
43 uterine: 21.7%)<sup>14</sup>. The validity of self-reporting is associated with individual  
44  
45  
46 characteristics<sup>15</sup>, and our cohort consisted entirely of nursing professionals. While  
47  
48  
49 evidence suggests that educational level has a negligible association with validity<sup>16</sup>, we  
50  
51  
52 partially attribute the high self-reporting accuracy to the uniformly high level of medical  
53  
54  
55 education and deeper knowledge of cancer in our cohort than in the general population.  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 Other studies support this argument <sup>17</sup>. However, sizeable percentages of nurses who  
7  
8 affirmed new incidences of cancer in the regular follow-up gave the opposite response on  
9  
10 the confirmation questionnaire (gynecological cancer: 25.6%–37.6%, breast cancer:  
11  
12 8.3%). Similarly, considerable percentages of respondents to the confirmation  
13  
14 questionnaire were verified not to have cancer (gynecological cancer: 41.2%–81.2%,  
15  
16 breast cancer: 9.2%). Many who corroborated their self-reported positive diagnosis were  
17  
18 eventually ruled negative by expert review, especially for cervical cancer (72.1%),  
19  
20 followed by ovarian (30.3%), endometrial (11.1%) and breast (1.0%). In summary, self-  
21  
22 reporting alone apparently fails to capture the real cancer incidence, even for this cohort  
23  
24 of nursing professionals with uniformly high medical knowledge. Additional inquiries to  
25  
26 confirm the details are needed.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

39 Compared to PPVs of self-report validity in other prospective cohort study datasets  
40  
41 <sup>16, 18, 19</sup>, our PPVs were comparable to the literature values for breast cancer but lower  
42  
43 than these values for uterine cancers. Many studies have shown that self-reporting of  
44  
45 breast cancer has high PPVs <sup>10, 16, 19</sup>. Some evidence has linked higher educational levels  
46  
47 with a greater risk of breast cancer <sup>20</sup>, which may also be true for our cohort. Additionally,  
48  
49 breast cancer diagnoses included ductal carcinoma *in situ*, which may have led to less  
50  
51 confusion than with gynecological cancers that excluded stage 0 cases and borderline  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 tumors.  
7  
8

9  
10 Studies outside of Japan have also found self-reporting to yield lower PPVs for  
11  
12 uterine cancers than for other cancers<sup>18,21</sup>, for several possible reasons. One is inaccurate  
13  
14 memory of precancerous cervical lesions, which are rarely addressed immediately by  
15  
16 surgical intervention. Additionally, age and sex may have some association; for example,  
17  
18 participants >50 years old in a Native American cohort were more likely to report  
19  
20 incorrectly<sup>22</sup>. Further, a study from Australia found that self-reported breast cancer had  
21  
22 lower PPVs in women aged 70–75 years<sup>23</sup>. Disease-specific considerations may also be  
23  
24 relevant. One study noted that many cases of women's cancers, especially cervical cancer,  
25  
26 are not recorded in cancer registries<sup>22</sup>, while another estimated false-negative rates of  
27  
28 43.8%, 28.6%, and 20.8% for self-reports of uterine, ovarian, and breast cancers,  
29  
30 respectively<sup>24</sup>. Differences in incidence must also be considered. Because gynecological  
31  
32 cancers are >5 times less prevalent than breast cancer, a difference of one case would  
33  
34 produce a proportionally larger change in PPV.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

48 One problem specific to Japan regarding the self-reporting of women's cancers is  
49  
50 how the results of cytological screening tests are reported for cervical and endometrial  
51  
52 cancers. Today, Pap smear results are recorded using the Bethesda system, the standard  
53  
54 international format, but these results previously followed a class-based system. Class II  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 status, which shows within the normal range is sometimes confused with stage II cervical  
7  
8  
9 cancer. Similarly, atypical endometrial hyperplasia was previously classified as stage 0  
10  
11  
12 endometrial cancer, which may be confused with non-atypical endometrial hyperplasia.  
13  
14

15 We suspect that another reason the self-report validity in our cohort was so poor  
16  
17  
18 for certain cancers was that subjects were recalling their past medical history during the  
19  
20  
21 regular follow-up, rather than the new incidence as intended. Additionally, ambiguous  
22  
23  
24 language in the questionnaire, such as “dysplasia” or “precancerous lesions”, may have  
25  
26  
27 reduced the self-report validity, as evidenced by the higher PPVs for borderline forms,  
28  
29  
30 such as cervical CIS, atypical endometrial hyperplasia, and borderline ovarian tumors  
31  
32  
33 included in the analysis. Among the three borderline forms, classifying cervical CIS as  
34  
35  
36 cervical cancer led to a greater increase in PPVs than did other cancers. Manjer et al. also  
37  
38  
39 found that self-reporting of malignant cervical cancer was less sensitive when the  
40  
41  
42 definition included cervical CIS<sup>25</sup>. These considerations suggest that compared with  
43  
44  
45 other cancers, diagnoses of cervical cancer and precancerous lesions have a greater risk  
46  
47  
48 of being inaccurately communicated or negatively interpreted by patients.  
49  
50

51 One of this study’s strengths was our meticulous review of subjects’ medical  
52  
53  
54 records and death-certificate surveys to establish self-report validity, circumventing the  
55  
56  
57 limitations presented by Japan’s lack of complete national cancer registries. Additionally,  
58  
59  
60

1  
2  
3  
4  
5  
6 we believe that our data better reflect the general Japanese population than did past  
7  
8  
9 findings for other regional cohorts because the nationwide scope of the JNHS minimizes  
10  
11  
12 the geographical variation. Moreover, our cohort was relatively homogenous in terms of  
13  
14  
15 sex and occupation, consisting entirely of female nursing professionals.  
16  
17

18  
19 The study also had some limitations. Cohort-specific characteristics may limit the  
20  
21 generalizability of our findings, especially the relatively young skew of the participants'  
22  
23  
24 ages. However, when converted to incidence rates, our rates seem most consistent with  
25  
26  
27 the 2015 statistics published by Japan's National Cancer Center <sup>13</sup>. Additionally, self-  
28  
29  
30 reported diagnoses could not be verified in some cases. Our expert panel made their  
31  
32  
33 judgments based on the specific language nurses used in the questionnaire to describe  
34  
35  
36 their treatments such as "hysterectomy" and "chemotherapy", but the panel still  
37  
38  
39 encountered cases that were difficult to definitively verify. However, we established a  
40  
41  
42 conclusive diagnosis based on all available information such as postmortem exam  
43  
44  
45 findings and supplemental details from primary-care providers. No indeterminate cases  
46  
47  
48 were found among those lacking medical records for verification.  
49  
50  
51

## 52 53 54 **CONCLUSION**

55  
56  
57 In Japan, gynecological cancer also cannot be accurately assessed by self-reporting  
58  
59  
60

1  
2  
3  
4  
5  
6 alone. However, external validity of these cancer incidence in JNHS with our method was  
7  
8  
9 confirmed. As the JNHS database covers all of Japan, this results allow the further  
10  
11  
12 investigation of risk factors for different cancers such as menopausal hormone therapy  
13  
14  
15 and lifestyle factors and their associations, with unaffected by information bias. We plan  
16  
17  
18 to continue our work by analyzing the respective contributions of different risk factors  
19  
20  
21 among confirmed cases of gynecological and breast cancer, as verified above.  
22  
23  
24  
25  
26

27 **Acknowledgments** The authors appreciate the cooperation of the Japanese nurses who  
28  
29  
30 participate in the JNHS and the GNHS. We also thank Ms. Satomi Shimizu at the JNHS  
31  
32  
33 Data Center for her help with data management.  
34  
35  
36  
37  
38  
39

40 **Contributors** KT was the panel member and wrote the initial draft of the paper to which  
41  
42  
43 all authors contributed. YI, NM, and KN collected data and analyzed. MK and TY were  
44  
45  
46 panel members and revised manuscript. KH designed the study, raised funding, and  
47  
48  
49 directed its implementation including quality assurance and control.  
50  
51  
52  
53

54 **Funding** This work was partly supported by a grant from the Japan Society for the  
55  
56  
57 Promotion of Science (JSPS KAKENHI Grant Number: 18H04069).  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 **Competing interests** None declared.  
11  
12  
13  
14

15 **Patient consent for publication** Not required.  
16  
17  
18  
19

20  
21 **Data sharing statement** The data are not publically available due to data transfer  
22  
23 agreements. No additional data is available.  
24  
25  
26  
27  
28  
29

30 **Provenance and peer review** Not commissioned; externally peer reviewed.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**REFERENCES**

- 1 Jackson JM, DeFor TA, Crain AL, et al. Validity of diabetes self-reports in the Women's Health Initiative. *Menopause* 2014;21(8):861-868.
- 2 Gentry-Maharaj A, Taylor H, Kalsi J, et al. Validity of self-reported hysterectomy: a prospective cohort study within the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *BMJ Open* 2014;4(3):e004421.
- 3 Skeie G, Mode N, Henningsen M, et al. Validity of self-reported body mass index among middle-aged participants in the Norwegian Women and Cancer study. *Clin Epidemiol* 2015;7:313-323.
- 4 Hayashi K, Mizunuma H, Fujita T, et al. Design of the Japan Nurses' Health Study: A prospective occupational cohort study of women's health in Japan. *Ind Health* 2007;45(5):679-686.
- 5 Fujita T, Hayashi K, Katanoda K, et al. Prevalence of diseases and statistical power of the Japan Nurses' Health Study. *Ind Health* 2007;45(5):687-694.
- 6 Maeno T, Ohta A, Hayashi K, et al. Impact of reproductive experience on women's smoking behaviour in Japanese nurses. *Public Health* 2005;119(9):816-824.



- 1  
2  
3  
4  
5  
6  
7 7 Kato C, Shimada J, Hayashi K. Sleepiness during shift work in Japanese nurses: A  
8  
9 comparison study using JESS, SSS, and actigraphy. *Sleep Biol Rhythm*  
10  
11  
12 2012;10(2):109-117.  
13  
14
- 15 8 Andrews EA, Avorn J, Bortnichak EA, et al. ; ISPE. Guidelines for Good  
16  
17 Epidemiology Practices for drug, device, and vaccine research in the United States.  
18  
19 Pharmacoepidemiol *Drug Saf* 1996;5(5):333-338.  
20  
21  
22
- 23 9 Ministry of Education, Culture, Sports, Science and Technology and Ministry of  
24  
25 Health, Labour and Welfare. Ethical Guidelines for Epidemiological Research.  
26  
27  
28 <https://www.niph.go.jp/wadai/ekigakurinri/guidelines.pdf>. Accessed August 24,  
29  
30  
31 2020.  
32  
33  
34
- 35 10 International Union Against Cancer. TNM Classification of Malignant Tumours 7<sup>th</sup>  
36  
37 ed. Eds. Sobin LH, Gospodarowicz MK and Wittekind Ch, Wiley, New York, 2009.  
38  
39  
40
- 41 11 FIGO committee on gynecologic oncology. Revised FIGO staging for carcinoma of  
42  
43 the vulva, cervix, and endometrium. *Int J Gynecol Obstet* 2009;105(2):103-104.  
44  
45  
46  
47
- 48 12 Ulm K. A simple method to calculate the confidence interval of a standardized  
49  
50 mortality ratio (SMR). *Am J Epidemiol* 1990;131(2):373-375.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- 13 Cancer Registry and Statistics. Cancer Information Service, National Cancer Center, Japan (Monitoring of Cancer Incidence in Japan (MCIJ)).  
[https://ganjoho.jp/reg\\_stat/statistics/stat/summary.html](https://ganjoho.jp/reg_stat/statistics/stat/summary.html). Accessed August 24, 2020.
- 14 Inoue M, Sawada N, Shimazu T, et al. Japan Public Health Center-based Prospective Study Group. Validity of self-reported cancer among a Japanese population: recent results from a population-based prospective study in Japan (JPHC Study). *Cancer Epidemiol* 2011;35(3):250-253.
- 15 Jackson CA, Mishra GD, Tooth L, et al. Moderate agreement between self-reported stroke and hospital-recorded stroke in two cohorts of Australian women: a validation study. *BMC Med Res Methodol* 2015;15:7.
- 16 Bergmann MM, Calle EE, Mervis CA, et al. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol* 1998;147(6):556-562.
- 17 Colditz GA, Manson JE, Hankinson SE. The Nurses' Health Study: 20-year contribution to the understanding of health among women. *J Womens Health* 1997;6(1):49-62.

- 1  
2  
3  
4  
5  
6  
7 18 Navarro C, Chirlaque MD, Tormo MJ, et al. Validity of self reported diagnoses of  
8  
9 cancer in a major Spanish prospective cohort study. *J Epidemiol Community Health*  
10  
11 2006;60(7):593-599.  
12  
13  
14  
15  
16 19 Loh V, Harding J, Koshkina V, et al. The validity of self-reported cancer in an  
17  
18 Australian population study. *Aust NZ J Public Health* 2014;38(1):35-38.  
19  
20  
21  
22 20 Goldberg M, Calderon-Margalit R, et al. Socioeconomic disparities in breast cancer  
23  
24 incidence and survival among parous women: findings from a population-based  
25  
26 cohort, 1964-2008. *BMC Cancer* 2015;15:921.  
27  
28  
29  
30  
31 21 Cho S, Shin A, Song D, et al. Validity of self-reported cancer history in the health  
32  
33 examinees (HEXA) study: A comparison of self-report and cancer registry records.  
34  
35  
36  
37 *Cancer Epidemiol* 2017;50(Pt A):16-21.  
38  
39  
40  
41 22 Nash SH, Day G, Hiratsuka VY, et al. Agreement between self-reported and central  
42  
43 cancer registry-recorded prevalence of cancer in the Alaska EARTH study. *Int J*  
44  
45  
46 *Circumpolar Health* 2019;78(1):1571383.  
47  
48  
49  
50 23 Stavrou E, Vajdic CM, Loxton D, et al. The validity of self-reported cancer  
51  
52 diagnoses and factors associated with accurate reporting in a cohort of older  
53  
54 Australian women. *Cancer Epidemiol* 2011;35(6):e75-80.  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4  
5  
6  
7 24 Desai MM, Bruce ML, Desai RA, et al. Validity of self-reported cancer history: a  
8  
9 comparison of health interview data and cancer registry records. *Am J Epidemiol*  
10  
11 2001;153(3):299-306.  
12  
13  
14  
15 25 Manjer J, Merlo J, Berglund G. Validity of self-reported information on cancer:  
16  
17 determinants of under- and over-reporting. *Eur J Epidemiol* 2004;19(3):239-247.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7 **Figure and table legends**  
8  
9

10  
11  
12 **Figure 1** Estimated incidence rates of cervical cancer for each age group in the JNHS  
13 cohort and the national data from regional cancer registries.  
14  
15

16  
17  
18 (error bars shows the 95% confidence intervals)  
19  
20

21  
22  
23  
24 **Figure 2** Estimated incidence rates of endometrial cancer for each age group in the  
25 JNHS cohort and the national data from regional cancer registries.  
26  
27

28  
29  
30 (error bars shows the 95% confidence intervals)  
31  
32

33  
34  
35  
36 **Figure 3** Estimated incidence rates of ovarian cancer for each age group in the JNHS  
37 cohort and the national data from regional cancer registries.  
38  
39

40  
41  
42 (error bars shows the 95% confidence intervals)  
43  
44

45  
46  
47  
48 **Figure 4** Estimated incidence rates of breast cancer for each age group in the JNHS  
49 cohort and the national data from regional cancer registries.  
50  
51

52  
53  
54 (error bars shows the 95% confidence intervals)  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6 **Table 1** Numbers and percentages of subjects in each age group at baseline in the JNHS  
7  
8  
9 cohort

10  
11  
12  
13  
14  
15 **Table 2** Estimated incidence rate of each cancer in patients aged 30–60 years in the  
16  
17  
18 JNHS cohort

19  
20  
21  
22  
23  
24 **Table 3** PPVs/NPVs for regular follow-up, regular follow-up plus confirmation  
25  
26  
27 questionnaire, and expert review for new incidences of gynecological and breast cancers  
28  
29  
30 in the JNHS cohort

31  
32  
33  
34  
35  
36 **Table 4** Corresponding PPVs/NPVs including those of cervical CIS, atypical endometrial  
37  
38  
39 hyperplasia, and borderline ovarian tumor in the JNHS cohort

40  
41  
42  
43  
44  
45 **Appendix legends**

46  
47  
48  
49  
50  
51 **Appendix Figure 1** Flow diagram illustrating the validation process for self-reported  
52  
53  
54 cervical cancer (BL: baseline, CIS: carcinoma in situ, DCO: death certificate only)

1  
2  
3  
4  
5  
6 **Appendix Figure 2** Flow diagram illustrating the validation process for self-reported  
7  
8  
9 endometrial cancer (BL: baseline, aEmH: atypical endometrial hyperplasia, DCO: death certificate  
10  
11  
12 only)  
13  
14  
15  
16  
17

18 **Appendix Figure 3** Flow diagram illustrating the validation process for self-reported  
19  
20  
21 ovarian cancer (BL: baseline, DCO: death certificate only)  
22  
23  
24  
25  
26  
27

28 **Appendix Figure 4** Flow diagram illustrating the validation process for self-reported  
29  
30  
31 breast cancer (BL: baseline, CIS: carcinoma in situ, DCO: death certificate only)  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 1 Numbers and percentages of subjects in each age group at baseline in the JNHS cohort

Age (years)	n	(%)
< 30	692	(4.4)
30 – 34	2,955	(18.8)
35 – 39	3,176	(20.2)
40 – 44	3,133	(19.9)
45 – 49	2,767	(17.6)
50 – 54	2,012	(12.8)
55 – 59	797	(5.1)
60 – 64	143	(0.9)
≥ 65	42	(0.3)



Table 2 Estimated incidence rate of each cancer in patients aged 30 to 60 years in the JNHS cohort

	Cancer cases	Person-years	Incidence rate (per 100,000 person- years)	Lower limit of 95% confidence interval	Upper limit of 95% confidence interval
Cervical cancer	29	131,658.50	22.0	14.8	31.6
Endometrial cancer	32	126041.0	25.4	17.4	35.8
Ovarian cancer	18	130662.5	13.8	8.2	21.8
Breast cancer	210	130960.5	160.4	139.4	183.6

Table 3 PPVs/NPVs of regular follow-up, regular follow-up plus confirmation questionnaire, and expert review for new incidences of gynecological and breast cancers in the JNHS cohort

		Validation sample	Positive diagnosis self-reported in regular follow-up	Positive status established by cause-of-death investigation (incl. DCO cases)	Regular follow-up		Regular follow-up plus confirmation questionnaire		Expert review	
					PPV	NPV	PPV	NPV	PPV	NPV
Cervical cancer	167	15,550	219	2 (0)	16.9%	100.0%	31.5%	92.9%	92.3%	100.0%
Endometrial cancer	31	15,686	83	7 (2)	54.2%	99.9%	88.9%	99.9%	100.0%	100.0%
Ovarian cancer	37	15,680	51	3 (3)	45.1%	99.9%	76.7%	99.9%	100.0%	100.0%

Breast cancer	138	15,579	365	5 (3)	81.4%	99.9%	99.0%	99.9%	100.0%	100.0%
---------------	-----	--------	-----	-------	-------	-------	-------	-------	--------	--------

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

Table 4 Corresponding PPVs/NPVs including cervical CIS, atypical endometrial hyperplasia, and borderline ovarian tumors in the JNHS cohort

	Positive history self-reported at baseline	Validation sample	Positive diagnosis self-reported in regular follow-up	Positive status established by cause-of-death investigation (incl. DCO cases)	Regular follow-up		Regular follow-up plus confirmation questionnaire	
					PPV	NPV	PPV	NPV
Cervical cancer	167	15,550	219	2 (0)	37.0%	100.0%	69.4%	99.9%
Endometrial cancer	31	15,686	83	7 (2)	57.8%	99.9%	95.6%	99.9%
Ovarian cancer	37	15,680	51	3 (3)	49.0%	99.9%	80.0%	99.9%

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

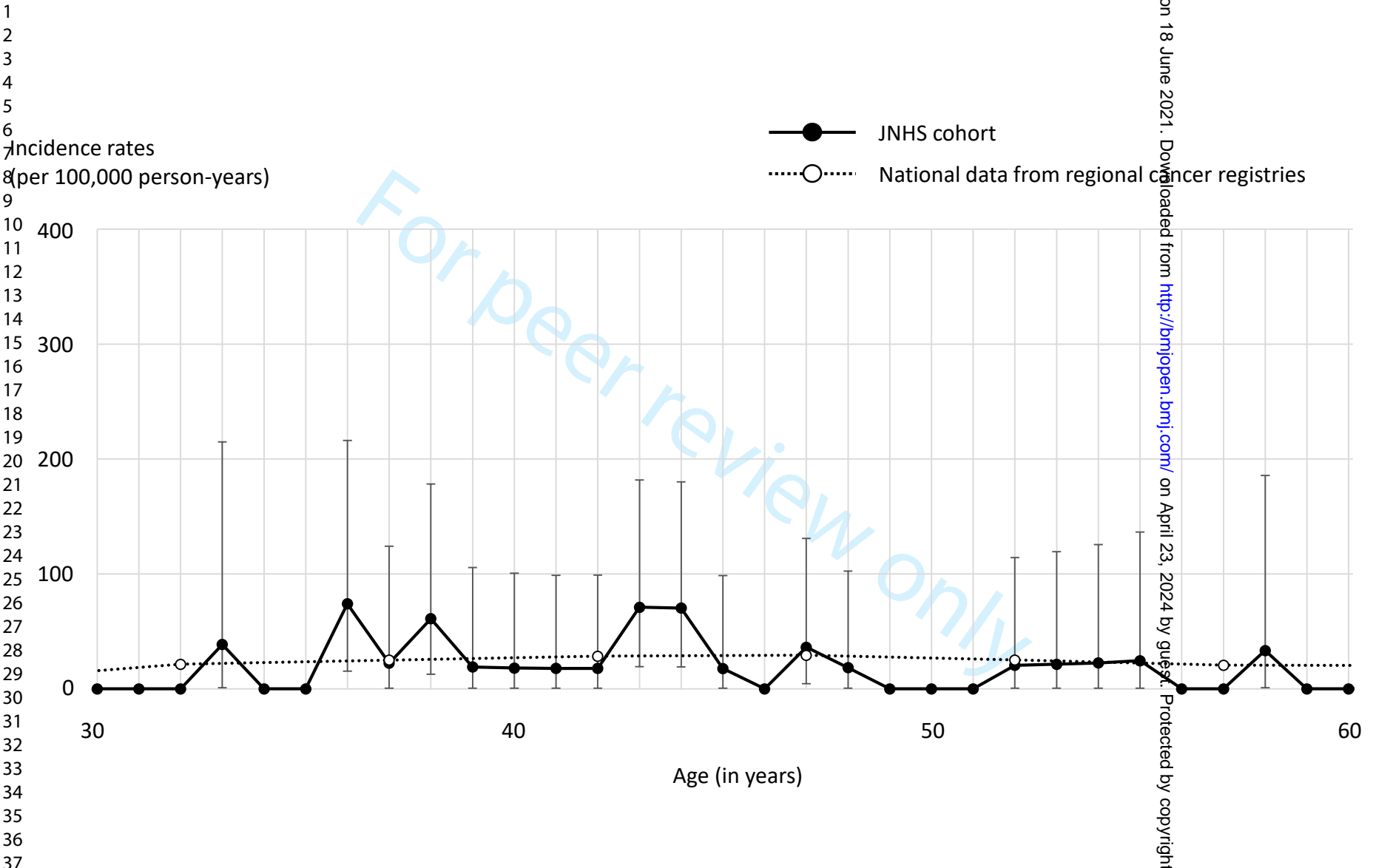
For peer review only

6/bmjopen-2020-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

**Figure 1** Estimated incidence rates of cervical cancer for each age group in the JNHS cohort and the national data from regional cancer registries (error bars shows the 95% confidence intervals)



20-00000-1 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

**Figure 2** Estimated incidence rates of endometrial cancer for each age group in the JNHS cohort and the national data from regional cancer registries

(error bars shows the 95% confidence intervals)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

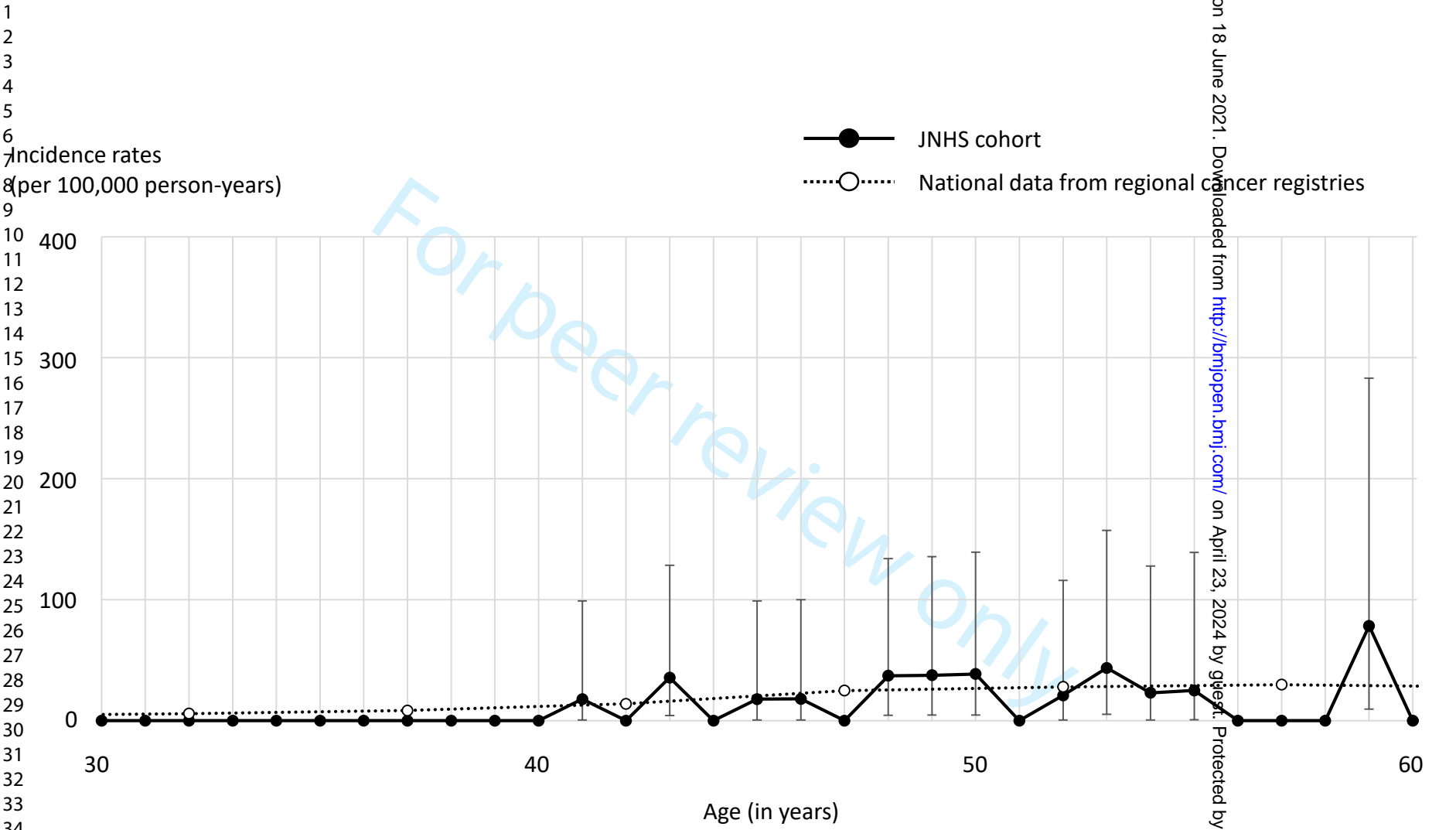
● JNHS cohort  
○ National data from regional cancer registries



20-000491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.



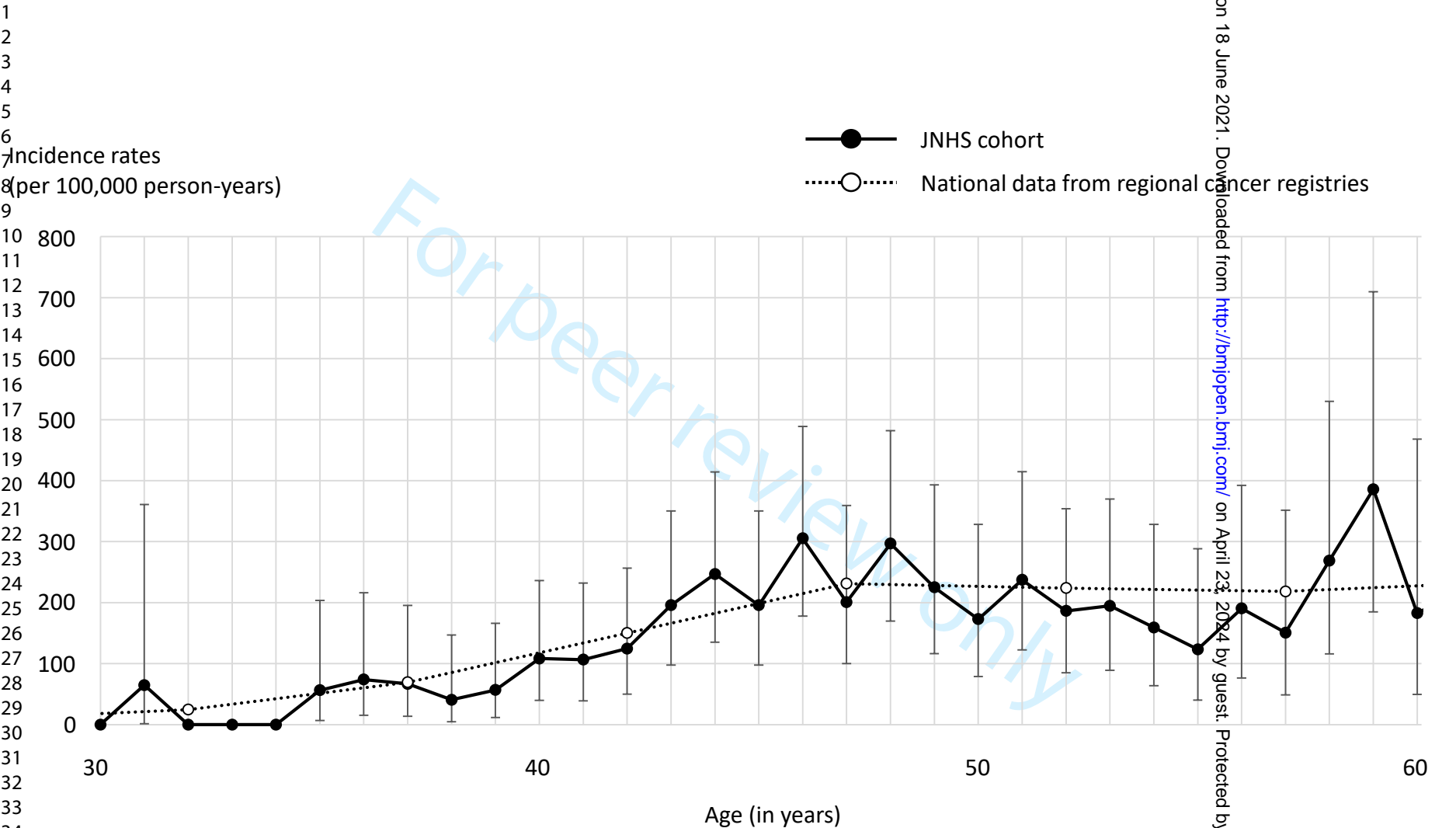
**Figure 3** Estimated incidence rates of ovarian cancer for each age group in the JNHS cohort and the national data from regional cancer registries (error bars shows the 95% confidence intervals)



For peer review only

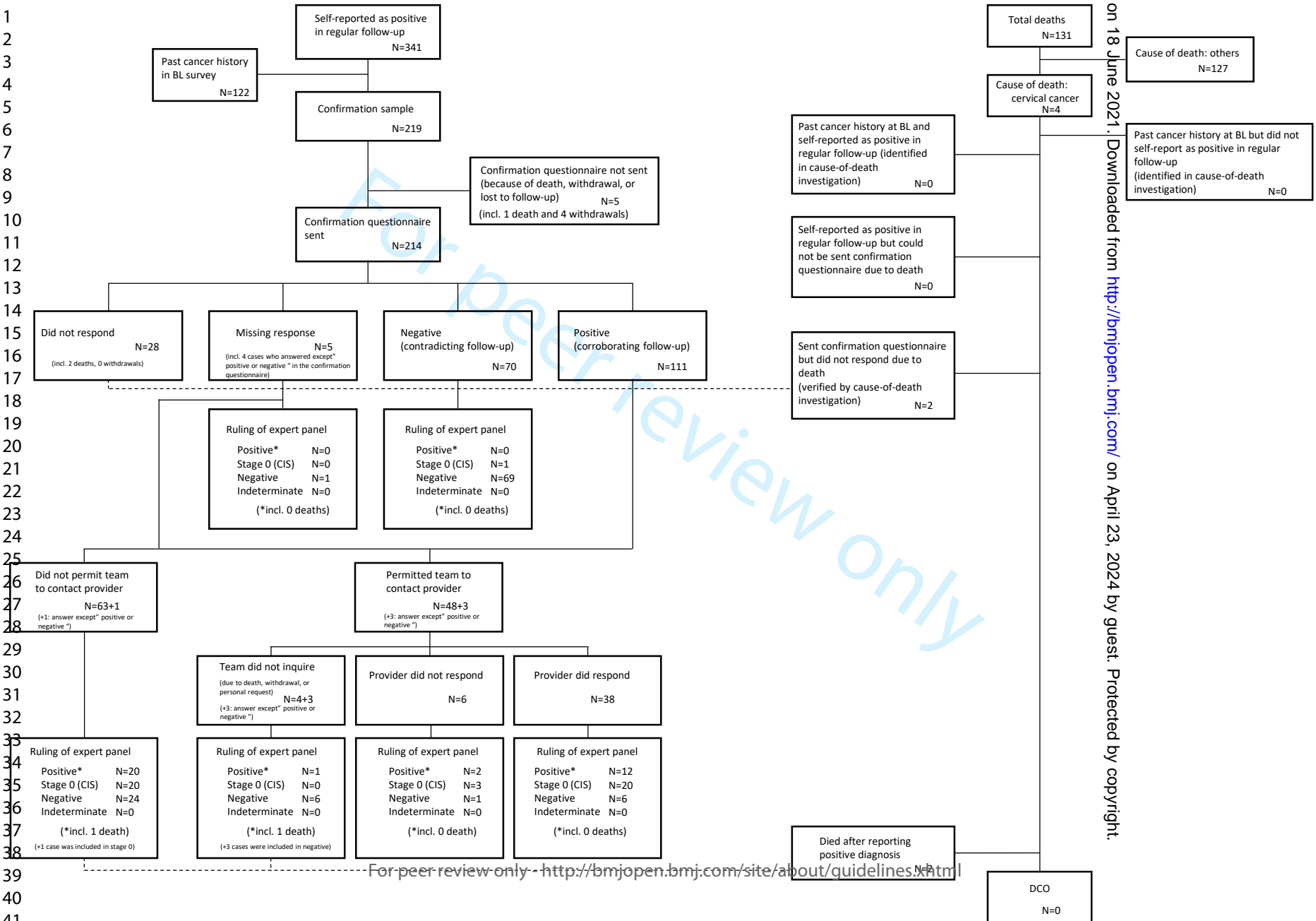
20-00000-191 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

**Figure 4** Estimated incidence rates of breast cancer for each age group in the JNHS cohort and the national data from regional cancer registries (error bars shows the 95% confidence intervals)



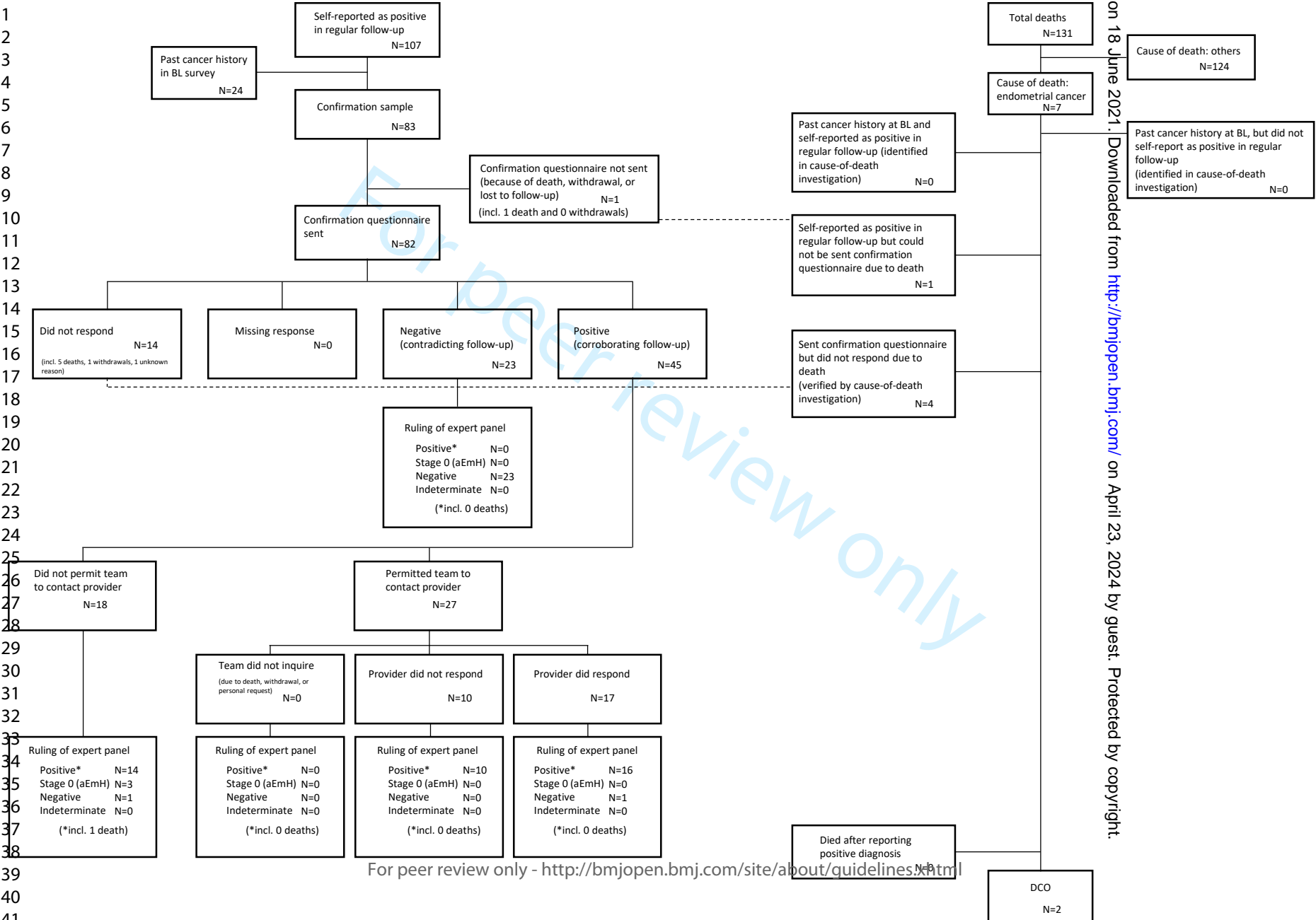
2020-06-19 11:51 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

Appendix Figure 1 Flow diagram illustrating the validation process for self-reported cervical cancer (BL: baseline, CIS: carcinoma in situ, DCO: death certificate only)



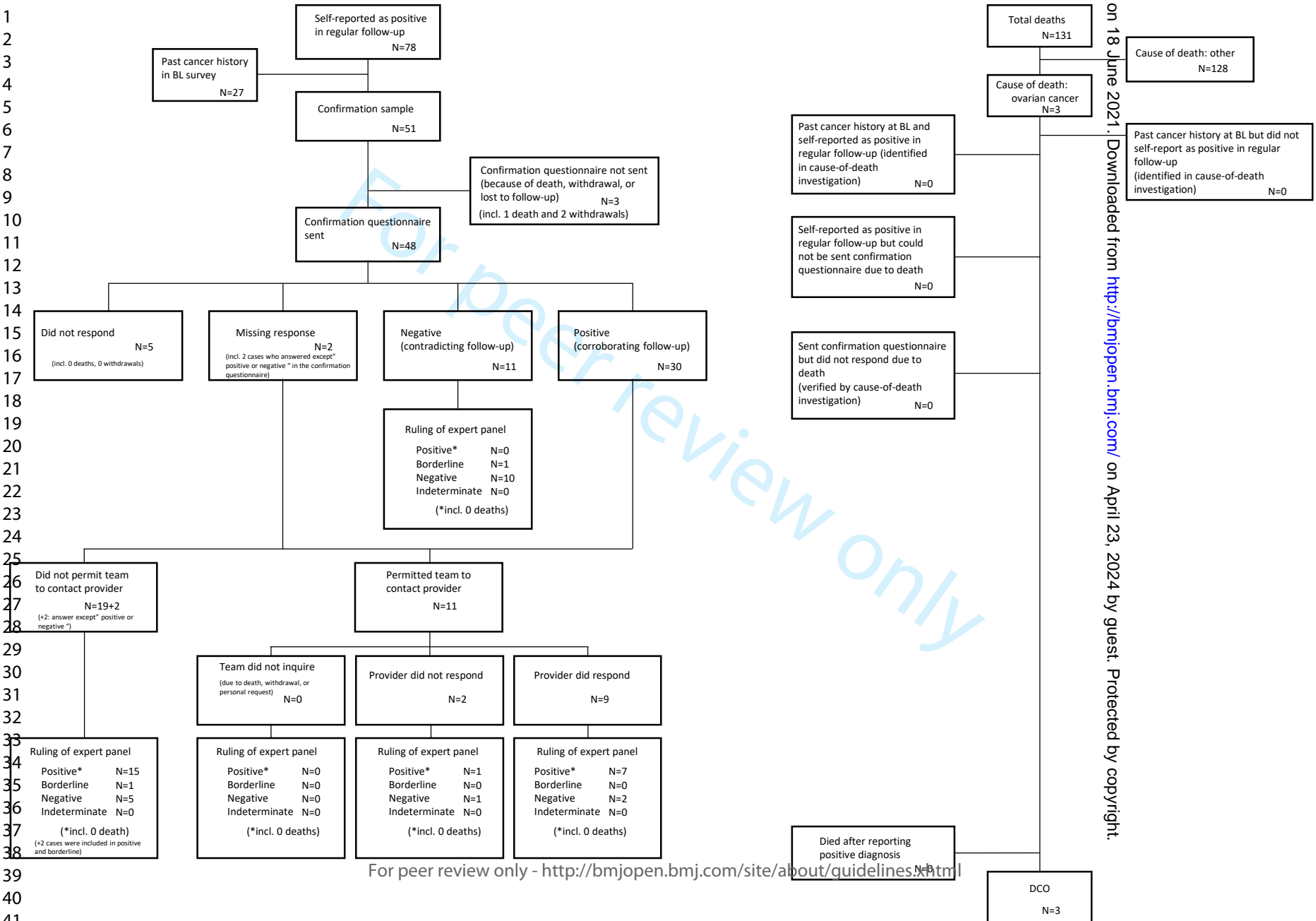
20-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

**Appendix Figure 2** Flow diagram illustrating the validation process for self-reported endometrial cancer (BL: baseline, aEmH: atypical endometrial hyperplasia, DCO: death certificate only)



20-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

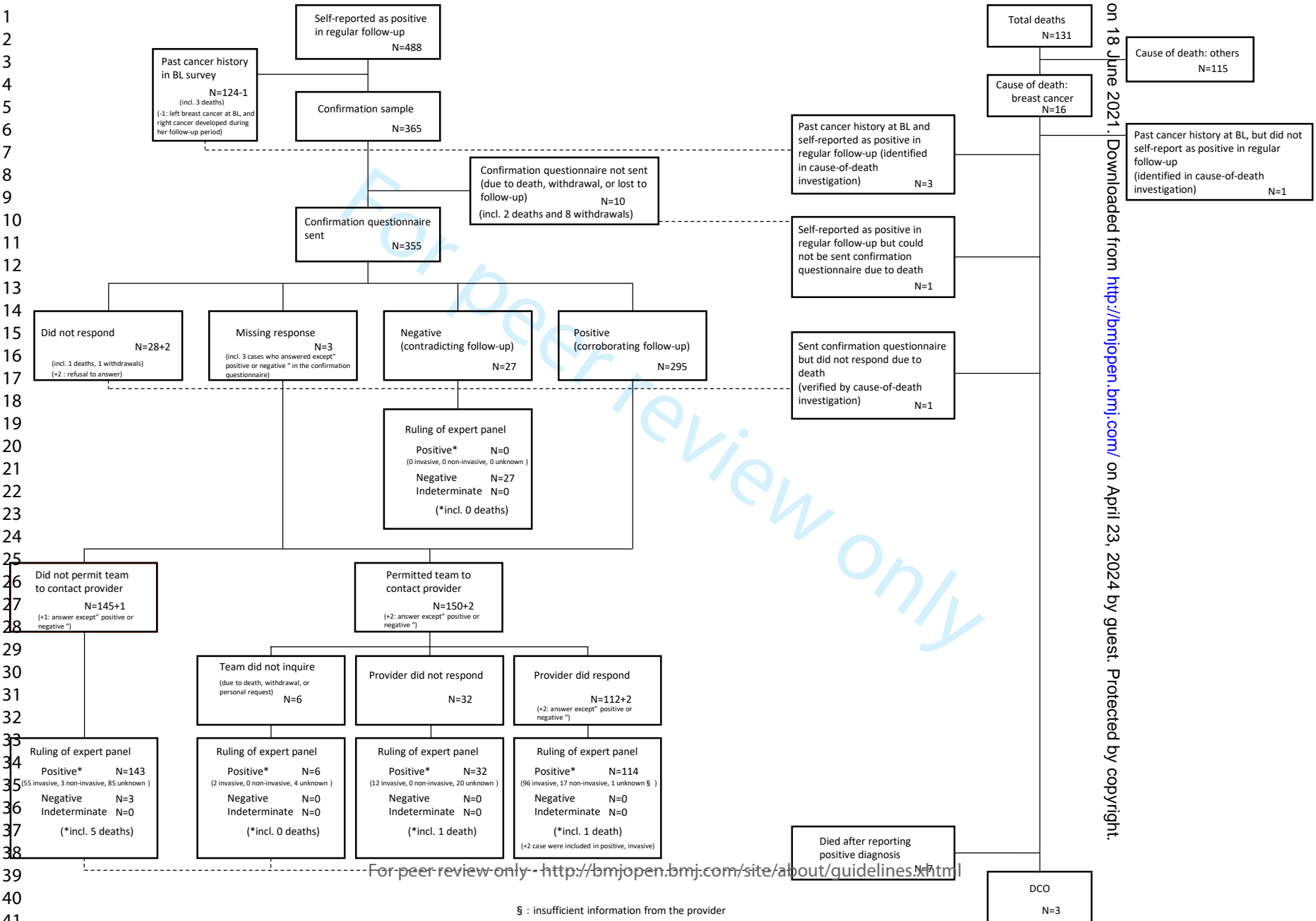
**Appendix Figure 3** Flow diagram illustrating the validation process for self-reported ovarian cancer (BL: baseline, DCO: death certificate only)



20-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.

**Appendix Figure 4** Flow diagram illustrating the validation process for self-reported breast cancer (BL: baseline, CIS: carcinoma in situ, DCO: death certificate only)

20-045491 on 18 June 2021. Downloaded from <http://bmjopen.bmj.com/> on April 23, 2024 by guest. Protected by copyright.



For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.html>

§ : insufficient information from the provider

**STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of cohort studies**

Section/Topic	Item #	Recommendation	Reported on page #
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5-6
		(b) For matched studies, give matching criteria and number of exposed and unexposed	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9-11
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	12
Bias	9	Describe any efforts to address potential sources of bias	7-9
Study size	10	Explain how the study size was arrived at	5-6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	9-11
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	11
		(b) Describe any methods used to examine subgroups and interactions	12
		(c) Explain how missing data were addressed	11-12, Appendix 1-4
		(d) If applicable, explain how loss to follow-up was addressed	11-12, Appendix 1-4
		(e) Describe any sensitivity analyses	N/A
<b>Results</b>			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	11-12, Appendix 1-4
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	Appendix 1-4
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	5-6
		(b) Indicate number of participants with missing data for each variable of interest	11-12, Appendix 1-4
		(c) Summarise follow-up time (eg, average and total amount)	7
Outcome data	15*	Report numbers of outcome events or summary measures over time	11-12, Appendix 1-4
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	12-14
		(b) Report category boundaries when continuous variables were categorized	N/A
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives	14
<b>Limitations</b>			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15-18
Generalisability	21	Discuss the generalisability (external validity) of the study results	17-18
<b>Other information</b>			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19

\*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).