PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Changes in weekly working hours, proportion of doctors with hours above the limitations of EUs working time directive (EWDT) and time spent on direct patient care for doctors in Norway from 2016 to 2019: A study based on repeated surveys
AUTHORS	Rosta, Judith; Rø, Karin

VERSION 1 – REVIEW

REVIEWER	Flinterman, Linda NIVEL
REVIEW RETURNED	14-Dec-2022

GENERAL COMMENTS	This a well written and clear paper on an relevant and up-to-date topic. I only have some minor points that would make the paper even better. In the methods you describe that you do a repeated measurement analysis. But it is unclear how many doctors did fill in one or both questionnaires from the results. As you have a dynamic cohort there should be participants that only were able to participate in one of the questionnaires. I would add this to results. An analysis separately on those that filled out the survey twice would also solve part of the problem of the self-reporting of the hours. As they most likely filled it out similarly both times.

REVIEWER	Weaver, Matthew Harvard Medical School
REVIEW RETURNED	31-Dec-2022

GENERAL COMMENTS	Thank you for the opportunity to review this work, which reports on the work hours of physicians in Norway from 2016-2019. I have several questions and suggestions for editorial revision. 1) Overarching methods question: It is stated several times, “There were similarities in the survey methods and measurements at two points in time”. Were the key items surrounding work hours the same at both time points? If so, the description should be modified accordingly. If the items differed, please describe how the approaches differed. 2) Overarching methods question: It is stated in the discussion that the same doctors answered at the same points in time. Is this is a before-after design, or is it more that the same general pool of potential participants was invited (some of these being prior respondents and some not)? I internalized that some doctors would be practicing and thus in the general pool of potential

	respondents at both time points, so a random intercept was included to account for this, but this is really two cross-sectional assessments. If so, I would de-emphasize the language regarding paired responses. 3) Overarching methods question: Related to the previous, it is also stated in the methods that the design allowed for each person to become his/her own control. This doesn't match the analysis as I understood it. The random intercept would account for the dependence between repeated measures, but it's not as though a within-person analysis was conducted to estimate the change within persons, or was it? 4) Abstract comment: The conclusion states that, "Regulating work hours can be a useful intervention for patient care and for doctors' wellbeing." I have concerns including this as the last line of the abstract as this was not investigated in the work. The sentence could be omitted, modified to highlight work hours as a potential target for interventions, or replaced with an alternative statement that is substantiated by data in the report. 5) Background comments: a. The final sentence of paragraph 1 seems out of place and could be omitted. The content is included in paragraph 3. b. Paragraph 2, the sentence with citation 17, should be edited for clarity. 6) Methods comments: a. Working hours for doctors in Norway: Please revise for clarity. Phrases including, "almost always" and "there is a possibility of" should be revised. In addition, the "list-patient" system is not widely known and should be described. Some of this content could alternatively be moved to the introduction or discussion and the rest omitted, as this does not seem to describe the methods of the study. b. Please include how the potential participants were sampled, e.g. randomly, quota-based, etc. c. Are you able to separate the part-time work from work at the primary job? d. Page 8 lines 37-41: Why were doctors over 70 excluded? What was the level of missingness after excluding responses for missing data? Would this be the 123 and 160 as a footnote in Table 1, out of the 1604 and 1511? e. Patient and public involvement: The last sentence of this section should be deleted (referring to more popular formats). 7) Discussion comments: a. The discussion as a whole should be substantially revised and shortened to focus less on your own findings and more comparing/contrasting elements of your findings with other work.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. Linda Flinterman, NIVEL

The authors wish to thank Dr Linda Flinterman for her valuable comments and suggestions. The manuscript has been revised along the indicated lines.

Comments to the Author:
In the methods you describe that you do a repeated measurement analysis. But it is unclear how many doctors did fill in one or both questionnaires from the results. As you have a dynamic cohort there should be participants that only were able to participate in one of the questionnaires. I would add this to results. An analysis separately on those that filled out the survey twice would also solve part of the problem of the self-reporting of the hours. As they most likely filled it out similarly both times.

Response
The samples in 2016 and 2019 were nearly identical with exception of n=111 doctors in 2019 that left the panel due to retirement, death or voluntary withdrawal (Table 1).

The majority of doctors did fill in both questionnaires n=1189 of 1481 (80%) in 2016 and 1189 of 1351 (88%) in 2019.

Due to the high percentage of doctors who reported at both time points we believe that a separate analysis for that group is not necessary.

We expanded the description of the "Respondents" and "Strengths and limitations".

Please see
-"Respondents", page 9, line 9 – page 10, line 3.
-"Strengths and limitations", Page 18, line 13-14.

Reviewer 2:

Dr. Matthew Weaver, Harvard Medical School

The authors wish to thank Dr Matthew Weaver for his valuable comments and suggestions. The manuscript has been revised along the indicated lines.

Comment 1
Overarching methods question: It is stated several times, “There were similarities in the survey methods and measurements at two points in time”. Were the key items surrounding work hours the same at both time points? If so, the description should be modified accordingly. If the items differed, please describe how the approaches differed.
Response 1
The key items on working hours were the same at both time points.

We now modified the description of "Depended variables".

Please see "Dependent variables", page 7, line 10.

Comment 2

Overarching methods question: It is stated in the discussion that the same doctors answered at the same points in time. Is this is a before-after design, or is it more that the same general pool of potential participants was invited (some of these being prior respondents and some not)? I internalized that some doctors would be practicing and thus in the general pool of potential respondents at both time points, so a random intercept was included to account for this, but this is really two cross-sectional assessments. If so, I would de-emphasize the language regarding paired responses.
Response 2
The samples in 2016 and 2019 were nearly identical with exception of n=111 doctors in 2019 that left the panel due to retirement, death or voluntary withdrawal (Table 1).

The majority of doctors did fill in both questionnaires n=1189 of 1481 (80%) in 2016 and 1189 of 1351 (88%) in 2019.

We now clarified the description of "Respondents".

Please see "Respondents", page 9, line 27 – page 10, line 3.

Comment 3
Overarching methods question: Related to the previous, it is also stated in the methods that the design allowed for each person to become his/her own control. This doesn't match the analysis as I understood it. The random intercept would account for the dependence between repeated measures, but it's not as though a within-person analysis was conducted to estimate the change within persons, or was it?

Response 3
As the vast majority of respondents were the same both in 2016 and in 2019 we assessed this as a within-subjects design; i.e. how individual male GPs on average had increased their work hours from 2016 to 2019, based on the results from linear mixed model analysis. For example, linear mixed model is used for analysis of cross-over trials and within-participant experiments (see references). In hindsight, stating that they are their own control might be confusing. We assessed changes over time with repeated measurements and not differences between a control and treatment as such.

In addition, since the same doctors were polled on each occasion largely, the reliability of changes over time in the data increases significantly and could be assessed as repeated measurements

References:
Reference on linear mixed models for within-participant experiments:
Frontiers | Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui) (frontiersin.org)

Reference on linear mixed models for cross-over studies:
Analysis of Data from a Cross-Over Trial | SpringerLink

We have rephrased the sentence (in strengths and limitations) to emphasize the repeated measures design.

Please see "Strengths and limitations", page 18, line 28 – page 19, line 2.

Comment 4

Abstract comment: The conclusion states that, “Regulating work hours can be a useful intervention for patient care and for doctors’ wellbeing.” I have concerns including this as the last line of the abstract as this was not investigated in the work. The sentence could be omitted, modified to highlight work hours as a potential target for interventions, or replaced with an alternative statement that is substantiated by data in the report.
Response 4
We modified the last sentence in the "Abstract".

Please see "Abstract", page 2, line 23-24.

Comment 5.a
Background comments:
a. The final sentence of paragraph 1 seems out of place and could be omitted. The content is included in paragraph 3.
Response 5.a
We omitted final sentence of paragraph 1 in "Background".

Please see "Background", page 4, paragraph 1.

Comment 5.b
Background comments:
Paragraph 2, the sentence with citation 17, should be edited for clarity.
Response 5.b
We clarified the sentence with citation 17 in "Background".

Please see "Background", page 4, line 18-19.

Comment 6.a
Methods comments:
Working hours for doctors in Norway: Please revise for clarity. Phrases including, “almost always” and “there is a possibility of” should be revised. In addition, the “list-patient” system is not widely known and should be described. Some of this content could alternatively be moved to the introduction or discussion and the rest omitted, as this does not seem to describe the methods of the study.
Response 6.a
We revised the description of "Working hours for doctors in Norway", added information about the "list-patient" system and moved it to the "Background".

Please see "Background", page 4, line 24 – page 5, line 8.

Comment 6.b
Methods comments:
Please include how the potential participants were sampled, e.g. randomly, quota-based, etc.
Response 6b
We included information on how the potential participants were sampled.

Please see Page 6, line 13-27.

Comment 6.c
Methods comments:
Are you able to separate the part-time work from work at the primary job?
Response 6c
Unfortunately no.
The items on an average working week include also on-call and any part-time job(s).

Please see description of "Dependent variables", page 7, line 13-14.

Comment 6.d
Methods comments:
Page 8 lines 37-41: Why were doctors over 70 excluded? What was the level of missingness after excluding responses for missing data? Would this be the 123 and 160 as a footnote in Table 1, out of the 1604 and 1511?
Response 6d
Doctors ≥70 year were excluded to maintain the sample`s representativity of practising doctors in Norway. The "Statistics on all Members of the Norwegian Medical Association" include doctors under 70 years as it is assumed that the real retirement age is just under 70 years (https://www.legeforeningen.no/om-oss/legestatistikk/yrkesaktive-leger-i-norge/)."

Please see the revised "Inclusion and exclusion criteria", Page 8, line 22-25.

Yes, the level of missingness after excluding responses for missing data were:
-in 2016: 1604-123=1481
-in 2019: 1351-160=1351

Please see "All(a)" in Table 1, Page 26

Comment 6.e
Methods comments:
Patient and public involvement: The last sentence of this section should be deleted (referring to more popular formats).
Response 6.e
We delated the last sentence of Patient and public involvement.

Please see " Patient and public involvement", Page 9.

Comment 7
Discussion comments:
The discussion as a whole should be substantially revised and shortened to focus less on your own findings and more comparing/contrasting elements of your findings with other work.
Response 7
We have substantially revised and shortened the "Discussion".

Please see "Discussion":
-page 13, line 10-13
-page 15, line 15 – page 16, line 6
-page 17, line 6-9
-page 18, line 13-14
-page 18, line 28 – page 19, line 2

VERSION 2 – REVIEW

REVIEWER	Weaver, Matthew Harvard Medical School I am partially supported by a grant to study physician work hours from the Centers for Disease Control and Prevention (CDC) National Institute for Occupational Safety and Health (NIOSH).
REVIEW RETURNED	04-Apr-2023

GENERAL COMMENTS	Thank you for addressing each of the concerns identified in the initial review. I suggest refining the strengths and limitations of the study consistent with the edits elsewhere in the paper. In particular, that the assessment was the same at both timepoints.

VERSION 2 – AUTHOR RESPONSE

To Reviewer: Dr Matthew Weaver, Harvard Medical School

The authors wish to thank Dr Matthew Weaver for his valuable suggestions. The manuscript has been revised along the indicated lines.

Comment 1

Thank you for addressing each of the concerns identified in the initial review. I suggest refining the strengths and limitations of the study consistent with the edits elsewhere in the paper. In particular, that the assessment was the same at both timepoints.

Response 1

The "Strengths and limitation" has been revised along the suggestion.

Please see "Strengths and limitations", page 18, line 8-9 and line 13-14.

"… There were similarities in survey methods and key items on working hours at both points in time…."

"… the majority of doctors answered at both points in time (80% in 2016 and 88% in 2019 - see Respondents), the changes in work hours documented in the study should be reliable…."