PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Disease history and risk of comorbidity in the women's life course: a comprehensive analysis of the Japan Nurses' Health Study baseline survey
AUTHORS	Nagai, Kazue; Hayashi, Kunihiko; Yasui, Toshiyuki; Katanoda, Kota; Iso, Hiroyasu; Kiyohara, Yutaka; Wakatsuki, Akihiko; Kubota, Toshiro; Mizunuma, Hideki

VERSION 1 - REVIEW

REVIEWER	Frederick Naftolin md phd New York university school of medicine, usa
REVIEW RETURNED	05-Sep-2014

GENERAL COMMENTS	The skewed age distribution may have an effect on the statistical significance. The authors should mention this in the discussion and furnish an estimate of the impact on their results and conclusion. The Journal's statistician should vet this. This is an important and useful article. Its durability will be supported by the continuation of data as time passes. The authors should include a discussion of the limitations of using diagnoses rather than prevalence in self-reported studied such as this one. They can only report diagnoses, but there could be considerable sub-clinical disease present which could be playing important roles in the correlations that may be found and in the conclusions that can be drawn. For example, the metabolic syndrome has many effects on inflammatory conditions, such as arteriosclerosis (ASVD) that may be consequentially present without being symptomatic, e.g. requiring angina which is usually a sign of ADVANCED ASVD is rather late to discern the impact of ASVD. There are a few corrections needed: L4 "...life course of women." L13: "...estimated using diseases... Figure 1b is it cholelithiasis?

REVIEWER	Seo Young Park University of Pittsburgh, USA
REVIEW RETURNED	15-Dec-2014

GENERAL COMMENTS	2 & 11) The abstract and the conclusion say there were significant associations among the early-onset diseases, but in their data only 4 out of 5 early-onset diseases are significantly associated each other.

	12) Most limitations are adquately described, but I would add one more: recall bias. Sick people tend to think about their disease history more thus result in reporting better about their disease history. This may increase the ORs in their study. 15) Some sentences are very hard to understand because of bad writing. - page 3: "To classify age at peak... in the life course of a women" should be "To classify diseases based on age a peak indidence to identify the risk factor in the life course of a woman", because they are classifying diseases, not age at peaks. Also it should be either a woman or women, not 'a women'. -page 3: "Incidence rate and peak incidence were estimated using and diseases..." I think 'using' should be deleted from this sentence. -page 3: "The risk of comorbidity with early-onset diseases...." seems to be a wrong sentence. The subject (risk) and the complement (ovarian cancer,...) do not match. All those diseases are not risk. - There were numerous grammatical mistakes in the body of the manuscript, including stratum vs. strata. - I'd like to see some explanation why they chose age 45 as a cut-off for early vs. late onset of the disease. - They calculated M-H common ORs and also tested for homogeniety of OR between age <=50 vs. older. Why did they need to look at them separately? Whey age 50? I'd like to see some justification.

VERSION 1 – AUTHOR RESPONSE

> Reviewer: 1
> Reviewer Name frederick naftolin md phd
> Institution and Country new york university school of medicine, usa
> Please state any competing interests or state ‘None declared’: none

> The skewed age distribution may have an effect on the statistical significance.
> The authors should mention this in the discussion and furnish an estimate of the
> impact on their results and conclusion. The Journal's statistician should vet this.

Thank you for your suggestion. As reviewer 1 pointed out, the skewed age distribution may have an effect on statistical significance. For example, statistical significance was unlikely for very late-onset diseases such as osteoporosis because of the small sample size in the older age group. Therefore, we stratified the study population by age at the time of survey into two strata (less than 50 years and 50 years of age or older) and examined for homogeneity of ORs between the age groups. Statistically significant heterogeneity was found for several comorbidities with the four early-onset diseases. We estimated the adjusted ORs instead of crude ORs to correct for the skewed age distribution.

We added to the discussion as follows; (P13, L200)

“In this study, we observed a skewed age distribution because of the smaller sample size of participants aged 50 years or older. We stratified the study population by age at the time of the survey into two strata (<50 years and ≥50 years of age) and examined the homogeneity of ORs between the age groups. In addition, we estimated the common ORs between the two age groups instead of

overall crude ORs to adjust for the skewed age distribution. However, statistical significance in the comorbidity of very late-onset diseases such as osteoporosis was unlikely because of the small sample size in the older age group."

>This is an important and useful article. Its durability will be supported by the
> continuation of data as time passes.

Thank you for your comments. Further research based on follow-up studies is needed to clarify the cause–effect associations between these diseases.

> The authors should include a discussion of the limitations of using diagnoses rather
> than prevalence in self-reported studied such as this one. They can only report
> diagnoses, but there could be considerable sub-clinical disease present which could
> be playing important roles in the correlations that may be found and in the
> conclusions that can be drawn. For example, the metabolic syndrome has many
> effects on inflammatory conditions, such as arteriosclerosis (ASVD) that may be
> consequentially present without being symptomatic, e.g. requiring angina which is
> usually a sign of ADVANCED ASVD is rather late to discern the impact of ASVD.

Thank you for your suggestion. We added to the limitation as follows; (P16 L265–269)
"In this study, we defined disease onset as a diagnosis by a medical doctor that was reported on the self-administered questionnaire. Participants could only report a diagnosis; asymptomatic or undiagnosed diseases were excluded. Use of diagnoses rather than self-reported prevalence may affect correlation in some diseases."

> There are a few corrections needed:
> L4 "...life course of women."

Thank you for pointing this out to us. We corrected the objective in the abstract (P3 L3–4) as "To classify diseases based on age at peak incidence to identify the risk factors for later disease in the women's life course."

> L13: "...estimated using diseases...
Thank you for pointing this out to us. We corrected the main outcome measures in the abstract (P3 L12–14) as "The incidence rate and peak incidence for diseases whose peak incidence occurred before the age of 45 years or before the perimenopausal period were selected as early-onset diseases."

> Figure 1b is it cholelithiasis?
Thank you for letting us know. We corrected to "cholelithiasis".

> Reviewer: 2
> Reviewer Name Seo Young Park
> Institution and Country University of Pittsburgh, USA
> Please state any competing interests or state 'None declared': None declared

> 2 & 11) The abstract and the conclusion say there were significant associations

> among the early-onset diseases, but in their data only 4 out of 5 early-onset diseases
> are significantly associated each other.

Thank you for pointing this out to us.
In the abstract, we corrected sentence (P4 L27–29) as “While there were significant associations between four early-onset diseases, women with a history of one or more of the early-onset diseases had a higher risk of other diseases later in life course” and we moved the sentence “Four early-onset diseases (endometriosis, anaemia, migraine headache, and uterine myoma) were significantly correlated with one another.” from page 4 line 25–26 to page 4 line 18–19.

We corrected the sentence in the conclusion as follows; (P17 L290–292)
“While there were significant associations between four early-onset diseases (endometriosis, anaemia, migraine headache, and uterine myoma), women with a history of one or more of the early-onset diseases had a higher risk of other diseases later in life course.”

> 12) Most limitations are adquately described, but I would add one more: recall bias.
> Sick people tend to think about their disease history more thus result in reporting
> better about their disease history. This may increase the ORs in their study.

Thank you for your suggestion. We added recall bias to the limitation. (P16 L282–285)

“Recall bias may have caused overestimation of ORs since sick people tend to report more about disease history. However, the participants were nurses we think that recall bias was minimized since they have medical knowledge and are more likely to have answered correctly.”

>15) Some sentences are very hard to understand because of bad writing.

Sorry for the many grammatical mistakes. We have corrected those sentences as follows;

> - page 3: "To classify age at peak... in the life course of a women" should be "To
> classify diseases based on age a peak indidence to identify the risk factor in the life
> course of a woman", because they are classifying diseases, not age at peaks. Also it
> should be either a woman or women, not 'a women'.

We corrected the sentence in objectives in the abstract (P3 L3–4) as “To classify diseases based on age at peak incidence to identify the risk factors for later disease in the women’s life course.”

> -page 3: "Incidence rate and peak incidence were estimated using and diseases..." I
> think 'using' should be deleted from this sentence.

We corrected the sentence in the main outcome measures in the abstract (P3 L12–14) as “The incidence rate and peak incidence for diseases whose peak incidence occurred before the age of 45 years or before the perimenopausal period were selected as early-onset diseases.”

> -page 3: "The risk of comorbidity with early-onset diseases...." seems to be a wrong
> sentence. The subject (risk) and the complement (ovarian cancer,...) do not match. All
> those diseases are not risk.

We were corrected the sentence in the results in the abstract (P3 L19) as “Late-onset diseases significantly associated (OR >2) with early-onset diseases included …”

> - There were numerous grammatical mistakes in the body of the manuscript,

> including stratum vs. strata.
Thank you for your suggestion. The revised manuscript was checked by a native-English-speaking science editor.

> - I'd like to see some explanation why they chose age 45 as a cut-off for early vs. late
> onset of the disease.

We defined 45 years of age as being before the perimenopausal period because menopause represents a major transition event in a woman's life course. Other epidemiological studies (1,2,3) have defined the beginning of the perimenopausal period as 45 years of age.

1. Dudley EC, Hopper JL, Taffe J, Guthrie JR, Burger HG, Dennerstein L. Using longitudinal data to define the perimenopause by menstrual cycle characteristics. Climacteric. 1998;1(1):18-25.
2. McKinlay SM, Brambilla DJ, Posner JG. The normal menopause transition. Maturitas. 2008;61(1-2):4-16.
3. Brambilla DJ, McKinlay SM, Johannes CB. Defining the perimenopause forapplication in epidemiologic investigations. Am J Epidemiol. 1994 15;140(12):1091-5.

> - They calculated M-H common ORs and also tested for homogeniety of OR between
> age <=50 vs. older. Why did they need to look at them separately? Whey age 50? I'd
> like to see some justification.

Thank you for your suggestion. As reviewer 1 pointed out, the skewed age distribution may have an effect on statistical significance. For example, statistical significance was unlikely for very late-onset diseases such as osteoporosis because of the small sample size in the older age group. Therefore, we stratified the study population by age at the time of survey into two strata (less than 50 years and 50 years of age or older) and examined for homogeneity of ORs between the age groups. Statistically significant heterogeneity was found for several comorbidities with the four early-onset diseases. We estimated the adjusted ORs instead of crude ORs to correct for the skewed age distribution.
We added to the discussion as follows; (P13 L200)
“In this study, we observed a skewed age distribution because of the smaller sample size of participants aged 50 years or older. We stratified the study population by age at the time of the survey into two strata (<50 years and ≥50 years of age) and examined the homogeneity of ORs between the age groups. In addition, we estimated the common ORs between the two age groups instead of overall crude ORs to adjust for the skewed age distribution. However, statistical significance in the comorbidity of very late-onset diseases such as osteoporosis was unlikely because of the small sample size in the older age group.”

In addition, we reported a median age at menopause of around 51 years* in this study population. Therefore we also used age 50 as the division for before menopause and after menopause.

*Yasui T, Hayashi K, Mizunuma H, Kubota T, Aso T, Matsumura Y, Lee JS, Suzuki S. Factors associated with premature ovarian failure, early menopause and earlier onset of menopause in Japanese women. Maturitas. 2012;72(3):249-55.

VERSION 2 – REVIEW

REVIEWER	Frederick Naftolin md phd New York university school of medicine, usa
REVIEW RETURNED	29-Jan-2015

GENERAL COMMENTS	I appreciate the authors' effort to improve the ms, and they have succeeded.