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)	Predicting new major depression symptoms from long working hours, psychosocial safety climate and work engagement: a population based cohort study
AUTHORS	Zadow, Amy; Dollard, Maureen F.; Dormann, Christian; Landsbergis, Paul

VERSION 1 – REVIEW

REVIEWER	Theorell, Tores Karolinska Institutet
REVIEW RETURNED	29-Sep-2020

GENERAL COMMENTS	This is a thought-provoking study of the possible association between number of working hours and risk of developing clinical depression (identified by means of a critical cut-off of scores in a self-administered Australian depression questionnaire). The authors expect the variable PSC (psychosocial safety climate) to be a potentially significant mediator; subjects with low PSC should have increased likelihood of working long hours and also increased risk of developing clinical depression. In univariate analysis they observe a strong relationship between good psychosocial safety climate and clinical depression (both cross-sectional at time 1 and prospective at time 2) as well as a weaker but significant relationship between long working hours and depression (increased risk, i e a relationship in the direction opposite to the expected one). However, when they move from univariate to multivariate analysis the relationship between long working hours and prospective (12 months follow-up) clinical depression disappears – it is even borderline significant in the direction opposite to the expected one with OR 0.45 for hours at or exceeding 55 hours a week whereas poor PSC (in one third of the participants) remains as a strong prospective predictor of clinical depression (odds ratio 3.22) after accounting for clinical depression at start. This is a very professional study. However, while the authors discuss possible sources of error in their main result, there are still some other possibilities that should have been mentioned. First of all, they actually have a very low participation rate particularly in the prospective part of this study. They mention that their participants are in general representing the Australian working population, but is that true also of number of working hours per week ? Secondly, one might ask what the reason for overtime work is among those who participate in the final step. The authors already discuss that question indirectly when they describe the possible complicated influence of PSC on number of working hours. In what kind of work situations are invited people willing or unwilling to participate? How does that relate to possible

	relationships between long working hours and depression? On the positive side (the authors mention positive PSC but there could be a lot of other factors involved) people who thrive at work may spend more happy hours at work than others and they are also more willing to fill out responses to the questions? The authors have used one indicator of social group, income, as possible confounder and found that unimportant. Type of work would have been much more interesting. The reasons for and consequences of overtime work are different in blue collar and white collar work, for instance. When the authors discuss the background literature I think they should be aware that working life is rapidly changing as a consequence of increasing home work for some grops. It may even be difficult for a lot of people to define “work”. In a lot of jobs nowadays emailing and writing on the home computer constitute a substantial part of the working day. How do reponders define the borders between work and leasure? This may have been different when this kind of epidemiological studies started. For instance, in a study from 1985 (Alfredsson et al Int J Epidemiology) a large-scale study based upon occupation and hospitalization for some illnesses, there was a clear relationship between a moderate amount of overtime work among women but not among men. The authors of the present article touch the possible effects of gender differences but I also feel that this should be put in a historical context and be related to occupation specificity.

REVIEWER	Peristera, Paraskevi Stressforskningsinstitutet
REVIEW RETURNED	01-Oct-2020

GENERAL COMMENTS	Predicting new clinical depression from long working hours and psychosocial safety climate: a population-based cohort study This article is investigating the association of long working hours and depression, which nowadays is a very important question. This paper requires a high revision in order to be a publishable paper. Please accept below some comments/suggestions. General Comments: The title including the expression clinical depression refers to clinical population or depressive symptoms? It was confusing I think More reflection is required on how to write the “Introduction” or first part of the article (lines 86 -169). More elaboration/ update of recent publications, in relation to existing literature and the research questions, is required for this part. For example, I could see a combination of arguments that belonged in the discussion part. In general, I think that this part should be written in a more cohesive way, so that the readers can understand what literature exists, what is lacking what are the aims of this study. See also more specific comments. In summary a clearer structure is needed, as well as a review of the recent literature, a connection to the research questions Depression: The authors give big emphasis on the different definitions of depression. Since their aim is not to compare different definitions, they should to a big extent revise/rewrite this section, focusing more on the findings themselves rather than the different thresholds.

	The authors provide as a limitation that they do not studied gender differences. I think they should include such analyses A sensitivity analysis by excluding those with those close to clinical depression is recommended The discussion and limitations parts should be revised, in the sense that several important issues are not discussed. See relevant comments Several terms are used to describe the outcome over the text, e.g. clinical depression, risk of depression etc. The authors should decide about which term to use and be consistent with that over the text. What is the difference between clinical depression and clinical population? Abstract: Lines 45-46: Should the authors include in the aims the association between PSC and new clinical depression? Lines 62-63. This seems like a repetition of what is already described in the results session. I would suggest to relate it more to practical implications. Strengths & Limitations: Lines 70-82: This paragraph repeats more or less what is already known from the abstract. The authors should focus more on the actual strengths and limitations of this study Introduction The problem: Lines 91-92: you give references about LWH and depression. Unclear if you refer to clinical or non-clinical depression Lines 92-94: Connect that to implications for health Lines 95-97: Now they refer to COVID-19. Do they study the associations due to COVID? How is this connected to their research questions? Lines 96-97: What are the practical gains of establishing guidelines? better scheduling, legislation and prioritization of preventive measures? Better health? I consider that the authors should revise the first three sessions and try to connect it more to literature review as well as societal aspects/needs. The authors have to review this part providing evidence of what we already know and what are the needs. Depression: Lines 9-112: How the different definitions contribute to this paper? If the authors want to support that the usefulness of measuring clinical depression, they should provide some evidence from literature also. It would be further interesting to refer to some findings from studies that are not only based in clinical population. Lines 118-120: This could be developed in the Discussion Part. Long-working hours

	Lines 122-123 I would suggest to write in a more cohesive way having together all the literature about associations between LWH and depression, Lines 124-139. In the way this is written one gets the message that this study will explore and find which threshold is more appropriate. A revision is needed focusing on the findings regarding the associations between LWH and outcome from different studies. The fact that they use different thresholds can be analyzed in the discussion session. Further the part about ILO / WHO definitions should be rather mentioned in the Methods session to justify the selection of the threshold Lines 140-142: Is that really a goal of this study? Focus rather what gives a better understanding of this relationship Lines 162-163: The authors should be more specific here. Methods Part Lines 213-215: Here should be given some details about the selection of this threshold as well as some references of studies that have use it before, see also previous comment Lines 232—It is rather confusing that you start referring to depression while planning to study clinical depression. Some rewriting – rearrangement is required Lines 254: Once has to refer which version of the software is used Results Is there possibility that these results indicate a suppression effect ? Table 4: It is not clear in the table which model is which. Add a column indicating, model 1, model 2, etc, Discussion In the discussion the authors provide a summary of their findings and some connection to previous findings. This section should be extended by explaining what do we gain from these, implications and importance for health/society What this study offers ? What are the theoretical and practical implications of these results/ implications for policy etc. ? How can employees promote better health ? Long working hours can affect the overall health of workers from different perspectives. Are other paths that should be studied in the future, discussed here? Is there any advantage of this definition of clinical depression Limitations Some issues that should be discussed in the limitations:

	Sex differences: See general comments. How the lack of longitudinal data affect these findings? What about the social desirability phenomenon (depressive individuals may be reluctant to admit their psychological problems even under conditions of anonymity and confidentiality) Line 320: should be long working hours? Line 323: I would suggest to run the analysis by excluding those with depressive symptoms close to the level of clinical depression as a sensitivity analysis Lines 335-336: It is not clear what the authors mean here. If they state that longitudinal studies are lacking is not correct. A better literature review is required in that case , and even explain what do they miss without longitudinal measures Implications for future research This part should be better connected to the limitations section

REVIEWER	Rivera, Adovich Northwestern University Feinberg School of Medicine
REVIEW RETURNED	06-Oct-2020

GENERAL COMMENTS	Study: Predicting new clinical depression from long working hours and psychosocial safety climate: a population based cohort study Detailed Comments to the authors:

Introduction	
99 to 100	DSM 5 is already available so I'm not sure why DSM 4 was the cited definition
	When discussing issues with current studies, they should also mention that a lot of previous studies used cross-sectional designs.
166	Their conceptual model suggests full mediation. Are they assuming only indirect effects of PSC on risk of depression?
Methods	
171 to 172	They mentioned using a random sample but did not clearly specify the sampling design. If they used multi-stage sampling, I would expect that the analysis mentions weighted models and some description on how weights were calculated. They need to add more details about sampling to support their claim that the sample is generalizable. Some results make it hard to believe. For example, their sex distribution was also skewed towards the female population when Australia's sex ratio is nearly 1.
198 to 199	They should be more specific as to how the public involvement affected the design of the study and how the pilot work affected the final methods.

	227 to 231	It would be good to see a sensitivity analysis where PSC is treated as a continuous rather than categorical variable.
	242 to 245	I am uncomfortable with the use of clinical depression based on just PHQ-9. I suggest use of terms like high risk for depression or high depressive symptoms instead.
		They should mention what tests of correlation were used in table 3.
	252 to 255	While I agree with their statements, it would still be good for them to actually run the test for mediation even if some of the models on their own do not show significance.
	Results	
		There results of their mediation analysis could be better presented using a mediation analysis diagram. For completeness, they should also include tests of direct and indirect effects.
		I got a bit confused in their variable specification. For example, they treated long working hours as a continuous variable when testing hypothesis 3 but converted it into a categorical variable in testing hypothesis 2.
		Table 4 is poorly constructed and hard to read. They should include appropriate headers to delineate the three models. They also include other information that aren't usually needed or redundant like beta when OR is already available, Wald statistic, and degrees of freedom.
	Discussion	
		While their conclusions are logically derived from presented results, they need to review the results after account for the sampling design in their statistical tests as well as perform a complete mediation analysis.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 (Töres Theorell)	Stockholm University Stress Research Institute Sweden	
	Reviewer Comments	Author Response
1.	This is a thought-provoking study of the possible association between number of working hours and risk of developing clinical depression (identified by means of a critical cut-off of scores in a self-administered Australian depression questionnaire).	Thank you. We agree.

2.	The authors expect the variable PSC (psychosocial safety climate) to be a potentially significant mediator; subjects with low PSC should have increased likelihood of working long hours and also increased risk of developing clinical depression. In univariate analysis they observe a strong relationship between good psychosocial safety climate and clinical depression (both cross-sectional at time 1 and prospective at time 2) as well as a weaker but significant relationship between long working hours and depression (increased risk, i e a relationship in the direction opposite to the expected one). However, when they move from univariate to multivariate analysis the relationship between long working hours and prospective (12 months follow-up) clinical depression disappears – it is even borderline significant in the direction opposite to the expected one with OR 0.45 for hours at or exceeding 55 hours a week whereas poor PSC (in one third of the participants) remains as a strong prospective predictor of clinical depression (odds ratio 3.22) after accounting for clinical depression at start.	This is correct.
3.	This is a very professional study. However, while the authors discuss possible sources of error in their main result, there are still some other possibilities that should have been mentioned. First of all, they actually have a very low participation rate particularly in the prospective part of this study. They mention that their participants are in general representing the Australian working population, but is that	Thank you. We have now added lines comparing the study sample with population data obtained from the Australian Bureau of Statistics to identify is the sample represents the general Australian working population (>35 hours) and also that participants are in general representing the Australian working population in terms of working hours per week (203-211).

	true also of number of working hours per week ?	
4.	Secondly, one might ask what the reason for overtime work is among those who participate in the final step. The authors already discuss that question indirectly when they describe the possible complicated influence of PSC on number of working hours. In what kind of work situations are invited people willing or unwilling to participate? How does that relate to possible relationships between long working hours and depression?	Thank you. We have now also completed additional sensitivity analyses examining the relationship between work engagement, long working hours and depression to identify whether employees who are engaged and spend long happy working hours are less likely to develop new major depression symptoms (291-295 and 371-383).
5.	On the positive side (the authors mention positive PSC but there could be a lot of other factors involved) people who thrive at work may spend more happy hours at work than others and they are also more willing to fill out responses to the questions?	Thank you for this suggestion. We have explored the role of work engagement and have identified some interesting findings - high levels of work engagement increase the odds of working long hours and subsequent levels of depression. These results have been included in this study (371-383)
6.	The authors have used one indicator of social group, income, as possible confounder and found that unimportant. Type of work would have been much more interesting. The reasons for and consequences of overtime work are different in blue collar and white collar work, for instance.	Thank you for this recommendation. We have included this in the 'Implications for Future Research' section (465-469).
7.	When the authors discuss the background literature I think they should be aware that working life is rapidly changing as a consequence of increasing home work for some groups. It may even be difficult for a lot of people to define "work". In a lot of jobs nowadays emailing and writing on the home computer constitute a substantial part of the working day. How do responders define the borders between work and leisure?	Thank you. We have now included this in the 'Implications for Future Research' section (480-484).

8.	This may have been different when this kind of epidemiological studies started. For instance, in a study from 1985 (Alfredsson et al Int J Epidemiology) a large-scale study based upon occupation and hospitalization for some illnesses, there was a clear relationship between a moderate amount of overtime work among women but not among men. The authors of the present article touch the possible effects of gender differences but I also feel that this should be put in a historical context and be related to occupation specificity.	Thank you. We have now included the historical context, effect of gender differences and occupation specificity in the introduction and discussion (138-150, 465-469). We have also completed additional sensitivity analyses examining gender differences (359-364).
	Reviewer 2 (Paraskevi Peristera)	Stockholm University Stress Research Institute Sweden
	Reviewer Comments	Author Response
	Introduction	
1	99 to 100 DSM 5 is already available so I'm not sure why DSM 4 was the cited definition	Many thanks for the recommendation. This has been removed.
2	When discussing issues with current studies, they should also mention that a lot of previous studies used cross-sectional designs.	Thank you. This has now been included in the introduction (123-136).
3	166 Their conceptual model suggests full mediation. Are they assuming only indirect effects of PSC on risk of depression?	Thank you for pointing this out. We have included a model showing full mediation in the introduction (176) and results (356) sections.
Method		
4	171 to 172 They mentioned using a random sample but did not clearly specify the sampling design. If they used multi-stage sampling, I would expect that the analysis mentions weighted models and some description on how weights were calculated. They need to add more details about sampling to support their claim that the sample is generalizable. Some results make it hard to believe. For	Thank you for this observation. The sex distribution is skewed because the sample only contained participants who were working >35 hours per week. In the sample and when reviewing the Australian Labour Force statistics released by the Australian Bureau of Statistics (ABS) there is a greater number of male Australian adults working more than 35 hours per week. We compared our sample

	example, their sex distribution was also skewed towards the female population when Australia's sex ratio is nearly 1.	with ABS data (>35 hours per week) and the gender ratio is similar: study data (61% male/ 39% female) compared to ABS data (62.4% male/ 37.6% female). We have also compared the amount of participants working over 35 hours per week in the study sample (37%) with the Australian workforce ABS data (32%). This information has been included in the methods section (203-211).
5	198 to 199 They should be more specific as to how the public involvement affected the design of the study and how the pilot work affected the final methods.	Thank you. More detail has been added (216-217).
6	227 to 231 It would be good to see a sensitivity analysis where PSC is treated as a continuous rather than categorical variable.	Thank you for this recommendation. We have completed additional analyses using multinominal regression using PSC measured as a continuous variable (288-291 and 339-345).
7	242 to 245 I am uncomfortable with the use of clinical depression based on just PHQ-9. I suggest use of terms like high risk for depression or high depressive symptoms instead.	Thank you. We have changed the term to "major depression symptoms" in accordance with the wording of the recent BMJ publication reviewing the same tool used in this paper: Levis B, Benedetti A, Thombs BD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019; 365.
8	They should mention what tests of correlation were used in table 3.	Thank you. The name of the test has been added (312).
9	252 to 255 While I agree with their statements, it would still be good for them to actually run the test for mediation even if some	Thank you for this recommendation. Mediation is supported if H1, and H3 hold, and H2 holds with PSC as an independent predictor. Hypothesis

	of the models on their own do not show significance.	1 proposed that PSC would be negatively related to new cases of depressive symptoms. In the second model we added PSC, and the effect was significant. Hypothesis 1 was supported; Low PSC was significantly associated with depression. Hypothesis 2 proposed a positive relationship between long working hours and new cases of depressive symptoms. In the final model we added long working hours categories and none were significant. Hypothesis 2 was not supported. Hypothesis 3 proposed that PSC is negatively related to long working hours. We ran a linear regression model, regressing long working hours on PSC accounting for age, gender and income. The effect was not significant; Hypothesis 3 was not supported. Hypothesis 4 proposed the relationship between PSC and depression would be mediated by long working hours; since H2 and H3 were not supported, this hypothesis was not supported. Following recommendations by the reviewers we did further sensitivity analyses examining whether work engagement and major depression symptoms would be mediated by long working hours. We found a significant mediation effect for the 41-48 long working hours group and the ≥ 55 hours category excluding those with mild symptoms of depression (PHQ-9 scores between 5-9) and controlling for age, gender, income and PSC (see Table 5) (371-388).
	Results	
10	There results of their mediation analysis could be better presented using a mediation analysis diagram. For	Thank you. A diagram has been added in the introduction (176) and results (356) section.

	completeness, they should also include tests of direct and indirect effects.	
11	I got a bit confused in their variable specification. For example, they treated long working hours as a continuous variable when testing hypothesis 3 but converted it into a categorical variable in testing hypothesis 2.	Thank you for this recommendation. We have now completed additional analyses using multinominal regression using PSC measured as a continuous variable (288-291 and 339-345).
12	Table 4 is poorly constructed and hard to read. They should include appropriate headers to delineate the three models. They also include other information that aren't usually needed or redundant like beta when OR is already available, Wald statistic, and degrees of freedom.	Thank you. Headers have been included to delineate the three models. We would prefer to leave beta in the table as this assists interdisciplinary collaboration with our colleagues in psychology who prefer these statistics. Please let us know if this would be possible otherwise we would be happy to remove (351).
	Discussion	
13	While their conclusions are logically derived from presented results, they need to review the results after account for the sampling design in their statistical tests as well as perform a complete mediation analysis.	We have reconducted the analysis, and since the logical steps - PSC to working hours to depression are not significant we draw the conclusion of no mediation. However we now report formal tests of mediation in relation to work engagement to long working hours to major depression symptoms.
	Reviewer 3 (Ado Rivera)	Northwestern University, United States
	Reviewer Comments	Author Response
1.	This article is investigating the association of long working hours and depression, which nowadays is a very important question. This paper requires a high revision in order to be a publishable paper. Please accept below some comments/suggestions.	Thank you for your comments.
2.	The title including the expression clinical depression refers to clinical population or depressive symptoms? It was confusing I think	Thank you for highlighting this. To reduce potential for confusion we have changed the term to "major depression symptoms" in accordance with the wording of the recent BMJ publication

		reviewing the same tool a used in this paper: Levis B, Benedetti A, Thombs BD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019; 365
3.	More reflection is required on how to write the “Introduction” or first part of the article (lines 86 - 169). More elaboration/ update of recent publications, in relation to existing literature and the research questions, is required for this part. For example, I could see a combination of arguments that belonged in the discussion part. In general, I think that this part should be written in a more cohesive way, so that the readers can understand what literature exists, what is lacking what are the aims of this study. See also more specific comments. In summary a clearer structure is needed, as well as a review of the recent literature, a connection to the research questions	Thank you. The introduction has been rewritten to include what literature exists and what is lacking with a connection to the research questions.
4.	Depression: The authors give big emphasis on the different definitions of depression. Since their aim is not to compare different definitions, they should to a big extent revise/rewrite this section, focusing more on the findings themselves rather than the different thresholds.	Thank you. The definitions have been removed from the introduction.
5.	The authors provide as a limitation that they do not studied gender differences. I think they should include such analyses	Thank you. We have now completed sensitivity analyses studying gender differences in the results section (359-364).
6.	A sensitivity analysis by excluding those with those close	Thank you. We have completed the sensitivity analysis excluding those with mild symptoms of

	to clinical depression is recommended	depression (PHQ-9 scores between 5-9). These findings are reported in the results section (364-371).
7.	The discussion and limitations parts should be revised, in the sense that several important issuesare not discussed. See relevant comments	Thank you for these comments. We have revised the discussion and limitations.
8.	Several terms are used to describe the outcome over the text, e.g. clinical depression, risk of depression etc. The authors should decide about which term to use and be consistent with that over the text. What is the difference between clinical depression and clinical population?	Thank you for highlighting this. To reduce the confusion we have changed the term to “major depression” in accordance with the wording of the recent BMJ publication reviewing the same tool a used in this paper: Levis B, Benedetti A, Thombs BD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019; 365
	Abstract:	
9.	Lines 45-46: Should the authors include in the aims the association between PSC and new clinical depression?	Thank you. We have now included the relationship between long working hours, PSC, work engagement and new major depressive symptoms as the initial objective.
10.	Lines 62-63. This seems like a repetition of what is already described in the results session. I would suggest to relate it more to practical implications.	Thank you. We have now removed the statistics from this section.
	Strengths & Limitations:	
11.	Lines 70-82: This paragraph repeats more or less what is already known from the abstract. The authors should focus more on the actual strengths and limitations of this study	Thank you. This section has been revised.
	Introduction	
12.	The problem: Lines 91-92: you give references about LWH and depression. Unclear if you refer to clinical or nonclinical depression	Thank you. We have revised this section to differentiate between studies which have examined depressive symptoms and studies that have specified depressive disorders clinically diagnosed or assessed by a

		structured interview (92-93 and 123-136).
13.	Lines 92-94: Connect that to implications for health	Many thanks for this recommendation. Have made the link to health (100).
14.	Lines 95-97: Now they refer to COVID-19. Do they study the associations due to COVID? How is this connected to their research questions?	Thank you. Have linked the pandemic to the research question (96-97).
15.	Lines 96-97: What are the practical gains of establishing guidelines? better scheduling, legislation and prioritization of preventive measures? Better health?	Thank you. We have included these suggestions (97-101).
16.	I consider that the authors should revise the first three sessions and try to connect it more to literature review as well as societal aspects/needs. The authors have to review this part providing evidence of what we already know and what are the needs.	Thank you. We have revised the first three sections.
	Depression:	
17.	Lines 9-112: How the different definitions contribute to this paper? If the authors want to support that the usefulness of measuring clinical depression, they should provide some evidence from literature also. It would be further interesting to refer to some findings from studies that are not only based in clinical population.	Thank you. The different definitions have been removed and findings from current studies are provided.
18.	Lines 118-120: This could be developed in the Discussion Part.	Thank you. This has been added to the discussion.
	Long-working hours	
19.	Lines 122-123 I would suggest to write in a more cohesive way having together all the literature about associations between LWH and depression,	Thank you. This has been rewritten in a more comprehensive way including all of the relevant literature outlining associations between LWH and depression (123-150).
	Lines 124-139. In the way this is written one gets the message that this study will explore and find which threshold is more	Thank you. This has been changed as suggested.

	appropriate. A revision is needed focusing on the findings regarding the associations between LWH and outcome from different studies. The fact that they use different thresholds can be analyzed in the discussion session. Further the part about ILO / WHO definitions should be rather mentioned in the Methods session to justify the selection of the threshold	
20.	Lines 140-142: Is that really a goal of this study? Focus rather what gives a better understanding of this relationship	Thank you for this recommendation. This will be changed as suggested.
21.	Lines 162-163: The authors should be more specific here.	Thank you. This has been revised.
	Methods Part	
22.	Lines 213-215: Here should be given some details about the selection of this threshold as well as some references of studies that have use it before, see also previous comment	Thank you. This has been included (226-234).
23.	Lines 232—It is rather confusing that you start referring to depression while planning to study clinical depression. Some rewriting – rearrangement is required	Thank you. This has been modified to major depression symptoms (258).
24.	Lines 254: Once has to refer which version of the software is used	Thank you. This has been included (294-295).
	Results	
25.	Is there possibility that these results indicate a suppression effect ?	We did not have any evidence to indicate suppression effects and rather we thought that there might be a sensitivity issue, so we ran the sensitivity analyses and found some interesting results. These have been included in the abstract (61-63), results (358-383) and discussion (400-404).
26.	Table 4: It is not clear in the table which model is which. Add a column indicating, model 1, model 2, etc,	Thank you. This has been added (351).

	Discussion	
27.	In the discussion the authors provide a summary of their findings and some connection to previous findings. This section should be extended by explaining what do we gain from these, implications and importance for health/society	Thank you for this suggestion. This section has been extended (404-407).
28.	What this study offers? What are the theoretical and practical implications of these results/ implications for policy etc. ? How can employees promote better health ?	Thank you. More information has been provided (407-410).
29.	Long working hours can affect the overall health of workers from different perspectives. Are other paths that should be studied in the future, discussed here?	Thank you. Recommendations for further paths are suggested (460-469 and 471-478 and 480-491).
30.	Is there any advantage of this definition of clinical depression	Thank you for highlighting this. To reduce the confusion we have changed the term to “major depression” in accordance with the wording of the recent BMJ publication reviewing the same tool used in this paper: Levis, B., Benedetti, A., & Thombs, B. D. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. bmj, 365.
	Limitations	
31.	Some issues that should be discussed in the limitations:	
32.	Sex differences: See general comments.	Thank you. We have now completed analyses examining gender differences (359-364).
33.	How the lack of longitudinal data affect these findings?	Thank you. This is now addressed in the introduction (123-125).
34.	What about the social desirability phenomenon (depressive individuals may be reluctant to admit their psychological problems even under conditions of anonymity and confidentiality)	Thank you. This is now raised in the discussion (433-43).

35.	Line 320: should be long working hours?	Thank you. This has now been changed.
36.	Line 323: I would suggest to run the analysis by excluding those with depressive symptoms close to the level of clinical depression as a sensitivity analysis	Thank you. This has now been completed (364-371).
37.	Lines 335-336: It is not clear what the authors mean here. If they state that longitudinal studies are lacking is not correct. A better literature review is required in that case , and even explain what do they miss without longitudinal measures	Thank you. This wording has been revised (426-431).
	Implications for future research	
38.	This part should be better connected to the limitations section	Thank you. This section has been revised.

VERSION 2 – REVIEW

REVIEWER	Theorell, Tores Karolinska Institutet
REVIEW RETURNED	02-Feb-2021

GENERAL COMMENTS	This is fine work and the statistical handling, the background description and the formal reasoning are excellent. It has become customary in psychiatric epidemiology to use standardized self-rating scales with different cut-offs for different degrees of depression. This is accepted although psychiatrists are still sceptical. In some places in the ms the authors are cautious and talk about major depression symptoms rather than major depression. I believe one way of countering criticism would be to mention again in the discussion the high sensitivity and specificity reported in ref 46 and to spell out the high percentage of subjects with a score ≥ 10 who would be considered to have major depression when examined by a psychiatrist and conversely that a low percentage below 10 that would be judged by a psychiatrist to have major depression. The readers of this article will be a mixed group and all are not aware of what a joint sensitivity and specificity of 88% means. One thing that worries me a little bit methodologically is that the cut-off of 37 for PSC was identified in relation to risk of developing major depression in previous work based upon the same data set. Thus, the relationship between PSC may have been “optimized”, leading to overestimation of the true association. The authors need to discuss this. Finally I lack a discussion about the obvious fact that the participation rate in this population study was low (as in most similar studies in similar countries). I think the reader needs a discussion about possible consequences of this for the findings.

	For instance, one could think of a situation in which subjects who work in a high PSC site and are depressed avoid participation. Table 1 seems to indicate that subjects with high PSC have a low representation.

REVIEWER	Peristera, Paraskevi Stressforskningsinstitutet
REVIEW RETURNED	24-Feb-2021

GENERAL COMMENTS	Predicting new clinical depression from long working hours and psychosocial safety climate: a population-based cohort study I would like to thank the authors for the revised version of the manuscript where they considered all previous suggestions/comments. Reading this revised manuscript, I think some new revisions are needed before accepting this manuscript for publication. Abstract A line about the sex differences results would be appreciated. The expression sensitivity analyses may not add much in the value of the findings. Instead could describe in words what was the case The results should be also more thoughtfully described/rephrased so as to distinguish to which sample/ scenario you refer to. Reading the Results session in abstract and this in main text one may wonder about the main results? Introduction The hypotheses are well stated however some theoretical reasoning seems to be missing or maybe a sentence relating that to the aims of this work and / or previous findings. Should the sensitivity analyses be also described as separate hypotheses, since you get some interesting results there? Nowhere is described work engagement although is used in the analyses Sample & Figure 2 Reading again the sample description I find that it is rather confusing and not in complete correspondence to Figure 2 Line 194: It is stated that the final sample was 1152 however this number is not seen in Figure 2 Lines 197& 208-209: You wrote: worked >=35 hrs. but in lines 208-209 it is written >35 hrs. Line 199: What does RR mean? Write the full word and in parenthesis (RR) Nowhere in the sample description the n=1084 is given while in Figure 2 as well as in Table 2 is shown as Sample 5 Finally, I wonder, if including in Figure 2: Sample 1, 2, etc. is rather confusing and one should only use it in the last box for the final sample Statistical Analysis I think you should even in text write which is model 1, model 2 etc. as shown in Table 4 Line 287: It would be useful to explain the purpose/motivation of doing these sensitivity analyses instead of writing as suggested by the reviewers Results

	Line 327: Include in parenthesis Model 1 In the sensitivity analysis: Given the mixed picture that you describe in the analyses by gender I wonder if it would be better to run a model with an interaction term by gender and see what you get in terms of significance. Usually this is the procedure to decide about gender specified analyses before deciding to run gender-specified analyses Lines 379-381: For first time you mention this bootstrap method. Should not this be part of the statistical analysis session? As well as explain in one line the advantage of using it Discussion Line 397 this finding refers to sensitivity analysis 2?

REVIEWER	Rivera, Adovich Northwestern University Feinberg School of Medicine
REVIEW RETURNED	14-Feb-2021

GENERAL COMMENTS	This revised manuscript shows much improvement. I do not have any major comments regarding how they conducted the study or analysis. Minor comments: (1) page 11, line 328-330 – I think the authors meant to report model 2 odds ratio of PSC <37 but they reported model 3 OR instead (model 2 is 3.14; model 3 is 3.22). (2) Table 4 - Please specify what B means in the table notes (I'm assuming it's the beta coefficient) and p-values should not be reported as "0.00". Those should be reported as <0.01 or whatever the lower limits are for the statistical software. (3) Figure 3 – Box should be Major Depression symptoms not just major depression. Consider splitting the figure into two parts, one for the PSC-long hours-depressive symptoms diagram and the other for the work engagement-long hours-depressive symptoms diagram. The figure in its current form sparks two questions: Does work engagement lie in the causal pathway of PSC to depressive symptoms? What's the association of PSC to depressive symptoms after adjusting for work engagement? I suggest reporting the full results of the models (similar to Table 4) for all the sensitivity analysis as part of supplementary content. Authors may need to conduct additional analysis and/or tackle these questions in the discussion. (4) Page 34, line 362 - Add "Neither" to improve clarity of sentence "For females, PSC nor working hours...

VERSION 2 – AUTHOR RESPONSE

Revisions		
	Reviewer 1 (Töres Theorell)	Stockholm University Stress Research Institute Sweden
	Reviewer Comments	Author Response
1.	This is fine work and the statistical handling, the background description and the formal reasoning are excellent.	Thank you for this excellent feedback. It is much appreciated.
2.	It has become customary in psychiatric epidemiology to use standardized self-rating scales with different cut-offs for different degrees of depression. This is accepted although psychiatrists are still sceptical. In some places in the ms the authors are cautious and talk about major depression symptoms rather than major depression. I believe one way of countering criticism would be to mention again in the discussion the high sensitivity and specificity reported in ref 46 and to spell out the high percentage of subjects with a score ≥ 10 who would be considered to have major depression when examined by a psychiatrist and conversely that a low percentage below 10 that would be judged by a psychiatrist to have major depression. The readers of this article will be a mixed group and all are not aware of what a joint sensitivity and specificity of 88% means.	We appreciate this suggestion and have included additional information in the discussion to address this (lines 513-519).
3.	One thing that worries me a little bit methodologically is that the cut-off of 37 for PSC was identified in relation to risk of developing major depression in previous work based upon the same data set. Thus, the relationship between PSC may have been “optimized”, leading to overestimation of the true association. The authors need to discuss this.	Thank you for raising this. We have now addressed this issue specifically in the discussion (lines 521-525). We have also completed additional analysis with PSC as a continuous variable (line 524-525).
4.	Finally I lack a discussion about the obvious fact that the participation rate in this population study was low (as in most similar studies in similar countries). I think the reader needs a discussion about possible consequences of this for the	Thank you for this recommendation. This has now been raised in the discussion (lines 527-532).

	findings. For instance, one could think of a situation in which subjects who work in a high PSC site and are depressed avoid participation. Table 1 seems to indicate that subjects with high PSC have a low representation.	
	Reviewer 2 (Ado Rivera)	Northwestern University, United States
	Reviewer Comments	Author Response
1	This revised manuscript shows much improvement. I do not have any major comments regarding how they conducted the study or analysis.	Thank you for this helpful feedback. It is much appreciated.
2	page 11, line 328-330 – I think the authors meant to report model 2 odds ratio of PSC <37 but they reported model 3 OR instead (model 2 is 3.14; model 3 is 3.22).	Thank you. You are correct. We initially changed this number in the manuscript but we have now reversed the analysis using high PSC (>37) rather than low PSC (<37) to assist with the interpretation (high PSC will be negatively related to major depression symptoms) so the entire section has now been modified. The new score is (B = -1.15, SE = .35, p < .001, OR = .32, 95% CI .16-.63) (line 376-379). This means that low PSC is associated with a 3 fold increase in risk for new major depression symptoms (lines 387-390)
3	Table 4 - Please specify what B means in the table notes (I'm assuming it's the beta coefficient) and p-values should not be reported as "0.00". Those should be reported as <0.01 or whatever the lower limits are for the statistical software.	Thank you. These have been revised (line 395).
4	Figure 3 – Box should be Major Depression symptoms not just major depression.	Thank you for highlighting this. The term 'symptoms' has been now included.
5	Consider splitting the figure into two parts, one for the PSC-long hours-depressive symptoms diagram and the other for the work engagement-long hours-depressive symptoms diagram.	Thank you. This has been completed (Figure 1, Figure 3 and Figure 4).
6	The figure in its current form sparks two questions: Does work engagement lie in the causal pathway of PSC to depressive symptoms? What's the association of PSC to depressive symptoms after adjusting for work engagement? I suggest reporting the full results of the models (similar to Table 4) for all the sensitivity	Thank you. We have now run the complete model and provide extensive supplementary material examining a range of models including both PSC and work engagement (Tables 4, 5, 6 and Supplementary Tables 1a, 2a, 3a). Engagement was not related to major depression (Hypothesis 5 was not supported) so we propose that engagement does

	analysis as part of supplementary content. Authors may need to conduct additional analysis and/or tackle these questions in the discussion.	not mediate the relationship between PSC and major depression symptoms.
7	Page 34, line 362 - Add "Neither" to improve clarity of sentence "For females, PSC nor working hours...	Thank you. We initially made this change but have now had to completely re-write this section (lines 466-474) due to the further analyses we have undertaken.
	Reviewer 3 (Paraskevi Peristera)	Stockholm University Stress Research Institute Sweden
	Reviewer Comments	Author Response
1.	I would like to thank the authors for the revised version of the manuscript where they considered all previous suggestions/ comments. Reading this revised manuscript, I think some new revisions are needed before accepting this manuscript for publication.	Thank you for this important feedback. It is much appreciated.
	Abstract	
2.	A line about the sex differences results would be appreciated.	Thank you. This has been included (line 62-63).
3.	The expression sensitivity analyses may not add much in the value of the findings. Instead could describe in words what was the case	Thank you. We have removed the term from the abstract.
4.	The results should be also more thoughtfully described/rephrased so as to distinguish to which sample/ scenario you refer to. Reading the Results session in abstract and this in main text one may wonder about the main results?	Thank you for this recommendation. We have rewritten the abstract to reflect the order of the separate hypotheses and scenarios (lines 57-66).
	Introduction	
5.	The hypotheses are well stated however some theoretical reasoning seems to be missing or maybe a sentence relating that to the aims of this work and / or previous findings.	Thank you. Additional text has now been included (lines 143-210).
6.	Should the sensitivity analyses be also described as separate hypotheses, since you get some interesting results there?	Thank you. Additional hypotheses have been added (lines 182-201).
7.	Nowhere is described work engagement although is used in the analyses	Thank you. Hypotheses for work engagement have been added (lines 186,194) and further

		description of the construct and previous findings have been included (lines 290-294).
	Sample & Figure 2	
8.	Reading again the sample description I find that it is rather confusing and not in complete correspondence to Figure 2	Figure 2 has been redone to make the samples and removals clearer to understand.
9.	Line 194: It is stated that the final sample was 1152 however this number is not seen in Figure 2	Thank you for identifying this. It has been reviewed and corrected.
10.	Lines 197& 208-209: You wrote: worked >=35 hrs. but in lines 208-209 it is written >35 hrs.	Thank you for having us check this. We have searched the document and removed any cases stating >35 hours as this needs to be ≥ 35 hours. We have not changed the cases stating < 35 which refers to cases removed.
11.	Line 199: What does RR mean? Write the full word and in parenthesis (RR)	Thank you. This has been amended with the term 'response rates' included (line 223).
12.	Nowhere in the sample description the n=1084 is given while in Figure 2 as well as in Table 2 is shown as Sample 5	Thank you for identifying this. It has been reviewed and corrected.
13.	Finally, I wonder, if including in Figure 2: Sample 1, 2, etc. is rather confusing and one should only use it in the last box for the final sample	Thank you. Figure 2 has been amended and the sample labels removed.
	Statistical Analysis	
14.	I think you should even in text write which is model 1, model 2 etc. as shown in Table 4	Thank you. These have been added in lines 313-343.
15.	Line 287: It would be useful to explain the purpose/motivation of doing these sensitivity analyses instead of writing as suggested by the reviewers	Thank you. This has been revised to explain the purpose of these sensitivity analyses in lines 348-353.
	Results	
16.	Line 327: Include in parenthesis Model 1	Thank you. We have now named all the models.
17.	In the sensitivity analysis: Given the mixed picture that you describe in the analyses by gender I wonder if it would be better to run a model with an interaction term by gender and see what you get in terms of significance. Usually this is the procedure to decide about gender	Thank you for this suggestion. This has now been conducted (lines 345-448 and 466-474).

	specified analyses before deciding to run gender-specified analyses	
18.	Lines 379-381: For first time you mention this bootstrap method. Should not this be part of the statistical analysis session? As well as explain in one line the advantage of using it	Thank you. An explanation has now been provided in the statistical analysis section (lines 335-339).
	Discussion	
19.	Line 397 this finding refers to sensitivity analysis 2?	Thank you. We have added in the sensitivity analysis result for the relationship between long working hours and depression (486-488).