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)	Association of Chronic Insomnia Symptoms and Recurrent Extreme Sleep Duration over 10 Years with Well-being in Older Adults: A Cohort Study
AUTHORS	Abell, Jessica; Shipley, Martin; Ferrie, Jane; Kivimäki, Mika; Kumari, Meena

VERSION 1 - REVIEW

REVIEWER	Mihai Teodorescu, MD University of Wisconsin Madison, Wisconsin, USA
REVIEW RETURNED	10-Aug-2015

GENERAL COMMENTS	Authors report on their study of the association between insomnia symptoms, sleep duration and well-being in a cohort of UK civil servants. Sleep data was obtained by self report 3 times over a period of 10 years; well-being was measured once. This study does not bring a significant amount of new information. Self reported sleep duration has been employed in studies for many years and more recent data correlates with some kind of objective measure (e.g. actigraphy). In addition, just 3 measurements in 10 years may not accurately describe real life scenarios for the involved population. Sleep disorders are not accounted for. Drop-outs have been significant, and statistical analysis appears to suggest that a different population remained enrolled in the study compared with the drop-outs. Therefore, extrapolating conclusions to a general population is questionable.

REVIEWER	Xiaoli Chen Harvard School of Public Health, USA
REVIEW RETURNED	27-Aug-2015

GENERAL COMMENTS	General Comments This is a generally clear and well-written manuscript with excellent tables. The content was well organized. This interesting paper was based on an important, large longitudinal study of 4491 UK civil servants. This study was very well designed and executed. Some thoughts to consider or need further explanation: Title ‘Recurrent Sleep Duration’ could be changed to: ‘Recurrent Short and Long Sleep duration’ or ‘Recurrent Extreme Sleep Duration’

	Abstract What's the percentage of female participants? To be consistent with the title order (insomnia, short and long sleep duration), it would be helpful to provide information on chronic insomnia first, and then describe short and long sleep duration in Abstract, Methods, Results, and Discussion sections. Please spell out 'CASP-19' and 'SE' when they came out in the text for the first time. Methods Please provide the information on the percentage of female participants in Methods or Results section. What's the response rate in this study? The follow-up rate from phase 5 to phase 9? What's the correlation between CASP-19 (QOL measure for old adults) and SF-36 (general QOL measure) in this study? How did the authors define 'vigorous activity'? what types of activities were called 'vigorous activity'? References for GHQ should be provided. Does high GHQ score indicate more depressive symptoms? In the statistical analysis, which software (e.g., SAS, SPSS) did the authors use? What's the reference group for sleep duration in the models? 7 hours or 7-8 hours? Any literature for the reference? Did the authors conduct any trend test to examine the dose-response relationships for chronic insomnia, short sleep, and long sleep with poor well-being? Table 1. Please provide notes for the abbreviations including CASP-19, SF-36, PCS, MCS, ADL, IADL, GHQ, BMI. % (N) should be indicated. In Table 1, the order of variables in the first column could be changed as: sociodemographic factors (age, women, married, employed, wealth), lifestyle factors (alcohol consumption, physical activity, smoking, BMI), comorbidities (long-term illness, ADL, IADL, depression), insomnia, and QOL measures (CASP-19, SF-36) Table S3, what's the definition of recurrent short sleep duration? <6 hrs? <=5 hours?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Mihai Teodorescu, MD, University of Wisconsin, Madison, Wisconsin, USA

Authors report on their study of the association between insomnia symptoms, sleep duration and well-being in a cohort of UK civil servants. Sleep data was obtained by self report 3 times over a period of 10 years; well-being was measured once.

1) This study does not bring a significant amount of new information. Self reported sleep duration has been employed in studies for many years and more recent data correlates with some kind of objective measure (e.g. actigraphy). In addition, just 3 measurements in 10 years may not accurately describe real life scenarios for the involved population.

We agree with the reviewer that 3 measurements over 10 years may not accurately describe real-life sleep scenarios and we have now added this reservation to the limitations section of the manuscript (page 13). However, to the best of our knowledge no other studies have published data on repeat measurements of sleep indicators over a 10 year period and well-being. In addition, our outcome measures, which have been designed primarily for use with older people, have allowed us to examine

associations between extreme sleep duration or insomnia symptoms and different aspects of well-being (mental, physical, overall). Our finding that in older adults short sleep duration alone does not have a negative effect on overall or mental well-being, when the effects of insomnia or depressive symptoms are taken into account is, we feel, of interest. We apologise for not highlighting these features clearly enough in our manuscript and have now corrected this in the introduction (page 4) and discussion (14).

As observed by the reviewer 'objective' measures of sleep such as actigraphy are extremely useful tools in sleep research. However, these devices have only been available for observational research in the last few years. Therefore, it is not yet possible to examine the association of long-term sleeping patterns and well-being using sleep assessed by objective measures in large observational studies. We agree with the reviewer that recent data has shown correlation between self-reported sleep duration and objective measures of sleep [1-3]. Self-reported sleep duration has also been shown to be associated with objective health outcomes [4 5]. Such findings underscore the validity of self-reported measures of sleep, which are also the measures generally employed for the assessment of sleep duration and insomnia in the primary care setting. We have drawn attention to these considerations in the section of the manuscript which considers the limitations of our study (page 13).

2) Sleep disorders are not accounted for.

We agree with the reviewer that not being able to account for sleep disorders is a limitation for our study. We have highlighted this in the limitations section of the manuscript (page 13). Sleep apnoea, one of the most prevalent sleep disorders, is strongly correlated with obesity. We hope that our adjustment for body mass index may go some way to deal with this disorder. We also note that further research in this area should collect information on a range of sleep disorders.

3) Drop-outs have been significant, and statistical analysis appears to suggest that a different population remained enrolled in the study compared with the drop-outs. Therefore, extrapolating conclusions to a general population is questionable.

We thank the reviewer for the highlighting this issue. We acknowledge that as with any long running cohort, drop-out is always an issue. Despite differences between the two populations being observed (in Table S3) it is likely that those participants who remain in the study will be younger and healthier than those who leave [6]. This has been observed previously in this cohort. Therefore, it is possible that poor sleep is under-reported in these participants. Furthermore, although caution should always be exercised before generalizing to the general population the findings from an occupational cohort of middle-aged, white-collar civil servants (who are the population in Whitehall II), we have shown previously that standard risk factor-cardiovascular disease associations found in this study are in close agreement with those observed in a UK-wide general population study (British Regional Heart Study) and the community-based Framingham study [7]. We have highlighted both of these limitations in the appropriate section of this manuscript (pages 13-14).

Reviewer: 2 (Xiaoli Chen, Institution and Country, Harvard School of Public Health, USA)

This is a generally clear and well-written manuscript with excellent tables. The content was well organized. This interesting paper was based on an important, large longitudinal study of 4491 UK civil servants. This study was very well designed and executed.
We thank the reviewer for these positive comments.

Some thoughts to consider or need further explanation:

Title ‘Recurrent Sleep Duration’ could be changed to: ‘Recurrent Short and Long Sleep duration’ or ‘Recurrent Extreme Sleep Duration’
We agree that the suggested amendment to the title would make the subject of the paper clearer and have now changed the title to “Association of Chronic Insomnia Symptoms and Recurrent Extreme Sleep Duration over 10 Years with Well-being in Older Adults: A Cohort Study" (page 1)

Abstract

What’s the percentage of female participants?

We have now included the percentage of female participants (25.2%) in the abstract (page 2).

To be consistent with the title order (insomnia, short and long sleep duration), it would be helpful to provide information on chronic insomnia first, and then describe short and long sleep duration in Abstract, Methods, Results, and Discussion sections.

We thank the reviewer for this suggestion, which we agree would be helpful to the structure of the manuscript. We have changed we order in which we discuss each of the sleep exposures so that we provide information on insomnia first followed by sleep duration.

Please spell out ‘CASP-19’ and ‘SE’ when they came out in the text for the first time.

We thank the reviewer for spotting this omission; we have now spelt out CASP-19 and SE in the abstract. (page 2)

Methods
Please provide the information on the percentage of female participants in Methods or Results section.

The percentage of female participants has now been added to the study sample section of the Methods section (page 5).

What’s the response rate in this study? The follow-up rate from phase 5 to phase 9?

The response rate in the study was 66% since phase 1 and 86% from those eligible at phase 9 and the follow-up rate from phase 5 to phase 9 was 85.9%. This has been added to the study sample section of the Methods section (page 5).

What’s the correlation between CASP-19 (QOL measure for old adults) and SF-36 (general QOL measure) in this study?

The correlation between CASP-19 and SF-36 (MCS) is ($r=0.64$, $p\leq0.001$). This has been added to the section of the Methods section which discusses the SF-36 questionnaire (page 5).

How did the authors define ‘vigorous activity’? What types of activities were called ‘vigorous activity’?

We thank the reviewer for pointing out that this was not defined adequately. We have included the following information and appropriate references in the covariates section of the Methods section and given example of types of vigorous activity (page 6):

Physical activity was assessed using a questionnaire which asked participants about the number of hours spent undertaking a range of physical activity (both leisure- time and job-related activities).

Each activity was assigned a metabolic equivalent (MET) value[8]. Vigorous physical activity was defined as activities with a MET value of 6 or more[9] (e.g. swimming, mowing).

References for GHQ should be provided. Does high GHQ score indicate more depressive symptoms?

We have included a reference for the GHQ. Higher GHQ scores indicate more depressive symptoms and this has also been stated (page 7).

In the statistical analysis, which software (e.g., SAS, SPSS) did the authors use?

We used Stata 13.1 to conduct the statistical analysis; this information has been added to the statistical analysis section of the Methods section (page 7).

What's the reference group for sleep duration in the models? 7 hours or 7-8 hours? Any literature for the reference?

The reference group for sleep duration in the models in Tables 2 & 3 is 'no recurrent short sleep duration' or 'no recurrent long sleep duration'. In Table S1 the reference group is 7 hours of sleep. This has been added to the statistical analysis section and a reference justifying this choice provided (page 7).

Did the authors conduct any trend test to examine the dose-response relationships for chronic insomnia, short sleep, and long sleep with poor well-being?

We had reported the results of the trend test to examine the dose-response relationships for chronic insomnia, short sleep or long sleep with the well-being outcomes. We thank the reviewers for bringing this to our attention. We have reported these in the results section (pages 9:11) and also added a line to the methods section, describing the tests used (page 7). We used the Stata command nptrend which is an extension of the Wilcoxon rank-sum test.

Table 1. Please provide notes for the abbreviations including CASP-19, SF-36, PCS, MCS, ADL, IADL, GHQ, BMI. % (N) should be indicated.

In Table 1, the order of variables in the first column could be changed as: sociodemographic factors (age, women, married, employed, wealth), lifestyle factors (alcohol consumption, physical activity, smoking, BMI), comorbidities (long-term illness, ADL, IADL, depression), insomnia, and QOL measures (CASP-19, SF-36)

We thank the reviewers for pointing out some ways to make Table 1 clearer. We have followed the suggestion to re-order the variables in the first column. We have also indicated % (N) at the beginning of each relevant row rather than once. We have also provided notes for the abbreviations included in Table 1 (page 10).

Table S3, what's the definition of recurrent short sleep duration? <6 hrs? <=5 hours?

In Table S3 recurrent short sleep duration was defined as <=5hours, a footnote describing this has now been added to the supplementary table.

1. Signal TL, Gale J, Gander PH. Sleep measurement in flight crew: comparing actigraphic and subjective estimates to polysomnography. Aviat. Space Environ. Med. 2005;76(11):1058-63
2. Lockley SW, Skene DJ, Arendt J. Comparison between subjective and actigraphic measurement of sleep and sleep rhythms. J. Sleep Res. 1999;8(3):175-83 doi: 10.1046/j.1365-

2869.1999.00155.x[published Online First: Epub Date]|.
3. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? Epidemiology (Cambridge, Mass.) 2008;19(6):838-45 doi: 10.1097/ede.0b013e318187a7b0[published Online First: Epub Date]|.
4. Ferrie JE. A prospective study of change in sleep duration: associations with mortality in the Whitehall II cohort. Sleep 2007;30(12):1659
5. Hublin C, Partinen M, Koskenvuo M, Kaprio J. Sleep and mortality: a population-based 22-year follow-up study. Sleep 2007;30(10):1245-53
6. Ferrie JE, Kivimäki M, Singh-Manoux A, et al. Non-response to baseline, non-response to follow-up and mortality in the Whitehall II cohort. Int. J. Epidemiol. 2009;38(3):831-37 doi: 10.1093/ije/dyp153[published Online First: Epub Date]|.
7. Batty GD, Shipley M, Tabak A, et al. Generalizability of occupational cohort study findings: Epidemiology. 2014 Nov;25(6):932-3. doi: 10.1097/EDE.0000000000000184.
8. Sabia S, Dugravot A, Kivimaki M, Brunner E, Shipley MJ, Singh-Manoux A. Effect of Intensity and Type of Physical Activity on Mortality: Results From the Whitehall II Cohort Study. Am. J. Public Health 2011;102(4):698-704 doi: 10.2105/AJPH.2011.300257[published Online First: Epub Date]|.
9. WHO. Recommended Amount of Physical Activity. Switzerland: World Health Organization 2010. .

VERSION 2 – REVIEW

REVIEWER	Xiaoli Chen USA
REVIEW RETURNED	20-Oct-2015

GENERAL COMMENTS	General comments The revision is much improved. Specific comments SF-36 is a measure of health status across eight sections (health-related quality of life, HRQOL). On page 5, the authors mentioned 'The correlation between CASP-19 and SF-36 mental well-being was r=0.64'. The authors may want to provide the correlation between overall well-being measured by CASP-19 and health status/HRQOL measured by SF-36 (including both SF-36 PCS and MCS) rather than only SF-36 mental well-being. Statistical analysis: How were statistical significance levels defined? Is it 'Statistical significance levels were set at P <0.05 for two-sided analyses'? Results: 1st paragraph, it would be helpful that the authors describe the information on the percentages of participants with short and long sleep duration and chronic insomnia symptoms (two major exposure variables) in this study population. Table 1. The title is "Characteristics of participants by sleep duration 2007-2009 (N=4,491)". In the footnote, the authors provided the percentages with sample sizes of chronic insomnia symptoms under 4 subgroups: No occurrence, 1 occurrence, 2 occurrences, 3 occurrences. It would be helpful if the authors could provide detailed information about the 3 time points (1997-1999, 2003-2004, and 2007-2009) used to define high levels of chronic insomnia symptoms. The authors may want to revise the first column by placing mean (SD) or placing % (N) after each characteristic, e.g., Age (years),

	mean (SD); Women, % (N) There is no footnote for GHQ (General Health Question). On page 9, line 38, 'Table 2 shows the results for recurrent…' could be considered as a separate paragraph. Table 2. It would be helpful that in the footnote, the authors indicate that overall well-being was measured by CASP-19, while physical well-being and mental well-being were assessed by SF-36. Results for the trend tests could be added to Table 2. It is interesting that the sample sizes for Recurrent short sleep subgroups (no short sleep N=2,842); one occurrence n=782; two occurrences n=499; and three occurrences n=368) are exactly the same as those for Chronic insomnia symptoms. Please double check.

VERSION 2 – AUTHOR RESPONSE

1) SF-36 is a measure of health status across eight sections (health-related quality of life, HRQOL). On page 5, the authors mentioned 'The correlation between CASP-19 and SF-36 mental well-being was r=0.64'. The authors may want to provide the correlation between overall well-being measured by CASP-19 and health status/HRQOL measured by SF-36 (including both SF-36 PCS and MCS) rather than only SF-36 mental well-being.
As we understand it the reviewer is asking for a correlation of the SF-36 as a total score (including both SF-36 PCS and MCS together. It is our understanding that the SF-36 measure wasn't designed to be used as a total score, since it was developed as 8 separate scales which could be used as two separate summary measures (PCS and MCS) [1]. For these reasons we have not provided a correlation between overall well-being as measured by CASP-19 and HRQOL measured by SF-36. However, the section which introduces this scale in the methods section is quite misleading and does suggest that there is one 'health related well-being'. We have altered this part of the method to make this clearer:
"The Short Form 36 health survey (SF-36) is a 36 item questionnaire; these questions are used to construct the eight SF-36 scales: physical functioning, mental functioning, role limitations due to physical problems, social functioning, bodily pain, role limitations due to emotional problems, vitality, and general health perceptions. [21] These eight scales can be aggregated to form two summary scores - physical and mental functioning component scores - using a method based on factor analysis. They are considered to be conceptually distinct measures of physical (SF-36: PCS) and mental well-being (SF-36: MCS)." (p.5)
We have also included a correlation between CASP-19 and SF-36 physical well-being:
The correlation between CASP-19 and SF-36 physical well-being was r=0.39 (p≤0.001). (p.6)
2) Statistical analysis: How were statistical significance levels defined? Is it 'Statistical significance levels were set at P <0.05 for two-sided analyses'?
We thank the reviewer for pointing out this omission. Statistical significance levels were set at P<0.05 for two-sided analyses as suggested and this sentence has been added to the text. (p. 7)
3) Results: 1st paragraph, it would be helpful that the authors describe the information on the percentages of participants with short and long sleep duration and chronic insomnia symptoms (two major exposure variables) in this study population.
We have now added information on the percentages of recurrent short and long sleep duration and chronic insomnia symptoms in the first paragraph of the results section. (p.9)
Table 1
4) The title is "Characteristics of participants by sleep duration 2007-2009 (N=4,491)". In the footnote, the authors provided the percentages with sample sizes of chronic insomnia symptoms under 4 subgroups: No occurrence, 1 occurrence, 2 occurrences, 3 occurrences. It would be helpful if the

authors could provide detailed information about the 3 time points (1997-1999, 2003-2004, and 2007-2009) used to define high levels of chronic insomnia symptoms.

We have amended the footnote of Table 1 to include further information about the 3 time points used to define occurrences of chronic insomnia symptoms:
c Number of times (1997-1999, 2003-2004, and 2007-2009) high level of insomnia symptoms reported. (p.8)
5) The authors may want to revise the first column by placing mean (SD) or placing % (N) after each characteristic, e.g., Age (years), mean (SD); Women, % (N)
We thank the reviewer for suggesting this change which we agree makes the table clearer, we have altered the order of the symbols in the column, adding them after the variable name as suggested. (p.8)
6) On page 9, line 38, 'Table 2 shows the results for recurrent…' could be considered as a separate paragraph.
This revision has been added and now line 38 is the beginning of a separate paragraph. (p.9)
Table 2
7) It would be helpful that in the footnote, the authors indicate that overall well-being was measured by CASP-19, while physical well-being and mental well-being were assessed by SF-36.
We have now added the following footnote to Table 2, which clarifies which measure assessed which aspect of well-being:
b Overall well-being (CASP-19); c Physical well-being (SF-36); d Mental well-being (SF-36). (p.11)
8) Results for the trend tests could be added to Table 2.
We accept that adding these results to Table 2 would be interesting. However, since Table 2 is already quite large we didn't want to add another column and because of the adjustment of age and sex in the first column (which aren't in the test for trend) we didn't want to add the results to this column. Therefore we have added these results as a footnote to Table 2.(p.11)
9) It is interesting that the sample sizes for Recurrent short sleep subgroups (no short sleep N=2,842); one occurrence n=782; two occurrences n=499; and three occurrences n=368) are exactly the same as those for Chronic insomnia symptoms. Please double check.
We thank the reviewer for pointing out this error in the table. The sample sizes for the short sleep subgroups in this table duplicated the chronic insomnia subgroups in error. This has now been corrected. (P.11)

1. Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol 1998;51(11):903-12