

## 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

<b>TITLE (PROVISIONAL)</b>	A Prospective Comorbidity-Matched Study of Parkinson's Disease and Risk of Mortality Among Women
<b>AUTHORS</b>	Winter, Anke; Rist, Pamela; Buring, Julie; Kurth, Tobias

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Angus Macleod University of Aberdeen, UK
<b>REVIEW RETURNED</b>	01-Apr-2016

<b>GENERAL COMMENTS</b>	<p>This is a well conducted study which reports mortality in women with PD using a matched cohort design. At least 17 studies have previously published data on mortality in women with PD and, in terms of sample size (55 deaths) this manuscript does not add a huge amount to the previous literature, but the study is well conducted and does deserve to be published.</p> <p>I have several comments:</p> <p>Methods:</p> <ol style="list-style-type: none"> <li>1. In terms of the methods re analyses, covariate adjustment for age is included, which is (to met) counter-intuitive in a study which is matched on age. I do think this is the correct way to analyse this study, but an explanation why would be helpful together with a reference e.g. Sjölander, A. and Greenland, S. (2013), Ignoring the matching variables in cohort studies – when is it valid and why?. <i>Statist. Med.</i>, 32: 4696–4708. doi:10.1002/sim.5879.</li> <li>2. It was only mentioned at the end of the discussion that the risk set for the analyses was age. This firstly needs to be in the methods but, secondly, an explanation of why this was chosen is also needed. Age is adjusted for as a covariate in the models and the standard way to analyse a cohort study of PD would be to use time from diagnosis (time from index date for controls). I would be interested to know how the choice of risk set affects the results.</li> <li>3. Why were Kaplan-Meier curves not presented?</li> <li>4. I am unclear what was analysed in the competing risk Cox model(s). What was the outcome of interest and what were the competing outcomes?</li> <li>5. There is no mention of ethical approval.</li> </ol> <p>Results</p> <ol style="list-style-type: none"> <li>6. The word controls is missing from line 23 on page 7.</li> </ol>
-------------------------	--

	<p>7. You say that the interaction between smoking and PD did not reach statistical significance (<math>p=0.08</math>). However, interaction tests have low power and I think this suggests something potentially interesting, particularly as 2 previous studies have reported a similar interaction.</p> <p>8. However, re the analyses to look at interactions, multiple analyses of subgroups are generally more likely to lead to false positive findings rather than true findings and this caveat should be mentioned. See doi: 10.1136/bmj.h5651.</p> <p>9. The numbers lost to follow-up must be reported (see STROBE).</p> <p>Discussion</p> <p>10. It would be helpful to quote the hazard ratios from the two studies you quote in the first paragraph of the discussion.</p> <p>11. In the second paragraph, the authors discuss confounding by comorbidities. They argue that controlling this is important because co-morbidities are common in the age groups in which PD presents. However, this in itself does not lead to confounding. It is only an issue if co-morbidity burden is different between patients and controls, which they have not discussed. This requires further discussion.</p> <p>12. In line 36 and 38 the authors state: "In this case, lack of control for comorbidities may underestimate the true effect of PD on risk of mortality." I think this is correct, but the authors should explain why (e.g. cancer and smoking-related comorbidities may be less common in PD).</p> <p>13. The authors should discuss other issues apart from race why generalisability may be less (e.g. socio-economic status being higher, those who consent to a trial unlikely to be representative of the general population). There is clear evidence the population is not representative of the general population with PD as fewer than 10% died over mean 6 years, which is much lower mortality rate than in population-based studies. However, it is unclear how this unrepresentativeness influences a hazard ratio as the same issues will apply to both cases and controls.</p> <p>Table</p> <p>14. Ages are presumably mean ages. Please state.</p> <p>15. The time-point for mean duration of PD is unclear.</p> <p>16. Charlson score total is an odd measure to give. I think a mean would be more meaningful.</p>
--	--

<b>REVIEWER</b>	Walter Pirker Head of Dept. Department of Neurology, Wilhelminenspital der Stadt Wien Vienna, Austria
<b>REVIEW RETURNED</b>	12-Apr-2016

<b>GENERAL COMMENTS</b>	This is an excellent epidemiological paper on the mortality of PD in women. The methods are sound. The results are in line with previous findings in male patients and represent a relevant addition
-------------------------	--

	<p>to the published literature by focussing on a large sample of women with PD and controls drawn from a prospective population based study. Patients and controls were carefully matched for clinical and demographic variables including comorbidities.</p> <p>The length of the paper is adequate, the results are discussed appropriately.</p> <p>A very minor point: The second sentence of the methods section seems to need editing: After the end of the trial in 2004, women were asked if they would be willing to continued to be followed on an observational basis.</p>
--	--

<b>REVIEWER</b>	<p>Giancarlo Logroscino MD PhD Center for Neurodegenerative Diseases Department of Basic Medicine Neuroscience and Sense Organs Department of Clinical Research in Neurology of the Fondazione Giovanni Panico University of Bari Aldo Moro.</p>
<b>REVIEW RETURNED</b>	15-Apr-2016

<b>GENERAL COMMENTS</b>	<p>Matching is a difficult business. What are the possible problems of matching on comorbidity in a study on mortality? The authors should add some considerations/ thoughts on this topic in the discussion.</p> <p>Did the authors explore the effect of PD age of onset on mortality?</p> <p>Comorbidities are not all equal. Did vascular or metabolic diseases like diabetes 2 a role in chenging the mortality risk?</p> <p>Are results generalizable?</p>
-------------------------	--

### VERSION 1 – AUTHOR RESPONSE

Reviewer: 1  
 Reviewer Name: Angus Macleod  
 Institution and Country: University of Aberdeen, UK  
 Competing Interests: None

This is a well conducted study which reports mortality in women with PD using a matched cohort design. At least 17 studies have previously published data on mortality in women with PD and, in terms of sample size (55 deaths) this manuscript does not add a huge amount to the previous literature, but the study is well conducted and does deserve to be published.

I have several comments:

Methods:

1. In terms of the methods re analyses, covariate adjustment for age is included, which is (to me) counter-intuitive in a study which is matched on age. I do think this is the correct way to analyse this study, but an explanation why would be helpful together with a reference e.g. Sjölander, A. and Greenland, S. (2013), Ignoring the matching variables in cohort studies – when is it valid and why?. Statist. Med., 32: 4696–4708. doi:10.1002/sim.5879.

Thank you for providing this reference. As mentioned in the reference, in a matched cohort study, when other variables besides the matching factors are included in the multivariable model, you must

also include the matching variables in the multivariable model. We have revised our methods to adjust for both age at index date and Charlson comorbidity score at index date. Additionally, we have added the following sentence to explain why we adjusted for age and Charlson comorbidity score at index date: “Although we matched on age and Charlson comorbidity score at index date, when additional covariates are included in the model, the matching factors also need to be adjusted for to obtain unbiased effect estimates (Sjölander A et al.)” (page 7)

2. It was only mentioned at the end of the discussion that the risk set for the analyses was age. This firstly needs to be in the methods but, secondly, an explanation of why this was chosen is also needed. Age is adjusted for as a covariate in the models and the standard way to analyse a cohort study of PD would be to use time from diagnosis (time from index date for controls). I would be interested to know how the choice of risk set affects the results.

We have added the following sentence to the methods: “Because age is a strong risk factor for death, we used age instead of time-on-study as our time scale for these analyses (reference added for Korn et al. AJE 1997).” The citation provided explains why it is often preferable to use age as the time scale. (page 7)

We ran a sensitivity analysis using time-on-study as the time scale and found similar results to the model with age as the time scale (HR=2.04; 95% CI: 1.30-3.19). However, given the strong association between age and death, we believe that it is more appropriate to use age as the time scale and have shown the results with age as the time scale in the manuscript.

3. Why were Kaplan-Meier curves not presented?

We have added in Kaplan Meier curves to our results section: “The Kaplan-Meier curve for overall survival is displayed in Figure 1 and shows the PD cases have an increased risk of death compared to the comparators (p-value < 0.01).” (page 8 and Figure 1)

4. I am unclear what was analysed in the competing risk Cox model(s). What was the outcome of interest and what were the competing outcomes?

There were three possible outcomes: death from cardiovascular disease, death from cancer, and death from other causes (not cardiovascular disease or cancer). We have tried to clarify this in the methods by adding the following sentence: “We used a competing risk Cox model adjusted for age and smoking status to analyze the association between PD and three competing causes of death: cardiovascular disease, cancer, or other illnesses.” (page 8)

5. There is no mention of ethical approval.

We have added the following phrase to the methods section: “The Women’s Health Study was approved by the Institutional Review Board at Brigham and Women’s Hospital.” (page 6)

## Results

6. The word controls is missing from line 23 on page 7.

This has been added in. (page 8 in the revised version)

7. You say that the interaction between smoking and PD did not reach statistical significance (p=0.08). However, interaction tests have low power and I think this suggests something potentially interesting, particularly as 2 previous studies have reported a similar interaction.

We agree that interaction tests can have low power and have added this as potential limitation (From the discussion section: “Our analyses of potential effect modification by smoking or age at onset should be interpreted with caution given the low power to detect effects in strata and the possibility of false positive findings when several subgroups are analyzed”. (see page 11) However, since prior studies have shown some evidence of effect modification by smoking status, we thought that it was important to explore this option in our study and decided to report the stratum-specific results even

though the formal test of interaction did not reach statistical significance.

8. However, re the analyses to look at interactions, multiple analyses of subgroups are generally more likely to lead to false positive findings rather than true findings and this caveat should be mentioned. See doi: 10.1136/bmj.h5651.

We agree that false positive finding can occur when exploring several potential effect modifiers. We have added the following sentence to our discussion of the limitations of this study and cite the reference provided by the reviewer: "Our analyses of potential effect modification by smoking or age at onset should be interpreted with caution given the low power to detect effects in strata and the possibility of false positive findings when several subgroups are analyzed". (page 11)

Despite the potential for low power and false positive findings, we chose to explore potential effect modifiers which have been suggested as modifiers from prior work. This prior work guided our a priori hypotheses about potential effect modification by smoking status, duration of PD, or age at onset.

9. The numbers lost to follow-up must be reported (see STROBE).

We have added the following sentence to our methods section "Mortality follow-up is over 99% complete in the WHS." (page 7)

#### Discussion

10. It would be helpful to quote the hazard ratios from the two studies you quote in the first paragraph of the discussion.

These hazard ratios have been added: "Results are similar to those observed in a study among men which also performed age and comorbidity matching among men enrolled in the Physician's Health Study[21] (HR=2.32; 95% CI: 1.85-2.92)" (page 9)

11. In the second paragraph, the authors discuss confounding by comorbidities. They argue that controlling this is important because co-morbidities are common in the age groups in which PD presents. However, this in itself does not lead to confounding. It is only an issue if co-morbidity burden is different between patients and controls, which they have not discussed. This requires further discussion.

We have significantly revised our discussion of how confounding by comorbidities would impact our results. We have added the following discussion to our manuscript "Confounding by comorbidities could result in mainly an overestimate or underestimate of the true effect of PD on mortality. For example, some studies have suggested that individuals with PD are less likely to have a prior diagnosis of most cancers than individuals without PD.[26–28] Lack of adequate control for cancer comorbidities would result in an underestimate of the effect of PD on mortality. Similarly, smokers are less likely to develop PD[29], but smoking is linked with several other comorbidities. A higher prevalence of smoking-related comorbidities among the matched comparators without PD may also result in an underestimate of the true effect of PD on mortality." (page 10)

12. In line 36 and 38 the authors state: "In this case, lack of control for comorbidities may underestimate the true effect of PD on risk of mortality." I think this is correct, but the authors should explain why (e.g. cancer and smoking-related comorbidities may be less common in PD).

As mentioned in the response to comment #11, we have significantly revised our discussion of how confounding by comorbidities would impact our results. (see page 10)

13. The authors should discuss other issues apart from race why generalisability may be less (e.g. socio-economic status being higher, those who consent to a trial unlikely to be representative of the general population). There is clear evidence the population is not representative of the general population with PD as fewer than 10% died over mean 6 years, which is much lower mortality rate than in population-based studies. However, it is unclear how this unrepresentativeness influences a

hazard ratio as the same issues will apply to both cases and controls.

We have expanded our discussion of generalizability to indicate other factors besides race which might influence the generalizability of our results: "Our cohort was composed of female health professionals who are primarily white, which may limit the generalizability of our findings to other racial or ethnic groups or to populations of different socioeconomic status. Additionally, our results may not be generalizable to populations with younger ages at PD onset since the average age at PD onset in our population was 70 years of age." (page 11)

Table

14. Ages are presumably mean ages. Please state.

We have clarified this (see Table 1).

15. The time-point for mean duration of PD is unclear.

This is the time from PD onset to death, loss to follow-up, or end of study. We have clarified this in the methods section: "Mean duration of PD was calculated from PD onset to death, loss to follow-up, or end of study." (page 7)

16. Charlson score total is an odd measure to give. I think a mean would be more meaningful.

We agree with this suggestion and have changed Table 1 to include the mean Charlson score in the PD cases and comparators (see Table 1).

Reviewer: 2

Reviewer Name: Walter Pirker

Institution and Country: Head of Dept. Department of Neurology, Wilhelminenspital der Stadt Wien, Vienna, Austria

Competing Interests: none declared

This is an excellent epidemiological paper on the mortality of PD in women. The methods are sound. The results are in line with previous findings in male patients and represent a relevant addition to the published literature by focussing on a large sample of women with PD and controls drawn from a prospective population based study. Patients and controls were carefully matched for clinical and demographic variables including comorbidities.

The length of the paper is adequate, the results are discussed appropriately.

A very minor point:

The second sentence of the methods section seems to need editing:

After the end of the trial in 2004, women were asked if they would be willing to continued to be followed on an observational basis.

We have edited this sentences to "After the end of the trial in 2004, women were asked if they would be willing to continue to be followed on an observational basis." (page 6)

Reviewer: 3

Reviewer Name: Giancarlo Logroscino MD PhD

Institution and Country: University of Bari Aldo Moro, Italy

Competing Interests: None declared

Matching is a difficult business. What are the possible problems of matching on comorbidity in a study on mortality? The authors should add some considerations/ thoughts on this topic in the discussion.

We believe that it is important to match on comorbidities status in order to understand the true effect of PD on mortality. Unless comorbidities are adequately controlled for, any association seen between

PD and mortality may be due to the presence of other comorbidities and not to the effect of PD on mortality. We have revised our discussion section to explain how confounding by comorbidities may bias results (page 10).

One disadvantage to matching is that we are unable to explore the prevalence of various comorbidities in participants with PD compared to participants without PD. However, this was not a primary aim of this study.

Did the authors explore the effect of PD age of onset on mortality?

We explored the effect of PD age of onset on mortality and found no evidence of effect modification by age. We include this analysis in our methods and have added the following sentences to our results section:

“The association between PD and mortality was similar for those <70 years of age (HR=2.22; 95% CI: 1.04, 4.71) and those ≥70 years of age (HR=2.26; 95% CI: 1.27, 4.01). The interaction between age at onset and PD was not statistically significant (p-value = 0.81).” (page 9)

Comorbidities are not all equal. Did vascular or metabolic diseases like diabetes 2 a role in changing the mortality risk?

While we agree that not all comorbidities are equal, we did not explore the role of individual comorbidities on changing the mortality risk. Due to the low number of outcome events, we did not have enough power to run any meaningful analyses stratified by particular risk factors like diabetes.

Are results generalizable?

We have expanded our discussion of the generalizability of our results in response to yours and Reviewer 1’s comment. We have added the following sentences to the discussion: “Our cohort was composed of female health professionals who are primarily white, which may limit the generalizability of our findings to other racial or ethnic groups or to populations of different socioeconomic status. Additionally, our results may not be generalizable to populations with younger ages at PD onset since the average age at PD onset in our population was 70 years of age.” (page 11)

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Angus Macleod University of Aberdeen, UK
<b>REVIEW RETURNED</b>	27-May-2016

<b>GENERAL COMMENTS</b>	<p>The changes have improved the clarity of the manuscript and I recommend this paper be accept without further revisions.</p> <p>Only one small note: I disagree about the choice of risk sets for survival analysis, and think it is preferable to use a common time point where possible (time of diagnosis in this study). This seems to be what Korn suggests in the paper you have cited: "For example, in a randomized clinical trial or in a natural history study of a disease, time since randomization or diagnosis would be used".</p>
-------------------------	--

<b>REVIEWER</b>	GIANCARLO LOGROSCINO UNIVERSITY OF BARI "ALDO MORO" BARI - ITALIY
<b>REVIEW RETURNED</b>	13-Jun-2016

<b>GENERAL COMMENTS</b>	<p>The authors have improved the manuscript;</p> <p>The possible problems of matching ( even when properly done) should be more emphasized in the discussion;</p> <p>The clinical diagnosis of PD has problems also in tertiary centers with MD specialists as recently shown ( Rizzo et al Neurology 2016);</p>
-------------------------	--

## VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Angus Macleod

Institution and Country: University of Aberdeen, UK

Competing Interests: None

The changes have improved the clarity of the manuscript and I recommend this paper be accept without further revisions.

1. Only one small note: I disagree about the choice of risk sets for survival analysis, and think it is preferable to use a common time point where possible (time of diagnosis in this study). This seems to be what Korn suggests in the paper you have cited: "For example, in a randomized clinical trial or in a natural history study of a disease, time since randomization or diagnosis would be used".

RESPONSE:

Thank you very much for the thoughtful feedback. Age is a very strong predictor of PD onset and mortality so we have decided to control for age by matching on age, including it as a covariate in all analyses and using age as the primary time scale in our analyses. We have rerun all analyses using time since diagnosis/time on study as time scale to address the reviewers' concern. Our results did not change when using time since diagnosis/time on study as time scale HR=2.04; 95% CI: 1.30-3.19).

Reviewer: 3

Reviewer Name: Giancarlo Logroscino

Institution and Country: University of Bari "Aldo Moro" Bari - Italy

Competing Interests: None declared

The authors have improved the manuscript.

1. The possible problems of matching (even when properly done) should be more emphasized in the discussion;

RESPONSE

Thank you very much for the valuable feedback. We have added a brief paragraph acknowledging the limitations of our study design (please refer to page 11). As our study is a matched cohort as opposed to a case-control study and we performed exact matching, usual concerns associated with matching such as overmatching and residual confounding within m matching strata shouldn't apply to our study design.

2. The clinical diagnosis of PD has problems also in tertiary centers with MD specialists as recently shown (Rizzo et al Neurology 2016).

RESPONSE

Thank you very much for the meaningful feedback. We have expanded our limitation section to further highlight the possibility of misclassification of PD cases in our study and included the suggested reference (please refer to page 10).