PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The associations between shift work characteristics, shift work schedules, sleep and burnout in North American police officers: a cross-sectional study
AUTHORS	Peterson, Scott; Wolkow, Alexander; Lockley, Steven; O'Brien, Conor; Qadri, Salim; Sullivan, Jason; Czeisler, Charles; Rajaratnam, Shantha M. W.; Barger, Laura

VERSION 1 – REVIEW

REVIEWER	Jonathan Houdmont University of Nottingham, UK
REVIEW RETURNED	03-Apr-2019

| GENERAL COMMENTS | This manuscript reports on cross-sectional associations between shift work characteristics, shift work schedules, sleep, and burnout in a large sample of more than 3,000 US police officers. In my view the manuscript is written to a high academic standard and utilises appropriate measurement instruments and statistical analyses, which lead to appropriate interpretations of the data. Owing to the high standard of the manuscript I have only a small number of comments and recommendations.

1. Given the cross sectional, rather than longitudinal, nature of the investigation please avoid use of the word 'impact' in the title and introduction and 'effect' in the introduction. It would be more appropriate to refer to 'relationships' or 'associations' throughout.

2. Please clarify the basis on which long work hours were defined as ≥48 hours per week. I know that the UK policing research on relations between long work hours and burnout that you have cited set the threshold slightly higher at ≥49 hours per week on the basis that this is the definition applied in the European Working Time Directive. (Obviously EU legislation does not apply in the US.) Please indicate the basis on which ≥48 hours per week was selected.

3. I was a little confused about the dichotomisation of PA scores. The manuscript states "PA was defined as either high-moderate or low (≥39 points)." Isn't it the case that the PA dimension of the MBI is scored so that low scores indicate low PA? I.e., this dimension is scored in the opposite direction to the EE and DP dimensions.

4. I was surprised to see the p value set at p<.10 for the identification of covariates that were subsequently controlled for in the regression models. Please justify the decision to set the threshold at .10 rather than the more conventional .05 level. |
| --- | --- |

	5. Table 1: on the row for age there appears to be a letter 'y' that needs deleting.

REVIEWER	Philip Tucker Swansea University, United Kingdom
REVIEW RETURNED	09-Apr-2019

GENERAL COMMENTS	A large cross sectional survey of police officers is reported, examining associations between work schedule and burnout. Among the main findings, long work shifts and overtime are associated with aspects of burnout and overall burnout scores, as were non-fixed shifts (especially irregular shifts), short sleeps and high levels of sleepiness. This is a quite interesting study that makes a useful contribution to the literature. Studies of shiftwork and mental health are less common than studies of physical health, and the findings tend to be quite mixed; all the more so with respect to studies of burnout. So it is good to see a well reported and well conducted study in this area that uses appropriate measures, and is based on a quite large and relatively homogeneous sample. The key weaknesses, as the authors acknowledge, are the cross-sectional design and its reliance of self-report measures. In particular, the measure of work schedule is somewhat unsatisfactory. While there is some differentiation between different forms of work schedule, the nature of the schedules remain somewhat underspecified, which limits the interpretability of some of the findings. The introduction provides a comprehensive, reasonably well structured over view of the relevant literature. The rationale for the current study is reasonably well argued. The methods and results are mostly well-described. From my understanding of what is reported, each aspect of work schedule was examined separately and without taking into account the other aspects of work schedule. This could be a problem if there were correlations between the work schedule variables. For example, if working many long work shifts was associated with, say, working frequent night shifts, or working irregular shifts. This possibility for confounding between work schedule variables needs to be addressed / clarified. For example, it might be useful to report the correlations between the work schedule variables. There is a minor issue with how the PA dimension of burnout is described. "PA" is first introduced as "lack of personal accomplishment" – thus a high score should be 'bad'. In the sub-section measures, it is unclear whether the cut off of ≥39 points refers to the classification "high-moderate". In the results section, the text refers to "low PA" – by the original definition in the introduction, that implies a low lack of PA – a 'good' result. I am pretty sure that having "low PA" means being more likely to 'lack a sense of personal accomplishment' (i.e. a bad thing), but care is needed in describing this concept to avoid confusion. Confusion is compounded by the fact that PA ends up showing some negative results (i.e. in the opposite direction to other trends, and to what might be expected). Overall, the discussion makes a good attempt to explain quite a complex set of findings with some well argued and detailed analyses of the observed trends. I particularly liked the discussions

of the distinction between long shifts and long weekly hours, the distinction between voluntary and mandatory overtime and the interesting speculation on the occupational differences in factors promoting DP among nightworkers. The discussion of fixed and rotating shifts highlights a weakness of this study that the nature of the shift schedules were underspecified, making it hard to interpret the comparisons between types of shift schedule. In particular, the definition of 'fixed shifts' (the 'baseline' group for schedule comparisons) does not differentiate between day and night work – a potentially important aspect which could have substantial bearing on the interpretations of the comparisons between schedule types.

Minor points
Page 4, 2nd paragraph. The statement about which shifts police have traditionally worked probably needs to be qualified as referring to police in the USA?
Discussion 1st para. Insertion needed – "Working night shifts *nearly every day *was associated with DP".

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name
Jonathan Houdmont

Institution and Country
University of Nottingham, UK

Please state any competing interests or state 'None declared':
None declared.

Please leave your comments for the authors below
This manuscript reports on cross-sectional associations between shift work characteristics, shift work schedules, sleep, and burnout in a large sample of more than 3,000 US police officers. In my view the manuscript is written to a high academic standard and utilises appropriate measurement instruments and statistical analyses, which lead to appropriate interpretations of the data. Owing to the high standard of the manuscript I have only a small number of comments and recommendations.

1. Given the cross sectional, rather than longitudinal, nature of the investigation please avoid use of the word 'impact' in the title and introduction and 'effect' in the introduction. It would be more appropriate to refer to 'relationships' or 'associations' throughout.
Response: Thank you for this suggestion. The title has now been revised to use the word 'associations', rather than 'impact'. The use of 'impact' and 'effect' has also been replaced with 'associations' or 'relationships' in the Introduction, as well as elsewhere in the manuscript where appropriate.

2. Please clarify the basis on which long work hours were defined as ≥48 hours per week. I know that the UK policing research on relations between long work hours and burnout that you have cited set the threshold slightly higher at ≥49 hours per week on the basis that this is the definition applied in the

European Working Time Directive. (Obviously EU legislation does not apply in the US.) Please indicate the basis on which ≥48 hours per week was selected.

Response: Thank you for alerting us to this important missing detail. We would like to clarify that weekly work hours were based on 1) the national estimates of work organisation characteristics among workers in the US which defines long work weeks as 48 hours or more, and 2) prior literature which has found close associations between weekly work hours of 48 hours or more and burnout. By applying this definition of weekly work hours in the current study, we are able to compare our data with prior benchmarking research, and interpret our burnout findings in relation to other work and burnout literature. Paragraph 4 of the Methods section has now been updated to include the below text clarifying the basis on which we defined long work weeks.

"This categorisation was based on prior benchmarking literature in the US which defines long work weeks as 48 or more hours per week,21 and research which has shown this cut-off to be associated with burnout.22"

In addition, we have also revised the below sentence in Paragraph 2 of the Discussion section to highlight the previous research which has shown a relationship between working 48 or more hours per week and burnout, as well as specify that the prior burnout research in UK police was based on a slightly different categorisation of long weekly work hours (i.e., ≥49 hours).

"Working ≥48 hours per week was associated with reduced odds of low PA in police, which contrasts research in physicians using a similar definition of weekly work hours,22 as well as a prior study in UK police that defined long work weeks using a slightly higher threshold (≥49 hours per week).27"

While making the above revisions, we also noticed that the incorrect categories were originally reported in the Methods for the night shift frequency variable. This minor error has now been corrected in Paragraph 3 of the Methods section.

3. I was a little confused about the dichotomisation of PA scores. The manuscript states "PA was defined as either high-moderate or low (≥39 points)." Isn't it the case that the PA dimension of the MBI is scored so that low scores indicate low PA? I.e., this dimension is scored in the opposite direction to the EE and DP dimensions.

Response: Thank you for this comment, as it has alerted us to some ambiguity in the MBI scoring manual. Specifically, on Page 5 – Table 1 of the manual (Maslach et al.1996) we accidently mixed up our interpretation of the wording, thinking that "low" (≥39) on the PA scale in this Table was representative of low PA, rather than low on the range of experienced burnout. After re-reading the manual we realised that low PA should have been scored as a total score ≤31, as you correctly pointed out in your comment. This error has now been corrected in Paragraph 2 of the Methods section by inserting the below sentence.

"PA was also a dichotomous variable, but it was scored in the opposite direction to EE and DP, in which low (versus high-moderate) scores indicated low PA (PA ≤31 points)"

Fixing the scoring of low PA has impacted some of the results in the study. For instance, we found 40.7% of the sample had PA scores ≤31, indicating low PA. Table 1 and Paragraph 1 of the Results have now been updated to reflect this revised finding. Furthermore, we found the percentage of participants with at least one dimension of burnout also changed slightly to 65.9% (instead of 65.5%) after fixing the scoring for low PA. This finding has also been revised in Table 1 and in Paragraph 1 of the Discussion.

After applying the revised PA scoring to the shift work characteristics and burnout analyses, we found there was no longer a relationship between long shift frequency (>0 - <2, ≥2 - <4, ≥4 long shifts per week) and low PA, but there is now a relationship between 1 to 4 night shifts per week and low PA, and long work weeks (≥48 hours) and low PA. The unadjusted OR and adjusted OR, and the 'Positive

result in police with burnout, No./No. total (%)' sections of Table 2 have been updated to reflect these changes to the findings. The below two sentences have also been included in Paragraph 2 of the revised Results section to highlight these new findings.

"Working ≥48 hours per week was associated with reduced odds of low PA (0.83, 95% CI 0.71-0.97) compared to shorter weeks."

"Working night shifts 1 to 4 times per week was associated with lower odds of low PA (0.79, 95% CI 0.64-0.98) compared to never working nights."

There were other minor changes to the reported unadjusted OR, adjusted OR and 'Positive result in police with burnout, No./No. total (%)' sections of Table 2 for the relationships between long shifts and burnout, and overtime (both mandatory and voluntary) and burnout, but these changes did not impact the significance of the original findings. Table 2 has been updated to reflect these minor changes.

The revised scoring of PA also impacted the shift schedules and burnout results. Specifically, after applying this correction, there was a significant association between working rotating shifts and low PA. The below sentence in Paragraph 3 of the Results has been revised to highlight this finding, and Table 3 has also been updated.

"Furthermore, police working rotating schedules were more likely to have low PA (1.38, 95% CI 1.15-1.64) compared to fixed shifts."

Finally, the correction of PA in the sleep and burnout analyses resulted in a significant relationship between short sleep and burnout, but the relationship between sleepiness and burnout was no longer significant after this correction. The sleep and burnout findings have been updated in Table 4. Paragraph 4 of the Results section has also been updated with the below revised sentence describing the new findings between short sleep and low PA, while details relating to the sleepiness and low PA relationship have now been deleted from this paragraph.

"Police who reported, on average, short sleep (<6 hours) were more likely to have overall burnout (1.49, 95% CI 1.21-1.81), high EE (1.60, 95% CI 1.33-1.93) and low PA (1.24, 95% CI 1.05-1.47) compared to those sleeping ≥6 hours."

In reviewing the PA findings, we were also alerted to two additional aspects of the analysis that we would like to address in this revise and resubmit. Firstly, it was noticed that the description of the statistical analysis method in the original manuscript differed slightly to how this analysis was actually completed. Specifically, the original Statistical Analysis section stated that a backwards elimination method was used to remove the covariates in the final model. While this step of the analysis was completed, it was not the final step. Instead, we used the backwards elimination method to identify the variables that were then inserted into the first step of a hierarchal logistic regression model. The below sentences have been included in the Statistical Analysis section to clarify the correct analysis steps that were applied in this manuscript.

"Each confounder was tested individually with the burnout outcome, and we only included those that were significant (p<0.10) in initial models.28 A backwards elimination method was then used to remove those that were not significant at the <0.05 level when all variables were included in a single model. Factors that remained were used in the first step of a hierarchical logistic regression model. The predictor variables were then added in the second step of the model that included the selected confounding factors. Model fit was assessed using the omnibus test and goodness-of-fit indices (-2 log likelihood)"

Secondly, it was not highlighted in the original submission that for our multi-level predictor variables (e.g., night shift frequency and shifts schedules), we had a priori hypothesis comparisons of interest. This meant that we ignored the goodness of fit results relating to whether each block in the hierarchical logistic regression analysis improved the overall model fit, and instead we only concentrated on our comparisons of interest (e.g., never working night shifts compared to working night shifts nearly every day). The Statistical Analysis section has now been updated with the below sentences describing our a priori hypotheses. The full details of the model fit have also been added to the Supplement. Please note that these revisions have not impacted the reported findings.

"Our analysis included several multilevel predictors, such as night shift frequency and shift work schedules. For night shift frequency, we hypothesised a priori that a greater frequency of night shifts would be more strongly associated with burnout compared to lower frequencies. Therefore, we compared never/nearly never working night shifts to working night shifts nearly every day and 1-4 times per week to examine if high frequency compared to low frequency of night shifts were associated with worse burnout. Furthermore, we hypothesised a priori that more variable shift work would be related to high burnout, and thus compared fixed shifts to all other categories of shift work (i.e., irregular, rotating and other) to examine whether greater variation in schedules predicted higher burnout."

In addition, using track changes, we have also made a number of minor corrections to some of the previously reported values in the Results section. For instance, in Paragraph 1 of the Results section, we have corrected the percentage of participants that reported irregular or rotating schedules to 33.6%. Furthermore, there were some minor corrections made to the Overtime and DP findings in Table 2 and Paragraph 2 of the Results section. A minor correction was also made to the p-value for the Any Overtime and Overall Burnout findings in Table 2 in the Results section. Finally, due to incorrect rounding, slight corrections were made to the values reported for the Other Shift Schedules and Overall Burnout findings in Table 3 of the Results section, and the Sleep Duration and Overall Burnout findings in Table 4 of the Results section. Please note that these minor corrections have not impacted the significance of the reported findings.

4. I was surprised to see the p value set at p<.10 for the identification of covariates that were subsequently controlled for in the regression models. Please justify the decision to set the threshold at .10 rather than the more conventional .05 level.

Response: The p-value cut-off point of p<0.10 was used for the selection of covariates in this study in the initial models to ensure we did not miss any potential predictors. Ranganathan et al. 2017 recommend that in this step of the logistic regression analysis, this more liberal cut-off should be used instead of the conventional 0.05 level because the purpose here is to identify potential predictors, instead of testing a specific hypothesis. We have now supported our decision to apply the p<0.10 cut-off for significance by adding this supporting reference to the below sentence in the revised manuscript.

"Each confounder was tested individually with the burnout outcome, and we only included those that were significant (p<0.10) in initial models.[28]"

5. Table 1: on the row for age there appears to be a letter 'y' that needs deleting.

Response: Thank you for this comment. The letter 'y' refers to the unit 'years', which should have been spelt out in full or inserted in the footnote. To correct this error, we have spelled out the word 'years' in Table 1 in the revised manuscript.

Reviewer: 2

Reviewer Name
Philip Tucker

Institution and Country
Swansea University, United Kingdom

Please state any competing interests or state 'None declared':
None declared

Please leave your comments for the authors below
A large cross sectional survey of police officers is reported, examining associations between work schedule and burnout. Among the main findings, long work shifts and overtime are associated with aspects of burnout and overall burnout scores, as were non-fixed shifts (especially irregular shifts), short sleeps and high levels of sleepiness.

This is a quite interesting study that makes a useful contribution to the literature. Studies of shiftwork and mental health are less common than studies of physical health, and the findings tend to be quite mixed; all the more so with respect to studies of burnout. So it is good to see a well reported and well conducted study in this area that uses appropriate measures, and is based on a quite large and relatively homogeneous sample. The key weaknesses, as the authors acknowledge, are the cross-sectional design and its reliance of self-report measures. In particular, the measure of work schedule is somewhat unsatisfactory. While there is some differentiation between different forms of work schedule, the nature of the schedules remain somewhat underspecified, which limits the interpretability of some of the findings.

The introduction provides a comprehensive, reasonably well structured over view of the relevant literature. The rationale for the current study is reasonably well argued.

Response: Thank you for these positive comments about the Introduction section.

The methods and results are mostly well-described. From my understanding of what is reported, each aspect of work schedule was examined separately and without taking into account the other aspects of work schedule. This could be a problem if there were correlations between the work schedule variables. For example, if working many long work shifts was associated with, say, working frequent night shifts, or working irregular shifts. This possibility for confounding between work schedule variables needs to be addressed / clarified. For example, it might be useful to report the correlations between the work schedule variables.

Response: Thank you for raising this important point. To examine the possibility of confounding, we have now performed multicollinearity diagnostics between the shift work predictor variables included in the multiple logistic regression model analyses. In addition, we thought it was important to also assess potential confounding among the sleep predictors, and have therefore determined the multicollinearity diagnostics for these variables too. These diagnostics revealed that for each model, variance inflation factor (VIF) values are below 5 and the tolerance statistics are above 0.2. Furthermore, the Pearson correlation coefficients between variables were also examined, confirming that none of the variables had a correlation higher than 0.8, and thus multicollinearity was not an issue in this data. The below sentence has been added to Paragraph 1 of the Statistical Analyses section to indicate that based on this assessment, predictor intercorrelations and tolerance levels are within acceptable ranges indicating that there is minimal collinearity within our data. Table S1 and

Table S2 reporting the full details of the multicollinearity assessment has also been added to Supplement 1, and is referred to in the additional sentence below describing the collinearity findings.

"A multicollinearity assessment using variance inflation factor, correlation coefficient and tolerance values showed no collinearity in the predictor variables (Supplement 1, Table S1 and Table S2)."

We also wanted to point out a couple of changes to the tables in Supplement 1. Firstly, in the revised submission, we have removed the first (previously Table S1. High burnout outcomes associated with shift work characteristics in police (n=3140)) in Supplement 1 because this table is not cited in the main text. Additionally, Supplement table 3 (previously Table S3. High burnout outcomes associated with sleep and sleepiness in police regularly working night shifts (1-4 nights per week or night shifts nearly every day; n=1199)) has also been removed from the revised submission because we do not directly cite this table in the main text.

There is a minor issue with how the PA dimension of burnout is described. "PA" is first introduced as "lack of personal accomplishment" – thus a high score should be 'bad'. In the sub-section measures, it is unclear whether the cut off of ≥39 points refers to the classification "high-moderate". In the results section, the text refers to "low PA" – by the original definition in the introduction, that implies a low lack of PA – a 'good' result. I am pretty sure that having "low PA" means being more likely to 'lack a sense of personal accomplishment' (i.e. a bad thing), but care is needed in describing this concept to avoid confusion. Confusion is compounded by the fact that PA ends up showing some negative results (i.e. in the opposite direction to other trends, and to what might be expected).

Response: Thank you for raising this important issue. You are correct that low scores on the PA dimension indicate "low PA", which refers to being more likely to have a reduced sense of personal accomplishment at work (i.e., a bad thing). After reviewing this description of PA in the MBI manual (Maslach et al. 1996), we have now realised that that the cut-off of ≥39 originally used in this study was incorrect. This may have been due to some ambiguity in how this dimension was described in Table 1 of MBI scoring manual, which resulted in us incorrectly interpreting low (i.e., ≥39 points) on the range of experienced burnout as low PA. Instead, low PA should have been defined as a score ≤31. This error in the scoring of PA has now been corrected in Paragraph 2 of the Methods section by including the below sentence.

"PA was also a dichotomous variable, but it was scored in the opposite direction to EE and DP, in which low (versus high-moderate) scores indicated low PA (PA ≤31 points)"

The correction to PA has impacted on a number of the original results in the manuscript. These results affected by the PA scoring error have been addressed in the revised manuscript. Reviewer 1 also highlighted a similar issue with the scoring of low PA. We have therefore addressed each of these changes to the manuscript (due to the corrected PA scoring) in response to this comment from Reviewer 1 above (see Reviewer 1 – Comment 3).

Although the scoring for PA was corrected, there were still some negative results for this dimension of burnout. Specifically, we found 1 to 4 night shifts per week and long work weeks reduced the odds of low PA. Paragraph 2 of the Results section has now been revised to include these new results. Furthermore, Paragraph 2 of the Discussion has also been revised by including the below sentence that highlights these results were unexpected, and provides a possible explanation for the findings.

"Although these findings were unexpected, they may highlight certain aspects of shift work in policing that allow personnel to feel effective and a sense of accomplishment in their work, but further research that explores the relationships between PA and weekly night shifts (e.g., comparing 1-2 and

3-4 night shifts per week) and work hours (e.g., ≥49 or ≥60 hours) in more detail is needed to confirm these findings."

Overall, the discussion makes a good attempt to explain quite a complex set of findings with some well argued and detailed analyses of the observed trends. I particularly liked the discussions of the distinction between long shifts and long weekly hours, the distinction between voluntary and mandatory overtime and the interesting speculation on the occupational differences in factors promoting DP among nightworkers. The discussion of fixed and rotating shifts highlights a weakness of this study that the nature of the shift schedules were underspecified, making it hard to interpret the comparisons between types of shift schedule. In particular, the definition of 'fixed shifts' (the 'baseline' group for schedule comparisons) does not differentiate between day and night work – a potentially important aspect which could have substantial bearing on the interpretations of the comparisons between schedule types.

Response: Thank you for raising this concern about the shift schedule variable and how this could limit the interpretations of the findings. We agree that not being able to identify the specific types of fixed shifts and rotations is a weakness of the study. Although this was a limitation, we have attempted to address this by restricting the sample to those who worked fixed shifts only, and then examining the relationship between night work and burnout in this sub-sample. This additional analysis revealed that in fixed shift workers, those working night shifts nearly every day had an increased risk of depersonalisation. We have highlighted this additional finding in Paragraph 6 of the Discussion by including the below text. Full details of this additional analysis can be found in Supplement 1.

"Despite this limitation, when we restricted the analysis to police only working fixed shifts, those reporting a greater frequency of night shifts had an increased risk of DP (Table S9, Table S10 and Table S11), which aligns with previous findings for fixed night work.43"

Despite this additional finding, we agree under-specification of shift schedules still prevents us from specifically differentiating between the specific types of fixed shifts (e.g., day vs night vs evening etc.) and shift rotations worked by participants. We have therefore highlighted this in the revised manuscript by including the below text in Paragraph 6 of the Discussion. Furthermore, we have also revised this paragraph to highlight the need to utilise objective shift information in future research to better understand the interactions between specific shift schedules and burnout.

"Specific shift rotations and types of fixed shifts may have varying effects on burnout, thus underspecifying these schedules in the current study may limit the interpretation of our findings. Further research that utilises detailed objective shift work information will allow for a more nuanced approach when it comes to examining interactions between different shift schedules and burnout in police and other emergency services."

Minor points
Page 4, 2nd paragraph. The statement about which shifts police have traditionally worked probably needs to be qualified as referring to police in the USA?

Response: Thank you for this helpful suggestion. We agree that this statement about the shifts police have traditionally worked should refer to police in the USA, given the literature in this area is from the USA (e.g., Vila et al.). The below revised sentence has been included to clarify this point. We have also added the study by Amendola et al. in USA police to this statement to further support the statement.

"Police have traditionally worked 8- to 10-hour shifts in the United States, although increasingly, departments are implementing compressed weekly schedules which involve officers working 11-, 12- and 16-hour shifts.1 3 8"

Discussion 1st para. Insertion needed – "Working night shifts *nearly every day *was associated with DP".

Response: We have inserted this additional detail into the sentence.