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)	Sleep duration and mortality in the elderly: a systematic review with meta-analysis
AUTHORS	Silva, Andressa; Mello, Renato; Shaan, Camila; Fuchs, Flavio; Redline, Susan; Fuchs, Sandra

VERSION 1 - REVIEW

REVIEWER	Yoon Kong Loke Norwich Medical School, University of East Anglia, UK
REVIEW RETURNED	29-Apr-2015

GENERAL COMMENTS	Major comments: 1) The search seems to be at least two years out of date. It needs to be made more current. 2) The method of assessing study quality or validity needs to be stated, and the judgments should be explicitly presented. At present, I have no idea whether the studies were high, low, or of uncertain quality. 3) Although there are attempts to assess smal-study bias through funnel plot and 'trim and fill', there is no mention of possibility of selective outcome and analysis bias. By this, I mean that the authors may have specifically chosen certain cut-off points or categories for sleep duration so that significant results are found. Or they only reported outcomes that are statistically significant amongst the multitude of numerous outcomes they analysed. Although this is hard to detect, you should at least discuss the possibility. Minor comments: 1) Studies often recruite patients of diverse ages. Did your inclusion criteria focus on mean age, or median age in the study? It seems odd that you excluded studies that enrolled patients age > 60 years, when perhaps those studies might only have a few young patients and they could have large numbers of elderly? 2) I cannot understand Table 3. How did you do multivariate analysis in the meta-analysis? There is no established method for this, unless you had individual patient data that allowed you to re-calculate the risk ratios according to individual variables.

REVIEWER	Michelle Miller University of Warwick UK
REVIEW RETURNED	03-Aug-2015

GENERAL COMMENTS	General comments: This is a well conducted important and informative study. I have some minor comments (see below)

	Specific comments: Re PRISMA 2009 checklist. The page number - given in the last (reported on page #) column, is often incorrect. For example, the structured abstract is on Page 2 not page 3, and study selection point 9 is on pages 2 and 4, not 3 and 4. There are many other examples but it is not clear if this is a result of the journal formatting or would indeed reflect the final page numbering in a published article. Please can this be checked and addressed. Methods. What was the rationale for choosing age 60 as the cut off age? Please explain. Methods page 5 refer to the assessment of agreement at the 'pilot phase'. It is not clear what constituted the pilot phase and this is not reported on, with kappa statistics, in the results. Results. It is not clear how many studies, which also included participants younger than 60 years, were not included, as the corresponding author was not available / did not supply the information required. Please supply details. Was data on sleep quality available in any of the studies? With regards to Table 1, it would be useful to state to which database the search strategy pertains. Re Figure 3: Both Mallon cohorts (men and women) have wide confidence intervals compared with the other studies. Apart form the relatively smaller sample size, were there any reasons for this difference? What happens to the estimate if both cohorts are removed? Discussion: There is a robust discussion of the analysis and possible mechanisms underlying the observations. The limitations of the study have been well described.

REVIEWER	Bo Xi Shandong University
REVIEW RETURNED	08-Aug-2015

GENERAL COMMENTS	The article is well written but I have a few major concerns. First, I note that this review was updated to 2013. The search is not sufficient. Please update the included studies up to August, 2015. Second, the categories of sleep duration for each included study are much different. The dose-response analysis should be performed. Please refer to (Shan Z1, Ma H1, Xie M1, Yan P1, Guo Y2, Bao W1, Rong Y1, Jackson CL3, Hu FB4, Liu L5.Sleep duration and risk of type 2 diabetes: a meta-analysis of prospective studies.Diabetes Care. 2015 Mar;38(3):529-37.) Third, the meta analysis by Cappuccio FP et al. (Sleep. 2010 May;33(5):585-92.) reported that both long and short sleep duration are associated with risk of all cause mortaltiy in both young (<60 y) and old (>60 y) populations. Please compare your findings with them

	and explain the reasons. Fourth, does the sleep duration include daytime napping? Recent meta-analyses demonstrated that daytime napping is the risk factor for mortality (Sleep Med. 2015 Jul;16(7):811-819. Med Sci Monit. 2015 May 4;21:1269-75.). Fifth, another limitation is that the sleep quality were not assessed and this point might also be the limitaiton of the included original articles.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

We would like to thank to Dr. Loke for contributing to improve our manuscript. We hope that the revised version fulfill your requirements.

Reviewer Name Yoon Kong Loke
Institution and Country Norwich Medical School, University of East Anglia, UK
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below
Thank you for giving me the opportunity to comment on this interesting work.

Major comments:

1) The search seems to be at least two years out of date. It needs to be made more current.

Response to the reviewer:

We redid the literature search and contacted authors in order to provide information for the elderly population. We were able to incorporated eight studies.

Changes in the manuscript:

Please, see our new flowchart with the 27 studies and forest plots with all studies.

2) The method of assessing study quality or validity needs to be stated, and the judgments should be explicitly presented. At present, I have no idea whether the studies were high, low, or of uncertain quality.

Response to the reviewer:

The editor also emphasize this topic and we repeated here the answer:
In our meta-analysis, two independent assessors appraised all studies for internal validity, using the STROBE statement. The NIH provided an instrument for Quality Assessment of Systematic Reviews and Meta-Analyses, but it does not have a tool for rating quality of the studies, and there is no one commonly accepted (http://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/sr_ma). In the past, we used the Newcastle Ottawa Scale for assessing the quality of observational studies. However, the instrument was criticized and is no longer used. Therefore, a qualitative assessment of its items was held: ascertainment of exposure, demonstration that outcome of interest was not present at start of study (which is obvious since the outcome is death), comparability of cohorts on the basis of the design or analysis, assessment of outcome by

independent blind assessment (for CVD deaths), follow-up long enough for outcomes to occur, and adequacy of follow up of cohorts. We found that all studies provided enough information to ensure the quality of reporting.

Change in the manuscript:

We expanded a paragraph describing the Study selection and data extraction:

The assessors appraised all studies for internal validity, using the STROBE statement and a qualitative assessment of its items was held: ascertainment of exposure, demonstration that outcome of interest was not present at start of study (which is obvious since the outcome is death), comparability of cohorts on the basis of the design or analysis, assessment of outcome by independent blind assessment (for CVD deaths), follow-up long enough for outcomes to occur, and adequacy of follow up of cohorts. We found that all studies provided enough information to ensure the quality of reporting.

3) Although there are attempts to assess small study bias through funnel plot and 'trim and fill', there is no mention of possibility of selective outcome and analysis bias. By this, I mean that the authors may have specifically chosen certain cutoff points or categories for sleep duration so that significant results are found. Or they only reported outcomes that are statistically significant amongst the multitude of numerous outcomes they analyzed. Although this is hard to detect, you should at least discuss the possibility.

Response to the reviewer:

As you can see, there was several cutoffs for sleep duration and we did not have a hypothesis about a specific cutoff for short or long sleep duration. We decided a priori analyze extreme cutoffs for a long and short sleep, where several cuts were available, keeping the reference category chosen by the authors of individual articles. Statistical significance did not drive our decisions. Instead of testing different cutoffs, we decided to perform a subgroup analysis using all available cutoffs, shown in Figure 4. For short sleep duration, we found that ≤5, ≤6 and ≤7 hours were not associated with mortality, in comparison to long sleep duration, while ≥8, ≥9, and ≥10 hours of sleep duration were associated with mortality.

Change in the manuscript:

In order to describe what we did, we added a paragraph in the discussion.

Although there are several cutoffs for the duration of sleep, we did not have a hypothesis a priori on a specific cutoff for short or long sleep. The decisions taken before starting the analysis were to analyze the extreme cutoffs for long and short sleep, where several cutoffs points were available, keeping the reference category as presented in the individual articles. Recognizing the diversity of cutoffs, subgroup analyses were performed using all available cutoffs points. All cutoffs for long sleep duration were associated with mortality, but there was no association with the cutoffs for short sleep duration.

Minor comments:

1) Studies often recruite patients of diverse ages. Did your inclusion criteria focus on mean age, or median age in the study? It seems odd that you excluded studies that enrolled patients age > 60 years, when perhaps those studies might only have a few young patients and they could have large

numbers of elderly?

Response to the reviewer:

We selected articles that enrolled participants aged 60 years or over. We did not use an average age of 60 years as a criterion. In studies with a wider age range, we tried to obtain data for elderly population. However, some studies did not describe the age distribution and we were not able to detect how many non-elderly participants were enrolled. Therefore, we have standardized the age criterion to select individual studies.

2) I cannot understand Table 3. How did you do multivariate analysis in the meta-analysis? There is no established method for this, unless you had individual patient data that allowed you to recalculate the risk ratios according to individual variables.

Response to the reviewer:

We agree with the reviewer that the title of Table 3 is not clear. Therefore, we modified the title of Table 3. Table 3 shows the analysis of subgroups according to the models of multivariate analysis carried out for control of confounding factors in the individual studies. The summary estimates were calculated using the results of each set of multivariate analysis of individual studies using the random effects models.

Change in the manuscript:

Table 3. Analysis of subgroups for the association between sleep duration and all-cause mortality according to the models of multivariate analysis carried out for control of confounding factors in the individual studies.

Reviewer: 2

We would like to thank Dr. Miller for taking time in the revision of our manuscript. We hope that the revised version has reached your demands.

Reviewer Name Michelle Miller
Institution and Country University of Warwick
UK
Please state any competing interests or state ‘None declared’: None declared
Please leave your comments for the authors below

1.General comments:
This is a well conducted important and informative study. I have some minor comments (see below)

2.Specific comments:

a) Re PRISMA 2009 checklist.
The page number given in the last (reported on page #) column, is often incorrect. For example, the

structured abstract is on Page 2 not page 3, and study selection point 9 is on pages 2 and 4, not 3 and 4. There are many other examples but it is not clear if this is a result of the journal formatting or would indeed reflect the final page numbering in a published article. Please can this be checked and addressed.

Response to the reviewer:

Thank you for noticing that. We corrected the pages.

b) Methods.
What was the rationale for choosing age 60 as the cut off age? Please explain.

Response to the reviewer:

In Brazil, individual aged 60 years or over are identified as elderly. Therefore, we used this age as a criterion, but most studies enrolled participants above this age.

c) Methods page 5 refer to the assessment of agreement at the 'pilot phase'. It is not clear what constituted the pilot phase and this is not reported on, with kappa statistics, in the results.

Response to the reviewer:

A pilot phase was carried out before the starting the meta-analysis and its purpose was to training and standardize the assessment and data extraction among researchers. This phase was conducted using articles that were not included in the meta-analysis. Therefore, we let the information out of the manuscript.

d) Results.
It is not clear how many studies, which also included participants younger than 60 years, were not included, as the corresponding author was not available / did not supply the information required. Please supply details.

Response to the reviewer:

The flowchart shows that 40 studies enrolled only participants younger than 60 years, 7 studies reported results for a wider age range or described sleep hours as a continuous variable and the authors did not send additional results as requested. However, 5 corresponding authors positively respond to our requests.

Change in the manuscript:

We updated the results:

Authors of studies, which also included participants younger than 60 years, whose corresponding author was available, were contacted to request a reanalysis of data, restricted to elderly individuals. Five authors29-33 did a reanalysis and sent additional data. Those studies were included in the meta-analysis.

e) Was data on sleep quality available in any of the studies?

Response to the reviewer:

There was a diversity of additional sleep-related variables, including use of hypnotics, sleep complains, sleep disturbances, and sleep quality. However, there was no common information in order to use this information.

f) With regards to Table 1, it would be useful to state to which database the search strategy pertains.

Response to the reviewer:

The reported strategy of search was adopted for Meadline/Pubmed and it was described in the title. For EMBASE, a similar strategy was used (described as a note at the bottom of the table).

Change in the manuscript:

Table 1. Search strategy used to locate articles in the database of Meadline/PubMed*
* Similar strategies were used to perform the search in the EMBASE

g) Re Figure 3:
Both Mallon cohorts (men and women) have wide confidence intervals compared with the other studies. Apart from the relatively smaller sample size, were there any reasons for this difference? What happens to the estimate if both cohorts are removed?

Response to the reviewer:

The overall characteristics were similar to those of other studies. However, the investigators assessed the association between sleep duration and mortality in a population-based survey who responded to the questionnaire sent by mail, which represent only 70% of the eligible population.

We removed the data for men and women and the results did not change for short sleep duration, but the risk of long sleep duration on mortality was reduced to RR:1.22 (95%CI: 1.18-1.26). Please see the forest plots for long and short sleep duration below.

Short sleep duration

Long sleep duration

Change in the manuscript:

In order to highlight these findings, we included this information in the results.

Long sleep duration and mortality:

However, the removal of the study performed by Mallon et al. reduced the risk of long sleep duration

on mortality to RR:1.22 (95%CI: 1.18-1.26).

Short sleep duration and mortality:

Pooled analysis remained statistically significant after excluding Kronholm et al. study,29 which had the largest influence (relative weight: 31.6%) in the overall pooled estimate, as well as with the exclusion of Mallon et al. study.30

h) Discussion:
There is a robust discussion of the analysis and possible mechanisms underlying the observations. The limitations of the study have been well described.

Response to the reviewer: Thank you.

Reviewer: 3

Reviewer Name Bo Xi
Institution and Country Shandong University
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below

The article is well written but I have a few major concerns.
First, I note that this review was updated to 2013. The search is not sufficient. Please update the included studies up to August, 2015.

Response to the reviewer: Thank you for the compliment. We updated the review to August, 2015.

Second, the categories of sleep duration for each included study are much different. The dose response analysis should be performed. Please refer to (Shan Z1, Ma H1, Xie M1, Yan P1, Guo Y2, Bao W1, Rong Y1, Jackson CL3, Hu FB4, Liu L5. Sleep duration and risk of type 2 diabetes: a meta-analysis
of prospective studies. Diabetes Care. 2015 Mar;38(3):52937.)

Response to the reviewer:

We appreciate the reviewer's suggestion to conduct a dose-response analysis. However, our main aim was to assess the risk for short and long sleep duration. We have developed a protocol to extract the data and have been working to offer the best evidence to answer these questions. The suggestion of the reviewer is a completely different approach that can be conducted in another project, using another protocol and a different background.

In this study, the categories of sleep duration were agglutinated to allow performing the planned analysis. Such analysis provided fair results.

Third, the meta analysis by Cappuccio FP et al. (Sleep. 2010 May;33(5):58592.)
reported that both long and short sleep duration are associated with risk of all cause mortaltiy in both young (<60 y) and old (>60 y) populations. Please compare your findings with them and explain the

reasons.

Response to the reviewer:

This reference was incorporated in the current version of the Ms. and the results were discussed. Please see the paragraph below.

Change in the manuscript:

A meta-analysis conducted among adults and elderly population showed an increased risk of dying for both, short and long sleep duration. For short sleep duration, the magnitude of the effect was greater than the observed risk in this meta-analysis, but for long sleep duration the risk was similar.21

Fourth, does the sleep duration include daytime napping? Recent meta-analyses demonstrated that daytime napping is the risk factor for mortality (Sleep Med. 2015 Jul;16(7):811819. Med Sci Monit. 2015 May 4;21:126975.).

Response to the reviewer:

Long sleep duration did not include napping, but daytime did. We observe that two meta-analysis were published recently, one mentioned by the reviewer and the other led by Zhong et al. (Zhong G, Wang Y, Tao T, Ying J, Zhao Y. Daytime napping and mortality from all causes, cardiovascular disease, and cancer: a meta-analysis of prospective cohort studies. Sleep Med. 2015 Jul;16(7):811-9.). In both cases, statistically significant associations were found with all-cause mortality, but in a limited number of studies, aspect highlighted by the authors.

Change in the manuscript:

In two meta-analysis published recently, statistically significant associations were found with all-cause mortality, but in a limited number of studies.80,81

Fifth, another limitation is that the sleep quality were not assessed and this point might also be the limitation of the included original articles.

Response to the reviewer:

As another reviewer pointed out, quality assessment should be reported with sleep duration. However, there is a diversity of additional sleep-related variables, including use of hypnotics, sleep complains, sleep disturbances, in addition to sleep quality. However, there is no common information that enables the use of such information, except in the context of a randomized clinical trial.

Change in the manuscript:

We included this aspect as a limitation of the meta-analysis, carried out from the original studies which did assessed this aspect of sleep duration.

Another limitation of this study is the lack of assessment of sleep quality, which can interfere with sleep duration. However, this is a limitation of the included original articles and remains a hypothesis to be tested in future research.

VERSION 2 – REVIEW

REVIEWER	Yoon Kong LOKE University of East Anglia UK
REVIEW RETURNED	13-Oct-2015

GENERAL COMMENTS	I appreciate that the authors have not used a specific quality assessment tool. Yet, in the Discussion section, the authors point out Limitations in the primary studies, such as confounding, poor ascertainment of sleep duration or quality etc. These aspects should have been placed in the Results section, under a subheading of Risk of Bias. I suggest that the authors should actually have a table with 3 explicit columns - ascertainment of sleep, ascertainment of mortality, adjustment for confounders (this is currently in Table 1) - that describe the methods used in each study. This then allows appropriate text in the Risk of Bias section e.g. to say that X studies did not adjust for confounders, Y only adjusted for age and gender etc. Readers can then get an idea of the major weaknesses within this set of included studies.

REVIEWER	Bo Xi Shandong University, China
REVIEW RETURNED	24-Oct-2015

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.

VERSION 2 – AUTHOR RESPONSE

We would like to thank again to Dr. Loke for contributing to improve our manuscript. It is considerably better written.

Please leave your comments for the authors below
I appreciate that the authors have not used a specific quality assessment tool. Yet, in the Discussion section, the authors point out Limitations in the primary studies, such as confounding, poor ascertainment of sleep duration or quality etc.

These aspects should have been placed in the Results section, under a subheading of Risk of Bias.
I suggest that the authors should actually have a table with 3 explicit columns - ascertainment of sleep, ascertainment of mortality, adjustment for confounders (this is currently in Table 1) - that describe the methods used in each study. This then allows appropriate text in the Risk of Bias section e.g. to say that X studies did not adjust for confounders, Y only adjusted for age and gender etc.
Readers can then get an idea of the major weaknesses within this set of included studies.

Response to the reviewer:

We included a new Table 3 with ascertainment of sleep, ascertainment of mortality, and adjustment

for confounders.

It was also included a new subheading in the results section - Risk of Bias - which contains part of the text that previously was in the subheading limitation of the study. We agree that now readers can get an idea of the major weaknesses of included studies.

VERSION 3 - REVIEW

REVIEWER	Bo Xi Shandong University
REVIEW RETURNED	14-Dec-2015

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.