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) | Protocol for the National Nurse Health Study (NNHS): A Web-based Ambispective Cohort Study |
|---------------------|-----------------------------------------------------------------------------------------|
| AUTHORS             | Zhuo, Lin; Zhang, Heli; Geng, Rongmei; Wang, Panfeng; Zeng, Lin; Che, Ying; Wang, Peng; Li, Pengfei; Huang, Tao; Li, Baohua; Zhan, Siyan |

VERSION 1 – REVIEW

| REVIEWER          | Ramalho, André |
|-------------------|----------------|
| University of Porto, Department of Community Medicine, Information and Health Decision |
| REVIEW RETURNED   | 21-Mar-2021    |

GENERAL COMMENTS

Dear Editor and authors,

First, I would like to thank you for the opportunity to review such an interesting protocol regarding early-life events, daily habits, occupational and environmental risk factors, and health outcomes of a specific and vital subset of healthcare professionals.

I have a few suggestions that authors should improve in the protocol:

The authors refer that the study's generalization may be limited due to the small sample size in terms of limitations. As we know, the sample size will determine the validity in terms of the statistical power necessary to reject or approve the hypothesis. Adequate sample size will make it easier to avoid random errors in the study results. The protocol does not make it clear how they intend to calculate/estimate the sample size.

Recruitment and follow-up make me a little concerned, especially as the authors intend to minimize possible selective survival bias. As we are well aware, this situation can occur when exposed individuals have had the condition for some time (prevalent cohort), so the probability of expressing the outcome is more significant than those recently exposed (incident cohort). This effect is known as left truncation or time-related bias (immortal time bias, time interval bias).

Another point to be clarified in the protocol is how the authors will deal with the possible missing data. There may be missing data due to the record's poor quality or due to variables that were not...
considered to be recorded in advance. In both cases, the source of the missing information can lead to information bias.

I had some doubts about the analyses to be made, namely if there will be any subgroups (by sex, age, the existence of advanced nursing roles, standard working hours, ratio nurses x beds, etc.). If such sub-analyzes are foreseen, they must be included in a protocol. I suggest that if they are not, that the protocol bring this information with due justification. Interestingly, the authors should consult the National Health Workforce Accounts Handbook by WHO. This handbook can be helpful in some analyzes.

Best Regards,

REVIEWER
Atnafu, Desta
Bahir Dar University, Health System Management& Health Economics

REVIEW RETURNED
24-Mar-2021

GENERAL COMMENTS
1. I think the background in the abstract should display (summarize) at least four conditions of the whole manuscript content: What is known globally about the diseases, injuries, and impairments…; what is unknown about Nurses’ physical and mental health, medical care group and quality of medical services in globally…; the major gaps that this study was focused, and finally the aims of conducting this research. Thus, in my view the first two conditions are not yet expressed.
2. Is there any kind of sampling in taking Nurses’ physical and mental health, medical care group and quality of medical services among countries studied…?
3. The data collection procedures were not explained, since it is a quantitative data, the sample size needs to be clarified in the study in the case of data extracted using longitudinal survey.
4. Even though I could not see the template for journal, objectives and related has to be put on as part of introduction section, not methods section.

The reviewer provided a marked copy with additional comments. Please contact the publisher for full details.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: André Ramalho
First, I would like to thank you for the opportunity to review such an interesting protocol regarding early-life events, daily habits, occupational and environmental risk factors, and health outcomes of a specific and vital subset of healthcare professionals. I have a few suggestions that authors should improve in the protocol:

Comment 1: The authors refer that the study’s generalization may be limited due to the small sample size in terms of limitations. As we know, the sample size will determine the validity in terms of the statistical
power necessary to reject or approve the hypothesis. Adequate sample size will make it easier to avoid random errors in the study results. The protocol does not make it clear how they intend to calculate/estimate the sample size.

Authors' response: We acknowledge the reviewer's comment. We missed in our previous manuscript indeed and now have mentioned the sample size calculation in the 'Participants and study site' of 'Method.' Considering the situation that there are no many studies concerning nurses' health situation, especially mental health, we estimated the sample size quoting an investigation on nurses in 136 hospitals in China (Shi-chao W, et al. Health Economics Research 2019;8:62-65 ). A sample size of 401 produces a two-sided 95% confidence interval with a width equal to 0.100 when the sample proportion is 0.525 (line 152-154, page 7).

Comment 2: Recruitment and follow-up make me a little concerned, especially as the authors intend to minimize possible selective survival bias. As we are well aware, this situation can occur when exposed individuals have had the condition for some time (prevalent cohort), so the probability of expressing the outcome is more significant than those recently exposed (incident cohort). This effect is known as left truncation or time-related bias (immortal time bias, time interval bias).

Authors' response: We acknowledge the reviewer's comment. We planned two-step research: understand the baseline situation (including mental health) of the Peking University Third Hospital nurses and then try to figure out the risk factors for the diseases we are interested in using a subgroup of participants. We think the immortal time bias can be controlled in the study design and the analysis period, such as using matching or taking exposure as a time-varying covariate in the analysis stage and then incorporating it into the model analysis (line 242-244, page 10).

Comment 3: Another point to be clarified in the protocol is how the authors will deal with the possible missing data. There may be missing data due to the record's poor quality or due to variables that were not considered to be recorded in advance. In both cases, the source of the missing information can lead to information bias.

Authors' response: We acknowledge the reviewer's comment. About missing data, we decided to utilize the concept of 'minimal data set' based on the previous experience of the Peking University Third Hospital HIS system data. In our previous study, we extracted the health check data of physicians and found that the data performed well in terms of blood routine tests and biochemical tests. In the present study, we focus more on the mental health among nurses, and the workload data could be traced accurately. Besides, a sensitivity analysis will be done under different circumstances, such as deleting records with over 20% of missing data. It could result in a loss of sample size. However, it can still help us minimize the influence of the missing data capturing false positives in our study (line 248-250, page 10).

Comment 4: I had some doubts about the analyses to be made, namely if there will be any subgroups (by sex, age, the existence of advanced nursing roles, standard working hours, ratio nurses x beds, etc.). If such sub-analyses are foreseen, they must be included in a protocol. I suggest that if they are not, that the protocol bring this information with due justification. Interestingly, the authors should consult the National Health Workforce Accounts Handbook by WHO. This handbook can be helpful in some analyzes.

Authors' response: We acknowledge the reviewer's comment and have now mentioned the potential subgroup analyses in the 'Method.' We did careful research on the handbook the reviewer recommended, we decided to do some subgroup analysis to discover the differences between age groups, departments, the existence of advanced nursing roles, and working hours (line 236-238, page 10).

Reviewer 2: Desta Atnafu
Comment 1: I think the background in the abstract should display (summarize) at least four conditions of the whole manuscript content: What is known globally about the diseases, injuries, and impairments…; what is unknown about Nurses’ physical and mental health, medical care group and quality of medical services in globally…; the major gaps that this study was focused, and finally the aims of conducting this research. Thus, in my view the first two conditions are not yet expressed.
Authors’ response: We thank the reviewer for this comment and have modified the introduction part of the abstract.

Comment 2: Is there any kind of sampling in taking Nurses’ physical and mental health, medical care group and quality of medical services among countries studied…?
Authors’ response: We thank the reviewer for this comment. To our knowledge, there is no national or administrative level of large-scale sampling for nurses’ physical and mental health tracking in China currently.

Comment 3: The data collection procedures were not explained, since it is a quantitative data, the sample size needs to be clarified in the study in the case of data extracted using longitudinal survey.
Authors’ response: We thank the reviewer for this comment. We now have added the sample size calculation in the 'Participants and study site' of 'Method.' We estimated the sample size quoting an investigation on nurses in 136 hospitals in China (Shi-chao W, et al. Health Economics Research 2019:8:62-65 ). A sample size of 401 produces a two-sided 95% confidence interval with a width equal to 0.100 when the sample proportion is 0.525 (line 151-153, page 6). Moreover, according to a pilot study, we are sure that the sample we recruit can fulfill the study purpose.

Comment 4: Even though I could not see the template for journal, objectives and related has to be put on as part of introduction section, not methods section.
Authors’ response: Thank you – we have now moved the 'Objective' part right after the 'Introduction' (line 131-140, page 6).

VERSION 2 – REVIEW

| REVIEWER          | Ramalho, André  |
|-------------------|-----------------|
|                   | University of Porto, Department of Community Medicine, Information and Health Decision |
| REVIEW RETURNED   | 05-Jul-2021     |

GENERAL COMMENTS

Dear authors and editor,

My considerations were all considered. However, there is still a need for minor language corrections (English). At some points in the article, there are some hard-to-read sentences, fragmented sentences and minor spelling errors. I suggest that the reference suggested above and used by you is also referenced (NHWA - WHO).

I wish you success with the application of this protocol.

Best,

VERSION 2 – AUTHOR RESPONSE

Reviewer 1: André Ramalho

Comment 1: My considerations were all considered.
However, there is still a need for minor language corrections (English). At some points in the article, there are some hard-to-read sentences, fragmented sentences and minor spelling errors.

I suggest that the reference suggested above and used by you is also referenced (NHWA - WHO).

Authors’ response: We thank the reviewer again for the manuscript review and acknowledge the reviewer’s thoughtful comment on the first revision. We have obtained editorial assistance from a professional editor to improve the English language.