

Work time reduction via a four-day workweek finds improvements in workers' well-being

Corresponding Author: Dr Wen Fan

A version of this paper was originally rejected for publication by Nature Human Behaviour, however that decision was reconsidered after appeal by the authors.

Version 0:

Decision Letter:

6th March 2024

Dear Dr Fan,

Thank you once again for your manuscript, entitled "Does work time reduction improve workers' well-being? Evidence from global four-day workweek trials," and for your patience during the peer review process.

Your manuscript has now been evaluated by 4 reviewers, whose comments are included at the end of this letter. In the light of their advice, I regret that we cannot offer to publish your manuscript in Nature Human Behaviour.

While the reviewers find your work of some interest, they raise concerns about the appropriateness of the design and the soundness of the technical approach. We feel that these reservations are sufficiently important as to preclude publication of this work in Nature Human Behaviour.

I am sorry that we cannot be more positive on this occasion but hope that you will find our reviewers' comments helpful when preparing your paper for submission elsewhere.

Sincerely,



Nature Human Behaviour

Reviewer expertise:

Reviewer #1: labour economics, working and life satisfaction, well-being

Reviewer #2: sleep and stress, working time, health

Reviewer #3: organisational psychology, behavioural economics, work time and motivation

Reviewer #4: sociology of time, cultural sociology, working hours

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

I have carefully reviewed the manuscript, and while I find the research question raised to be highly significant and the research project ambitious, there are several critical issues that I believe need to be addressed before the manuscript can be considered for publication.

1) The manuscript's initial 13 pages appear to oversell the results and interpret them as causal. However, there is a glaring

absence of a control group, and consequently, no evidence is presented to support the assertion that the trial itself led to improvements in wellbeing. It's plausible that any observed improvements are simply a result of the passage of time rather than the intervention itself.

2) Much of the identification strategy relies on comparing workers who reduced their hours by different increments (e.g., 8 hours or more, 5 to 7 hours, 1 to 4 hours) with those who did not change their working time. However, it's unclear where these between-group variations originate. In a scenario where all firms switch to a four-day workweek trial, one would expect all workers to experience a reduction of 20% in their working time. The existence of groups of workers with no change in their working hours raises questions about the methodology.

3) The fact that there are workers who did not change their working time raise many concerns. Who are these workers? If they did not participate to the trial, this would introduce issues of self-selection into the treatment. If they participated to the trial but still work the same amount of time, then the coefficients in the first three lines of Table 3 just capture the effect of working time reduction but are missing the effect of working four days, which would limit the contribution of the paper.

4) A comprehensive description of the estimation sample and the distributions of key variables is absent from the manuscript. Providing such details is essential for readers to understand the robustness of the analysis.

5) The authors assert on page 14 that "the choice to join a trial is made by senior management," suggesting that self-selection bias is attenuated. However, it's worth noting that senior managers, especially in smaller companies, often consult with employees before making such decisions. Therefore, the extent to which self-selection bias is mitigated remains uncertain. Furthermore, while the authors acknowledge the limitations of their recruitment procedure, they should explicitly acknowledge that self-selection into the trial likely biases the results towards detecting an upper bound rather than the average true effect of a four-day workweek.

Reviewer #2:

Remarks to the Author:

This is a timely paper related to a very relevant question: can a four-day workweek (with retained salary) improve well-being, including subjective health. The paper is based on a large intervention study, and the results show that well-being improves after implementing reduced weekly working hours. The empirical findings support earlier research on working time reduction, namely introducing the 6-hour working day. The study has some strengths, such as a large study sample, and the manuscript is well-written and the empirical results are clearly presented.

However, the study also has some limitations that I have summarized below.

1. Design of the study: the authors acknowledge that the lack of a control group is a limitation. There are several studies related to 6-hour working days (30 hour working week) that could include a control group. I can understand that it is difficult to randomize work sites to working time reduction or control conditions, but it would be possible to recruit work sites that continue to work as usual (≈ 40 hours/week). Since there is no control group it is difficult (I would say almost impossible) to draw the conclusion that the changes are solely related to the intervention.

Another limitation is the short follow-up time, only 6 months. Previous research has usually had a much longer follow-up time (at least a year). It cannot be ruled out that the positive effect gradually diminishes (especially if work stress/intensity increases as a result of the employer not hiring more staff). Another problem with having a 6-month follow-up time is that the measurement waves differ with respect to season. It is quite likely that well-being can differ depending on the season (and the study design would have been stronger if the follow-up was carried out in the same season as the pre-intervention measurement).

2. Selection bias: the authors acknowledge this bias, that is the participating companies were self-selected into the intervention. The self-selection could have a potentially strong impact. For example, it might be that companies with strong economy are over-represented in the study. A strong economy could mean that the company could hire more employees to compensate for the loss of man-hours (capacity). There can of course also be other factors explaining the selection bias, such as a greater interest in the employees' psychosocial work environment. As a consequence, it is difficult to draw conclusions of whether the obtained results are representative for implementing a 4-day working week on a large scale.

3. The outcome data: the outcome data is only based on self-reports. This is understandable to some extent – and subjective well-being should not be overlooked – but it also creates problems. For example, it might be that the employees' responses are influenced by the trial itself and that the employees would like to keep the 4-day working week (and therefore don't want to express any negative opinions related to the intervention). The quality of the paper would be strengthened if actual working time data were used instead of self-reported data on weekly working hours – and to me it seems doable to record such data via the companies HR-systems (and it has been done in previous research).

Another concern related to the outcome variables is that the instruments (indicators) that were used has not been validated. Several of the instruments are very brief – for example, mental and physical health are measured with single items – and the quality of such data can be questioned since the questions don't capture the full complexity of mental and physical health. Burnout, which was measured with seven questions, consists according to Maslach (one of the pioneers in research about burnout) of three components, namely exhaustion, depersonalization and personal accomplishment. It is unlikely that the Copenhagen Burnout Inventory (that was used in the present study) could capture the full complexity of the burnout concept. Moreover, it is possible that the CPI-instrument mainly measured the exhaustion component, and if this is the case it would be better to label the variable "exhaustion". The consequence of using non-validated measures is that the precision and accuracy becomes low, and the obtained results might be biased.

4. Mediating variables: descriptive results should be presented for the mediating variables (similar as table 2). It is difficult to interpret the mediation analyses without having information of the study sample's means and SD for these variables (including both pre- and post-measurement). I also feel that the conclusions based on the mediation analysis is too strong. It is recommended to have at least three measurement waves in order to obtain a reliable and accurate analysis of mediation effects.

5. Increased staffing when implementing reduced working hours: it is unclear if the participating companies increased staffing during the intervention. To maintain the same productivity without employing more staff would lead to increased work intensity and higher work stress. The paper includes a question of "work smarter", but I seriously doubt that working smarter can compensate for an up-to 8 hours reduction in weekly working hours. For example, in human service occupations (such as health care), "working smarter", would probably have a very small effect on productivity. Introducing reduced working times, without employing more personnel, could in many occupations lead to worse well-being and I strongly recommend that this aspect is discussed when the results are interpreted. It would also make the paper stronger if the authors could report if increased staffing occurred or not on the company level.

6. Theory: the authors used the demand-job resource theory, which according to my view is sub-optimal. A more relevant theoretical model is the "effort recovery model" developed by Sonnentag and others. According to the ER-model one free day more per week would increase the recovery opportunities, which is likely to improve well-being. The ER-model is widely used in work psychology and it was surprising that this model was not mentioned in the introduction.

7. Health effects of long working hours: I recommend the authors to read a few recent systematic reviews and meta-analyses regarding the health consequences (see examples of references below). Long working hours is usually defined as +55 hours/week. The paper by Ervasti et al (2021) presents a nice table of long working hours and risk of 50 health conditions – and 9 out of 50 show a significant association. Hence, many health outcomes such as cancer were not associated with long working hours. Furthermore, the evidence became weaker for working only +40 hours per week. It is still an open question if a reduction of weekly working hours from 40 to 30-35 hours has any impact on health and mortality. Thus, the statement on page 3 (line 66-68) "....., consistently finding that long working hours are associated with poor health outcomes such as the prevalence of heart disease and stroke, cancer, chronic pain, and psychological distress" should be rephrased and presented in a more balanced way. Also, it is debated whether long working hours is an independent risk-factor or if the association with health outcomes is determined only for high work stress individuals (occupations). This might also be true for reduced working hours and health: it might be employees with high work stress that benefits of having shorter working days or 4-day working week.

Ervasti J, Pentti J, Nyberg ST, et al. Long working hours and risk of 50 health conditions and mortality outcomes: a multicohort study in four European countries. *Lancet Regional Health – Europe*, 2021;11: 100212

Pega F, et al (see ref no 25 in the paper).

Virtanen M, Jokela M, Madsen IEH, et al. Long working hours and depressive symptoms: a systematic review and meta-analysis of published and unpublished individual participant data. *Scand J Work Environ Health*, 2018; 44: 239-250.

8. A comment on the presentation of the results: the results are presented as means, which is OK. However, is the change in average well-being of significance? For example, how many participants showed a change of importance for the outcomes? Importance or significance could be defined according to a cut off score, or 2-3 step change score.

9. Finally, a minor comment: I don't see the point of using statements such as "ground-breaking research". The study has several limitations and more research is needed before it is possible to draw firm conclusions about the effects of reduced working hours based on introducing a 4-day working week. I agree that this is the first large-scale study on 4-day working week, but that is not the same as doing "ground-breaking research".

I thank the editor for giving me the opportunity to evaluate the intervention study.

Reviewer #3:

Remarks to the Author:

NATHUMBEHAV-23124238

Key result:

The present research is the first to examine, in a large scale intervention across multiple companies and industries, how reduction in work time affects workers' well-being. There is much to like about this research. The paper is well written, the data collection is impressive, and the topic is interesting and important. Understanding how reduction in work time affects well-being is topical and of value – most will be interested in the findings of this research.

Although some of the findings are conceptually replicating results of smaller scale interventions, this paper is the first to test these predictions on a larger scale and across multiple companies/organizations, giving it greater power to test for potential moderators and different factors that could influence the results. In addition, this paper extends beyond prior work by examining specific mediators, including the relationship between work hours and work ability, improved sleep, and reduced fatigue.

1 Validity

1a The biggest flaw is the absence of a control condition. The authors use paired sample t-tests to compare pre vs. post intervention within the same organization. Possibly, results can be strengthened by including period data, or utilizing the different times in when the intervention was rolled out to serve as a control or "onset" of the intervention. The authors also appear to have data in the middle of the intervention that they don't utilize in their analysis.

1b In addition to a lack of control condition, it is not clear that the reference group the authors are using is appropriate. For their main analysis, they use as a reference point companies with "no change in hours or an increase in hours." All companies that were included in this intervention were meant to reduce work hours, so why are there some companies with no change (or an increase in work hours)? It seems these companies may be outliers / may vary on other factors that could lead them to be worse off (e.g., employees at these companies may have believed their hours would be cut, but in reality they were not cut or were increased). A better analysis may be to focus on a greater reduction in work hours (8+) compared to a lower reduction in work hours (1-4). The supplemental materials (Table S4) suggests that when comparing to no change in hours, results don't hold for physical well-being and increased hours increase burnout, reduce job satisfaction, affect, and mental health, and marginally harm physical health.

1c The authors suggest that a reduction in work hours is linearly related to improved well-being. However, there is research suggesting a curvilinear relationship between free time and wellbeing (Sharif et al., 2021). Presumably too much time off can come at a cost.

2. Originality and significance

The paper is very clear on what is new and what has already been done in this space. Prior work has found improvements in work time reduction, but this has been limited in scope. A benefit of the current research is testing that in a larger and more robust way. I believe this represents a significant advancement in the field, but leave this to the editor to decide. This conclusion is due in part to my subjective evaluation that these results would be of immediate interest to many people in my own / related disciplines. In particular, I believe this paper contributes in two ways (1) applied / societal / policy related advance and (2) evidence based advance, which offers a contribution in the form of increased scale/rigor superseding existing literature.

3 Data and Methodology

3a As I mentioned above, the biggest limitation of this research is a lack of a control condition. It is possible that companies that self-select into this type of intervention, wherein they continue to pay workers for a 5-day work week, but expect only 4-days of work in return, are very different from other types of companies. It is possible that improvements observed are due to these differences – the fact that a company would take this initiative, rather than the initiative itself. For example, companies that engage in this behavior are sending a strong signal that they care about worker well-being. Employees may feel that they should prioritize their well-being as a result (e.g., due to perceived support from managers). Managers need to really trust their employees to implement this change (i.e., reduce work time, without reducing pay). To what extent are these results related to changes in the job itself, rather than changes in hours worked. Presumably, cutting out boring meetings/busy work will make people enjoy their job more, which helps with persistence (e.g., Giurge & Woolley, 2022) and is related to well-being.

3b. A limitation is that the data relies on self-report. For example, they observe people saying their physical health has improved from baseline to end, but they do not have objective data on this.

4 Preregistration

N/A (research was not pre-registered to my knowledge)

5 Appropriate use of statistics and treatment of uncertainties:

5a I believe the approach the authors used is appropriate. However, I believe the authors should report as their main results the analysis controlling for time period. I do not believe the paired sample t-test is sufficient as it is missing a lot of information. I would like to see the authors test different comparisons (i.e., higher vs. lower reduction in work time) rather than rely on the comparison with no change + increase in work hours.

5b The authors measured a small subset of well-being measures after three months that I did not see analyzed (unless I missed this). It would be interesting to see how baseline compares to three months and three months to six months.

6 Custom code:

I accessed the code but did not run it as I do not have access to Stata

7 Conclusions:

It would help to have a discussion of what types of jobs are most likely to benefit from a reduction in work time. The authors focus on benefits to well-being in a company where everyone in the organization reduces their work hours. How might this play out when individuals are striving for a promotion (e.g., jobs where people are climbing the ladder; trying to make partner at a law firm).

How does change in job (e.g., increased enjoyment or intrinsic motivation at work) relate to these findings?

8 References: Does this manuscript reference previous literature appropriately? If not, what references should be included or excluded?

Three references that are relevant here:

- Shin, J., & Grant, A. M. (2019). Bored by interest: How intrinsic motivation in one task can reduce performance on other tasks. *Academy of Management Journal*, 62(2), 415-436. >>> This paper found that people have certain parts of their job that they like more than others. In the current research, might people be cutting out the parts of their job they don't like, increasing enjoyment of work and well-being as a result.
- Giurge, L. M., & Woolley, K. (2022). Working during non-standard work time undermines intrinsic motivation. *Organizational Behavior and Human Decision Processes*, 170, 104134.>>> This work demonstrates the importance of working standard (vs. non-standard) work time and relates to the question about days off vs. hours worked. If people are able to take time off and coordinate with others taking a standard work week, this research suggests that could have benefits for work performance and motivation
- Sharif, M. A., Mogilner, C., & Hershfield, H. E. (2021). Having too little or too much time is linked to lower subjective well-being.

Journal of Personality and Social Psychology, 121(4), 933.>>> this paper finds that the effect of free time on wellbeing is curvilinear – having too little free time is bad, but having too much free time is also bad (e.g., being too busy vs. too bored). Presumably, reduced work hours benefit well-being in the current study because they do not give people too much free time.

9 Clarity and context:

9a The abstract was clear and engaging. The abstract, introduction, and conclusions were appropriate. One thing that could use clarifying in the front end is # of employees and organizations. The abstract stated 2,134 employees and 123 organizations. The paper mentions that trials have involved 211 companies and 7,200 employees, which could be confusing to readers as this is not reflective of the dataset the authors analyze. Similarly, the authors later write that “four out of five” companies gave one full day off, which is confusing as it makes it seem like they only had data for five companies.

9b Is this a paper on the effects of a four-day work week or on the effects of work time reduction? That is, what is the primary IV of interest? The front end and abstract make this seem like it is about a four day work week trial, but the data and methods are more about hours worked per week, not necessarily days worked per week. The authors have information on whether people had a three day weekend that may be able to address this point about days off vs. hours worked. This is important as some professions people work fewer days but very long hours (e.g., nurses in the medical profession may work three 12-hour shifts a week).

Other comments;

- How does intensity drop but productivity increase or stay the same when working fewer hours?
- What about the role of hedonic adaptation – will people adapt to the 32 hour work week, erasing these benefits? How much are these results due to reflecting on the comparison with the 40 hour work week? How much is this about what is happening at work (e.g., change in job demands) vs. outside of work (e.g., having more free time to take care of oneself)?
- I believe the authors may be able to use more sophisticated models to control for time effects that they are not currently doing; however I am not an expert on those models.

I wish the authors best with this fascinating line of research.

Reviewer #4:

Remarks to the Author:

The paper reports the impact of a shorter workweek on workers' well-being. It uses unique, large scale data of workers over many different organisations in different countries. This is an important study in showing the large-scale impact of a working time reduction with wage retention.

I do have some remarks that I would like to see addressed:

- I would like to see a little more context on the setting of the trials. They took place in Anglo-Saxon countries known for long work hours and strong neoliberal cultures. After COVID-19, it might be no surprise that it is exactly in these countries that most resistance to work as they knew it grew?

On page 2 is stated that also companies from Continental Europe and South Africa participated. However, later on it seems that only Ireland, US, Canada, Australia, NZ and UK are included in the data analysis. This is not explained well in the text.

- The paper is very ambitious in that it includes many measures and different types of analyses, yet this is also a pitfall. There is little room for nuance and I would like to see a little more of this:

o Make more of the difference between the groups of decreased hours (≥ 8 hours, 5-7 hours, 1-4 hours). I think this is an important aspect that the paper should address more: do all decreases matter to well-being or is it only the large decreases. The paper points out that there is a difference, but interesting might be to test the significant differences between groups as well.

Also include the number of respondents included in the decreased work hours groups.

o The mediating variables are interesting to use and shine a light on how a shorter workweek impacts well-being. However, the organizational changes that were made in many of the companies might also contribute to those mediating factors and these effects cannot be separated. I would like to see somewhat more reflection on this.

o Make more of the difference between well-being indicators. 5 indicators are included and there is some difference between these, some are job related, others are more general. I think there is room for nuance between these.

o I wonder if it might be interesting to include the start point values of the mediating variables in the analyses of Table 3? In addition to the change over time.

- The limitations section can include some more reflection on the Hawthorne effect and hedonic adaption, especially since the trial was only six months it is hard to know how this would change in the long run.

Smaller comments:

- p 5, line 131: 'low- and zero-productivity activities': can you give more examples of these?

- P 13, line 354: work relatedness also seems unaltered by reductions (not really significant)

- Table 2: include N

Concluding:

The paper is an interesting broad study on the impact of an organisational working time reduction on workers' well-being. Although it might need some more nuancing here and there, it does give a good first general impression of the impact of these trials. In the future, the inclusion of the qualitative findings and more in-depth focus on one aspect will be very valuable as well. I am looking forward to reading more on these trials.

**Following suitable revisions, you may want to consider transferring your manuscript. Although we cannot offer to publish your manuscript, I suggest that you consider Humanities and Social Sciences Communications as a suitable venue for this work. To transfer your manuscript, please use our [manuscript transfer portal](#). You will not have to re-supply manuscript metadata and files, unless you wish to make modifications. For more information, please see our [manuscript transfer FAQ](http://www.nature.com/authors/author_resources/transfer_manuscripts.html?WT.mc_id=EMI_NPG_1511_AUTHORTRANSF&WT.ec_id=AUTHOR) page.

Version 1:

Decision Letter:

21st January 2025

Dear Dr Fan,

Thank you once again for your revised manuscript, entitled "Does work time reduction improve workers' well-being? Evidence from global four-day workweek trials," and for your patience during the re-review process.

Your manuscript has now been evaluated by Reviewers 1, 2, and 4 from the original round of review. All reviewer feedback is included at the end of this letter. Although the reviewers found your manuscript to have improved during revision, they also raise some important outstanding concerns. We remain interested in the possibility of publishing your study in Nature Human Behaviour, but would like to consider your response to these outstanding concerns in the form of a revised manuscript before we make a decision on publication.

In particular, Reviewers 1 and 4 both raise questions regarding the statistical approach. We would ask you to perform the recommended analysis, including adding an omnibus test rather than pairwise comparisons only.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

In sum, we invite you to revise your manuscript taking into account all reviewer and editor comments. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

We hope to receive your revised manuscript within 4-8 weeks. I would be grateful if you could contact us as soon as possible if you foresee difficulties with meeting this target resubmission date.

With your revision, please:

- Include a "Response to the editors and reviewers" document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be used by the editors and reviewers to evaluate your revision.
- Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Sincerely,


Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #1: labour economics, working and life satisfaction, well-being

Reviewer #2: expertise: C; sleep and stress, working time, health

Reviewer #4: sociology of time, cultural sociology, working hours

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

I would like to commend the authors for their thorough and thoughtful responses to my previous comments, as well as those of the other reviewers.

I have one comment regarding the nature of the control group. In the current version, the authors used a sample of companies that were willing to implement working time reductions. It is unlikely that these companies are fully representative of the broader population of firms, as they may possess unique characteristics (e.g., a specific workforce composition).

I recognize that the authors have likely made the best possible use of the available data, and I do not expect them to include random firms that might be unwilling to adopt working time reductions. Nonetheless, the paper would benefit from a more in-depth discussion of the extent to which the control firms are representative of the wider business landscape. Such a discussion would strengthen the manuscript and provide valuable insights for policymakers.

Reviewer #2 (Remarks to the Author):

The quality of the manuscript has improved, in particular the interpretation of the results is more balanced. The manuscript contains a solid presentation of results (5 tables, 7 tables in supplement and 3 figures) and the editor will have to decide if some tables can be moved to supplement. I have three minor comments, and one major comment. I start with the major comment.

1. I think that the statistical analysis including the control group (table 2) should be based on a statistical model including group (intervention group vs control group) and time (baseline vs endpoint). The key statistical test for a significant effect of the intervention is the interaction between group and time. An interaction analysis means that only one test per outcome variable is needed, which reduces the risk of type 1 errors (compared with the t-test analysis). Thus, I recommend that the analysis based on t-tests are replaced with a "real" interaction-based analysis.

2. It is a strength that the study design has included a control group. However, the control group is very small and represents about 10% of the participants in the study. In addition, the control group is collected later than most other trial groups in the study. These shortcomings should be mentioned as a limitation in the discussion.

3. The third comment is about conflict of interest. Do the authors have any formal connection or link with 4DWG group? This connection should be described (or have I read carelessly and missed this paragraph).

4. Finally, the abstract should end with a general conclusion, preferably related to the question related to the title.

Reviewer #4 (Remarks to the Author):

As mentioned in the previous review, the paper's strength is its the unique, large scale data of workers over many different organisations in different countries. This is an important study in showing the large-scale impact of a working time reduction with wage retention.

The paper has been adapted quite heavily although the original aim remains. The challenge with a paper like this, including many analyses, variables and levels is keeping it orderly and clear. I think the paper may want to show too much and loses focus here and there. However, I do think the authors did a great job trying to summarize the findings in the discussion section.

I still believe the paper adds important insights to the literature on working time reduction and do think it is fit for publication if the following points of feedback are considered:

Introduction/literature review

- P6 line 114-121: This part seems to fit better in the result section as it does discuss empirical findings of non-published research.
- P. 6-8: consider moving 'Hypothesis 2' to right below line 130
- P. 6-8: I think the structure of this text could be improved and made more clear to clearly state what you are talking about in each paragraph and outlining this before the start of the three paragraphs on (1) company level changes, (2) work experiences and (3) health behaviors.

I also think Hypothesis 3 is too broad. Could you be more specific in what work experiences and what health behaviors you think might be enhanced?

- I know the space is limited, but I miss some differentiation between wellbeing indicators. What does wellbeing entail exactly in this paper and why were these indicators chosen? How do you expect shorter working hours to impact these indicators – do you expect the same improvement for all – why not?

Results:

- Can we have some more information on the control companies – what country are these based in?
- Can we have some more information on the data collected at 12 months? Were the employees still working less at that point?
- Line 267 should be 'less likely'?
- Be consistent in the vocabulary that you use. What does health behaviors entail and why is it called like that?
- How big are the different work hours groups? And how big is the reference group of control companies/no reduction? Are there many employees part of trial companies but had no reduction in hours? Maybe I missed this information, but it seems strange to me to include both groups in the reference group, as these are very different.
- I am not an expert in multilevel analyses, but I do wonder if the two level model is necessary here?
- I appreciate the attention paid to differences between work hours groups in the analyses, yet I would like to read more about this in the discussion. Maybe it is more interesting to compare these, as I have doubts about the references group used now (see above).

A paper that might be interesting and similar, about a longitudinal working time reduction and its impact on three domains of wellbeing, is the following: Mullens, F., & Laurijssen, I. (2024). An organizational working time reduction and its impact on three domains of mental well-being of employees: a panel study. BMC public health, 24(1), 1727. Although with its limited numbers and only one company differs in various aspects from this study.

Version 2:

Decision Letter:

Our ref: NATHUMBEHAV-23124238B

16th February 2025

Dear Dr. Fan,

Thank you for submitting your revised manuscript "Work time reduction via a four-day week finds improvements in workers' well-being" (NATHUMBEHAV-23124238B). After a careful consideration, We are happy in principle to publish it in Nature Human Behaviour, pending minor revisions to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements within three weeks. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Please do not hesitate to contact me if you have any questions.

Sincerely,



Nature Human Behaviour

Version 3:

Decision Letter:

Dear Dr Fan,

We are pleased to inform you that your Article "Work time reduction via a four-day workweek finds improvements in workers' well-being", has now been accepted for publication in Nature Human Behaviour.

Authors may need to take specific actions to achieve [a href="https://www.springernature.com/gp/open-research/funding/policy-compliance-faqs">compliance](https://www.springernature.com/gp/open-research/funding/policy-compliance-faqs) with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to [a href="https://www.springernature.com/gp/open-research/funding/policy-compliance-faqs">compliance](https://www.springernature.com/gp/open-research/funding/policy-compliance-faqs) with funder and institutional open access mandates).

<https://www.springernature.com/gp/open-research/plan-s-compliance>>Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including <https://www.springernature.com/gp/open-research/policies/journal-policies>>self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see <http://www.nature.com/nathumbehav/info/gta>). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at <https://www.nature.com/reprints/author-reprints.html>><https://www.nature.com/reprints/author-reprints.html>. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Human Behaviour as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Human Behaviour logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

We look forward to publishing your paper.

With best regards,



Nature Human Behaviour

P.S. Click on the following link if you would like to recommend Nature Human Behaviour to your librarian
<http://www.nature.com/subscriptions/recommend.html#forms>

** Visit the Springer Nature Editorial and Publishing website at http://editorial-jobs.springernature.com?utm_source=ejp_NHumB_email&utm_medium=ejp_NHumB_email&utm_campaign=ejp_NHumB>[www.springernature.com/editorial-and-publishing-jobs](http://editorial-jobs.springernature.com?utm_source=ejp_NHumB_email&utm_medium=ejp_NHumB_email&utm_campaign=ejp_NHumB) for more information about our career opportunities. If you have any questions please click [here](mailto:editorial.publishing.jobs@springernature.com).

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source. The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

November 26, 2024

We thank the editor for the opportunity to submit an appeal regarding our paper, “Does work time reduction improve workers’ well-being? Evidence from global four-day workweek trials.” The reviewers’ comments provided very helpful guidance. All the comments and suggestions have been seriously considered and addressed. We believe these revisions have resulted in a significantly improved manuscript that is suitable for publication in *Nature Human Behaviour*. To facilitate reading, we underline our changes in the manuscript.

[Reviewer 1]

I have carefully reviewed the manuscript, and while I find the research question raised to be highly significant and the research project ambitious, there are several critical issues that I believe need to be addressed before the manuscript can be considered for publication.

1. The manuscript’s initial 13 pages appear to oversell the results and interpret them as causal. However, there is a glaring absence of a control group, and consequently, no evidence is presented to support the assertion that the trial itself led to improvements in wellbeing. It’s plausible that any observed improvements are simply a result of the passage of time rather than the intervention itself.

Response: We appreciate the reviewer’s comments regarding causality. We agree that having a control group would considerably strengthen the evidence. Since submitting our initial manuscript, we have collected data from 12 control companies that attended an information session about the trials but did not join a trial. Similar to the trial organisations, employees in these control organisations completed baseline and endpoint surveys six months apart. Importantly, these employees are unaware of the link between these surveys and the four-day workweek trials; the surveys were presented to them as a way to assess their well-being.

Results from these companies show that their work hours remained largely unchanged, and their well-being did not improve over the six months. By contrast, there was a significant reduction in work hours and a significant improvement in employee well-being among trial companies. We discuss these findings on pages 10–11. All our analyses now use the control companies as the reference group to allow for stronger causal inference.

As discussed in the manuscript (page 18), it is extremely difficult to locate companies willing to randomly assign employees to trial versus control conditions. Even when companies are willing to do so, the relatively small size of many companies in our sample would constrain the actual implementation of four-day week trials that involve changes at the organisational level. The next best option, we believe, is to compare the trial and control companies in terms of their respective pre-post changes in employee well-being, which resembles a difference-in-differences design and provides stronger evidence for causal inference than relying on trial companies alone. This also addresses the concern you rightly pointed out, that the effect we identified might simply reflect the passage of time. We have added extra information on pages 9–10 to ensure the reader fully understands the nature and limitations of our research design. We also call for future research to

investigate, when such opportunities arise, a group randomised trial on shorter workweeks to continue accumulating rigorous evidence on this topic (page 18).

2. Much of the identification strategy relies on comparing workers who reduced their hours by different increments (e.g., 8 hours or more, 5 to 7 hours, 1 to 4 hours) with those who did not change their working time. However, it's unclear where these between-group variations originate. In a scenario where all firms switch to a four-day workweek trial, one would expect all workers to experience a reduction of 20% in their working time. The existence of groups of workers with no change in their working hours raises questions about the methodology.

Response: As noted above, our methodology has been revised to include a control group, which adds to the methodological rigour of our design. Part of our analysis still relies on comparing workers in trial companies with varying reductions in hours (relative to those in control companies). Therefore, we appreciate the reviewer raising the issue of noncompliance. We are sure the reviewer is aware; noncompliance is common in intervention studies—researchers or management cannot compel participants to follow their assigned treatment. Rather than being a methodological weakness, this is representative of the real-world conditions under which such interventions are adopted.

We conducted multinomial logit regression models to examine differences across employees whose hours changed to varying degrees, and we found that certain workers such as women, Australian and New Zealand respondents, or employees in larger companies are less like to reduce their hours. However, reductions in hours are not dependent on age, college attainment, supervisor status, remote work status, or permanent contract status. These results are now discussed on pages 12–13. We choose to include employees regardless of their actual changes in hours, as variations in hours are what would realistically occur as more companies adopt four-day workweeks. We now clarify that this set of analysis approximates an “as-treated” approach, using the “treatment” respondents followed rather than the group they were originally assigned to (page 10).

3. The fact that there are workers who did not change their working time raise many concerns. Who are these workers? If they did not participate to the trial, this would introduce issues of self-selection into the treatment. If they participated to the trial but still work the same amount of time, then the coefficients in the first three lines of Table 3 just capture the effect of working time reduction but are missing the effect of working four days, which would limit the contribution of the paper.

Response: As mentioned above, we have run multinomial logit regression models to examine the differential changes in hours (pages 12–13). We acknowledge the reviewer’s comments regarding the distinction between working four days and reducing working hours. We now first use the intention-to-treat (ITT) analysis to identify the effect of working four days by comparing trial and control companies. ITT analyses participants according to their original group assignments, regardless of whether they complied with the treatment protocol. This approach is commonly used to evaluate trials, as it maintains the integrity of the randomisation¹ (even though we recognise that ours is not a randomised trial). To reflect real-world conditions where noncompliance is common, we have also conducted an as-treated analysis based on the actual changes in work hours at both the employee and organisation levels. We have clarified our identification strategy on pages 9–10.

4. *A comprehensive description of the estimation sample and the distributions of key variables is absent from the manuscript. Providing such details is essential for readers to understand the robustness of the analysis.*

Response: We have included a description of the sample at the end of the manuscript (pages 20–21) to comply with the requirements of *Nature Human Behaviour*. The distribution of our key variables is shown in Table 2 and Supplementary Table 1.

5. *The authors assert on page 14 that "the choice to join a trial is made by senior management," suggesting that self-selection bias is attenuated. However, it's worth noting that senior managers, especially in smaller companies, often consult with employees before making such decisions. Therefore, the extent to which self-selection bias is mitigated remains uncertain. Furthermore, while the authors acknowledge the limitations of their recruitment procedure, they should explicitly acknowledge that self-selection into the trial likely biases the results towards detecting an upper bound rather than the average true effect of a four-day workweek.*

Response: We thank the reviewer for this comment. In response, we have added on page 19 that our findings likely overestimate the average true effect of a four-day workweek due to self-selection. Additionally, we conducted a supplementary analysis and found that our main results do not vary by company size (Supplementary Table 3, Panel 8). Therefore, selection based on company size is unlikely to bias our findings.

[Reviewer 2]

This is a timely paper related to a very relevant question: can a four-day workweek (with retained salary) improve well-being, including subjective health. The paper is based on a large intervention study, and the results show that well-being improves after implementing reduced weekly working hours. The empirical findings support earlier research on working time reduction, namely introducing the 6-hour working day. The study has some strengths, such as a large study sample, and the manuscript is well-written and the empirical results are clearly presented. However, the study also has some limitations that I have summarized below.

6. *Design of the study: the authors acknowledge that the lack of a control group is a limitation. There are several studies related to 6-hour working days (30 hour working week) that could include a control group. I can understand that it is difficult to randomize work sites to working time reduction or control conditions, but it would be possible to recruit work sites that continue to work as usual (≈ 40 hours/week). Since there is no control group it is difficult (I would say almost impossible) to draw the conclusion that the changes are solely related to the intervention.*

Response: We appreciate the reviewer's comments regarding causality, and the understanding of the challenges in randomising work sites. Since submitting our initial manuscript, we have collected data from 12 control companies that attended an information session about the trials but did not join a trial. Similar to the trial organisations, employees in these control organisations completed baseline and endpoint surveys six months apart. Importantly, these employees are

unaware of the link between these surveys and the four-day workweek trials; the surveys were presented to them as a way to assess their well-being.

Results from these companies show that their work hours remained largely unchanged, and their well-being did not improve over the six months. By contrast, there was a significant reduction in work hours and a significant improvement in employee well-being among trial companies. We discuss these findings on pages 10–11. All our analyses now use the control companies as the reference group to allow for stronger causal inference. To continue accumulating rigorous evidence on this topic, we also call for future research to investigate, when such opportunities arise, a group randomised trial on shorter workweeks (page 18).

7. Another limitation is the short follow-up time, only 6 months. Previous research has usually had a much longer follow-up time (at least a year). It cannot be ruled out that the positive effect gradually diminishes (especially if work stress/intensity increases as a result of the employer not hiring more staff). Another problem with having a 6-month follow-up time is that the measurement waves differ with respect to season. It is quite likely that well-being can differ depending on the season (and the study design would have been stronger if the follow-up was carried out in the same season as the pre-intervention measurement).

Response: We appreciate the reviewer’s comment regarding the sustainability of the well-being benefits, which was also a concern of ours. To address this, we continued collecting data from the more than 90% of organisations that chose to maintain the shorter schedule after the trials. The results, which examine changes at 12 months post-baseline, are provided in Supplementary Table 4 and discussed on pages 14–15. As shown, the well-being benefits are largely durable, suggesting that the findings are not simply due to seasonality or initial enthusiasm.

8. Selection bias: the authors acknowledge this bias, that is the participating companies were self-selected into the intervention. The self-selection could have a potentially strong impact. For example, it might be that companies with strong economy are over-represented in the study. A strong economy could mean that the company could hire more employees to compensate for the loss of man-hours (capacity). There can of course also be other factors explaining the selection bias, such as a greater interest in the employees’ psychosocial work environment. As a consequence, it is difficult to draw conclusions of whether the obtained results are representative for implementing a 4-day working week on a large scale.

Response: We agree that self-selection is a challenging issue to address, not just for our study but for most (if not all) intervention studies, as it is impractical to force companies to participate in a study if they are not interested in doing so to begin with. However, we believe there is still much to be learned from these trials, as they precipitated actual organisational changes, in contrast to much prior research that relies on naturally occurring changes in hours, which are susceptible to more methodological issues. We would also like to point out that not all organisations participated to improve worker well-being; many wanted to address retention and hiring issues, and some participated with management deeply divided on the four-day workweek trials.

Companies also embarked on these trials in a variety of macroeconomic conditions. In the UK and Australasia, the trials took place during difficult business periods. In the U.S., the economy has

gone from operating with high unemployment to a more robust state. As we mentioned in response to comment #6, we have also recruited control companies, which were exposed to the same economic environment as the trial companies. The different patterns identified for trial and control companies suggest that the findings are attributable to the trials, rather than the economy. In terms of hiring, we have data from the companies that show that hiring is reduced on average over the trial period (not shown in the paper), likely due to the decline in resignations that occurs after this schedule is implemented.

That said, we agree that selection remains an issue. We have added on page 19 that our findings likely overestimate the average true effect of a four-day workweek due to self-selection. As the idea of four-day workweeks continues to gain traction among a growing number of companies, we call for more future studies to provide as generalisable evidence as possible.

9. The outcome data: the outcome data is only based on self-reports. This is understandable to some extent – and subjective well-being should not be overlooked – but it also creates problems. For example, it might be that the employees’ responses are influenced by the trial itself and that the employees would like to keep the 4-day working week (and therefore don’t want to express any negative opinions related to the intervention). The quality of the paper would be strengthened if actual working time data were used instead of self-reported data on weekly working hours – and to me it seems doable to record such data via the companies HR-systems (and it has been done in previous research).

Response: We agree that it is important to evaluate the impact of the trials on objective measures. However, as you rightly pointed out, subjective well-being is significant in its own right. As the Thomas theorem suggests, “If men define situations as real, they are real in their consequences.” As such, evaluating subjective well-being holds significant value.

We understand the concern that employees might not provide valid reports in the hope of keeping the four-day week. To the extent that this occurs, the fact that they wish to maintain the trial already suggests they likely benefit from it. That said, the baseline and endpoint surveys are six months apart, making it difficult for employees to remember what they answered six months ago. Results are also durable over 12 months (see our response to Comment #7), a period for which recall is even less likely. Additionally, we do not find evidence that employees are systematically choosing the highest (or lowest) categories for the well-being measures. The patterns of responses suggest that employees are responding based on their actual circumstances rather than out of a desire to manipulate the data. For example, we asked a set of satisfaction questions—regarding household finances, relationships, time, and life as a whole. If respondents were systematically overreporting, we would not expect them to carefully distinguish across domains. However, the results show that satisfaction with time experienced the largest increase (2 points on a 0-10 scale), much larger than the other domains (0.5–0.8 points on the same scale). We now include this discussion on page 19.

Regarding weekly working hours, not every company maintains such records, and practices vary across industries and countries, making it difficult to obtain reliable data across the 153 companies in our data set from multiple countries. We have asked the companies for data on sick days and absences, but roughly half do not report it. Furthermore, because most of these employees are salaried, HR records will typically register their hours as 40 rather than reflecting actual hours

worked. HR records may also not capture the time employees spend working at home or in other locations. Lastly, given that we are evaluating *changes* in work hours, as long as any reporting bias remains constant within individuals, it should not pose a significant threat to our estimation.

10. Another concern related to the outcome variables is that the instruments (indicators) that were used has not been validated. Several of the instruments are very brief – for example, mental and physical health are measured with single items – and the quality of such data can be questioned since the questions don’t capture the full complexity of mental and physical health. Burnout, which was measured with seven questions, consists according to Maslach (one of the pioneers in research about burnout) of three components, namely exhaustion, depersonalization and personal accomplishment. It is unlikely that the Copenhagen Burnout Inventory (that was used in the present study) could capture the full complexity of the burnout concept. Moreover, it is possible that the CPI-instrument mainly measured the exhaustion component, and if this is the case it would be better to label the variable “exhaustion”. The consequence of using non-validated measures is that the precision and accuracy becomes low, and the obtained results might be biased.

Response: We appreciate this comment. Given the amount of information we aim to collect via surveys while also protecting response rate and quality, we took great care to rigorously operationalise well-being outcomes and keep the survey reasonably short, in line with best practice².

All the well-being measures included in our surveys are validated and, when possible, derived from large-scale national or cross-national surveys. The single-item health questions are standardized questions recommended by the World Health Organization. They are used in the European Health Interview Survey (EHIS) and EU Statistics on Income and Living Conditions (EU-SILC). The sources of these measures are listed in the Measures section (page 21).

We chose the Copenhagen Burnout Inventory (CBI) as it has been found to perform similarly to longer scales³. We believe that the scale captures a wider range of experiences beyond personal exhaustion through items such as “Does your work frustrate you?” We acknowledge that the Maslach Burnout Inventory (MBI) includes additional components not captured by the CBI. Therefore, we are open to renaming the measure “emotional exhaustion,” leaving the final decision to the Editor. However, we would like to note that MBI has limitations. It was originally developed in the U.S. for use in human service professions (e.g., healthcare, education), which may limit its applicability to non-service-oriented occupations in different cultural contexts. Additionally, the three subscales of MBI are often highly correlated, raising concerns about whether they represent truly distinct constructs³. In contrast, the CBI is designed to be applicable across a wide range of occupations and with a focus on cross-cultural relevance, which is crucial for our study involving multiple occupations across countries.

11. Mediating variables: descriptive results should be presented for the mediating variables (similar as table 2). It is difficult to interpret the mediation analyses without having information of the study sample’s means and SD for these variables (including both pre- and post-measurement). I also feel that the conclusions based on the mediation analysis is too strong. It is recommended to have at least three measurement waves in order to obtain a reliable and accurate analysis of mediation effects.

Response: We have provided descriptive results for the mediating variables in Table 2. As a robustness check, for a subset of the mediators, we used changes from the baseline to midpoint waves (page 15, Supplementary Table 5). The midpoint survey was intentionally kept short, as it included an intensive time diary instrument; therefore, not all mediators, such as work experiences, were assessed in that wave. Furthermore, control companies were not surveyed at the midpoint, so this set of sensitivity analyses is conducted among trial companies only.

12. Increased staffing when implementing reduced working hours: it is unclear if the participating companies increased staffing during the intervention. To maintain the same productivity without employing more staff would lead to increased work intensity and higher work stress. The paper includes a question of “work smarter”, but I seriously doubt that working smarter can compensate for an up-to 8 hours reduction in weekly working hours. For example, in human service occupations (such as health care), “working smarter”, would probably have a very small effect on productivity. Introducing reduced working times, without employing more personnel, could in many occupations lead to worse well-being and I strongly recommend that this aspect is discussed when the results are interpreted. It would also make the paper stronger if the authors could report if increased staffing occurred or not on the company level.

Response: We thank the reviewer for this comment, which has prompted us to include more discussion on the different ways organisations can benefit from four-day workweeks in terms of company performance. White-collar firms are more likely to improve productivity by reducing inefficiencies, while human service occupations such as health care or restaurants experience productivity gains and cost reductions from reduced attrition—a prevalent issue in these fields that negatively affects productivity. We have added discussions on this point on page 6.

As we responded to Comment #8, we have data from the companies showing that hiring is reduced over the trial period across companies. This is likely due to the decline in resignations that occurs after this schedule is implemented.

13. Theory: the authors used the demand-job resource theory, which according to my view is sub-optimal. A more relevant theoretical model is the “effort recovery model” developed by Sonnentag and others. According to the ER-model one free day more per week would increase the recovery opportunities, which is likely to improve well-being. The ER-model is widely used in work psychology and it was surprising that this model was not mentioned in the introduction.

Response: We have added the effort recovery model to the introduction (page 3), which we also use to support sleep and fatigue as potential mediators (page 8). We have chosen to retain the job demands-resources theory due to its strong connection with the job crafting literature, which is relevant for understanding the four-day week intervention (i.e., companies and employees are encouraged to optimise their work practices to meet changing job demands).

14. Health effects of long working hours: I recommend the authors to read a few recent systematic reviews and meta-analyses regarding the health consequences (see examples of references below). Long working hours is usually defined as +55 hours/week. The paper by Ervasti et al (2021) presents a nice table of long working hours and risk of 50 health conditions – and 9 out of 50 show

a significant association. Hence, many health outcomes such as cancer were not associated with long working hours. Furthermore, the evidence became weaker for working only +40 hours per week. It is still an open question if a reduction of weekly working hours from 40 to 30-35 hours has any impact on health and mortality. Thus, the statement on page 3 (line 66-68) “....., consistently finding that long working hours are associated with poor health outcomes such as the prevalence of heart disease and stroke, cancer, chronic pain, and psychological distress” should be rephrased and presented in a more balanced way. Also, it is debated whether long working hours is an independent risk-factor or if the association with health outcomes is determined only for high work stress individuals (occupations). This might also be true for reduced working hours and health: it might be employees with high work stress that benefits of having shorter working days or 4-day working weeks.

Ervasti J, Pentti J, Nyberg ST, et al. Long working hours and risk of 50 health conditions and mortality outcomes: a multicohort study in four European countries. Lancet Regional Health – Europe, 2021;11: 100212

Pega F, et al (see ref no 25 in the paper).

Virtanen M, Jokela M, Madsen IEH, et al. Long working hours and depressive symptoms: a systematic review and meta-analysis of published and unpublished individual participant data. Scand J Work Environ Health, 2018; 44: 239-250.

Response: We thank the reviewer for pointing us to the relevant references, which we have now added to the paper (pages 3–4). Additionally, we have revised the sentence in question to provide a more balanced perspective on the literature (page 3), including highlighting the previous focus on working more than 55 hours per week (page 4).

We consider the question of whether reduced working hours primarily benefit those with high work stress to be an empirical one. Accordingly, we have conducted a moderation analysis using baseline burnout as a moderator, distinguishing between those reporting a value of 4 or higher (out of 5) and others. Only 5 out of 24 interaction terms between reductions in hours and high baseline burnout are marginally significant (Supplementary Table 3, Panel 10), suggesting that the benefits of four-day weeks extend broadly, not just to those experiencing high work stress.

15. A comment on the presentation of the results: the results are presented as means, which is OK. However, is the change in average well-being of significance? For example, how many participants showed a change of importance for the outcomes? Importance or significance could be defined according to a cut off score, or 2-3 step change score.

Response: We thank the reviewer for this important point. We have calculated the effect size using Cohen’s d to assess the substantive importance of our findings (page 11).

16. Finally, a minor comment: I don’t see the point of using statements such as “ground-breaking research”. The study has several limitations and more research is needed before it is possible to draw firm conclusions about the effects of reduced working hours based on introducing a 4-day working week. I agree that this is the first large-scale study on 4-day working week, but that is not the same as doing “ground-breaking research”.

Response: Following the suggestion, we no longer use the term “ground-breaking” in the manuscript.

[Reviewer 3]

Key result: The present research is the first to examine, in a large scale intervention across multiple companies and industries, how reduction in work time affects workers’ well-being. There is much to like about this research. The paper is well written, the data collection is impressive, and the topic is interesting and important. Understanding how reduction in work time affects well-being is topical and of value – most will be interested in the findings of this research.

Although some of the findings are conceptually replicating results of smaller scale interventions, this paper is the first to test these predictions on a larger scale and across multiple companies/organizations, giving it greater power to test for potential moderators and different factors that could influence the results. In addition, this paper extends beyond prior work by examining specific mediators, including the relationship between work hours and work ability, improved sleep, and reduced fatigue.

17. Validity

a. The biggest flaw is the absence of a control condition. The authors use paired sample t-tests to compare pre vs. post intervention within the same organization. Possibly, results can be strengthened by including period data, or utilizing the different times in when the intervention was rolled out to serve as a control or “onset” of the intervention. The authors also appear to have data in the middle of the intervention that they don’t utilize in their analysis.

Response: We appreciate the reviewer’s comments regarding the absence of a control condition. Since submitting our initial manuscript, we have collected data from 12 control companies that attended an information session about the trials but did not join a trial. Similar to the trial organisations, employees in these control organisations completed baseline and endpoint surveys six months apart. Importantly, these employees are unaware of the link between these surveys and the four-day workweek trials; the surveys were presented to them as a way to assess their well-being.

Results from these companies show that their work hours remained largely unchanged, and their well-being did not improve over the six months. By contrast, there was a significant reduction in work hours and a significant improvement in employee well-being among trial companies. We discuss these findings on pages 10–11. All our analyses now use the control companies as the reference group to allow for stronger causal inference. We also control for period indicators (i.e., when the trial started, page 23). We do not use the midpoint survey data for simplicity and because the midpoint survey is intentionally brief, lacking coverage of all the measures needed for our analyses, such as mental and physical health. However, we provide the descriptive results for the baseline-to-midpoint and midpoint-to-endpoint changes for the available measures in Supplementary Table 6 (page 20).

b. In addition to a lack of control condition, it is not clear that the reference group the authors are

using is appropriate. For their main analysis, they use as a reference point companies with “no change in hours or an increase in hours.” All companies that were included in this intervention were meant to reduce work hours, so why are there some companies with no change (or an increase in work hours)? It seems these companies may be outliers / may vary on other factors that could lead them to be worse off (e.g., employees at these companies may have believed their hours would be cut, but in reality they were not cut or were increased). A better analysis may be to focus on a greater reduction in work hours (8+) compared to a lower reduction in work hours (1-4). The supplemental materials (Table S4) suggests that when comparing to no change in hours, results don’t hold for physical well-being and increased hours increase burnout, reduce job satisfaction, affect, and mental health, and marginally harm physical health.

Response: We are grateful for this comment, which has made us realize that we need to distinguish between company-level and individual-level changes in hours. Previously, we examined changes in work hours at the individual level, including a group whose hours did not reduce because of noncompliance. We have now conducted a multilevel analysis to examine the respective roles of company-level and individual-level changes in hours (page 10; pages 11–12). We have also performed pairwise comparisons, allowing different reference groups to be used for comparisons across groups with varying changes in work hours (page 12).

c. The authors suggest that a reduction in work hours is linearly related to improved well-being. However, there is research suggesting a curvilinear relationship between free time and wellbeing (Sharif et al., 2021). Presumably too much time off can come at a cost.

Response: Thank you for raising this important point. We have now mentioned on page 4 that the relationship between free time and well-being can be nonlinear and have added the Sharif et al. (2021) reference. Our use of a four-category variable to measure varying changes in work hours allows us to identify such nonlinear patterns (page 11).

18. Originality and significance

The paper is very clear on what is new and what has already been done in this space. Prior work has found improvements in work time reduction, but this has been limited in scope. A benefit of the current research is testing that in a larger and more robust way. I believe this represents a significant advancement in the field, but leave this to the editor to decide. This conclusion is due in part to my subjective evaluation that these results would be of immediate interest to many people in my own / related disciplines. In particular, I believe this paper contributes in two ways (1) applied / societal / policy related advance and (2) evidence based advance, which offers a contribution in the form of increased scale/rigor superseding existing literature.

Response: We greatly appreciate the reviewer’s positive assessment of our paper!

19. Data and Methodology

a. As I mentioned above, the biggest limitation of this research is a lack of a control condition. It is possible that companies that self-select into this type of intervention, wherein they continue to pay workers for a 5-day work week, but expect only 4-days of work in return, are very different from other types of companies. It is possible that improvements observed are due to these differences – the fact that a company would take this initiative, rather than the initiative itself. For

example, companies that engage in this behavior are sending a strong signal that they care about worker well-being. Employees may feel that they should prioritize their well-being as a result (e.g., due to perceived support from managers). Managers need to really trust their employees to implement this change (i.e., reduce work time, without reducing pay). To what extent are these results related to changes in the job itself, rather than changes in hours worked. Presumably, cutting out boring meetings/busy work will make people enjoy their job more, which helps with persistence (e.g., Giurge & Woolley, 2022) and is related to well-being.

Response: In response to your comment #17a, we have now presented results from the control companies, which had expressed interest in implementing a four-day week trial (evidenced by their attendance at the information sessions). Presumably, many of these companies also prioritize employee well-being. However, the well-being improvements observed among trial companies are not evident in these control companies. We would also like to point out that not all organisations participated to improve worker well-being; many wanted to address retention and hiring issues, and some participated with management deeply divided on the four-day workweek trials.

That said, we agree that selection remains an issue. We have added on page 18 that our findings likely overestimate the average true effect of a four-day workweek due to self-selection. As the idea of four-day workweeks continues to gain traction among a growing number of companies, we call for more future studies to provide as generalisable evidence as possible.

We also agree that changes in the job itself are part of the story, which is the rationale for our mediation analysis—to assess how a reduction in work hours improves well-being through changes in other aspects of the job. The finding that work ability (at both individual and company levels) is an important mediator indicates that four-day weeks have indeed improved workers’ sense of performing their jobs well, which in turn benefits their well-being. These points are added on pages 7–8 and pages 16–18. We have also included the Giurge & Woolley (2022) reference (page 8) and theorised enjoyment through the elimination of boring meetings as a mediator (job crafting, page 7).

b. A limitation is that the data relies on self-report. For example, they observe people saying their physical health has improved from baseline to end, but they do not have objective data on this.

Response: We agree that evaluating the impact of the trials on objective measures is important. Unfortunately, collecting such measures is costly and beyond the scope of this study, which we acknowledge as a limitation (page 19). We believe that subjective well-being is significant in its own right and the multiple domains of well-being measures used in our analyses strengthen our findings. As the Thomas theorem suggests, “If men define situations as real, they are real in their consequences.” In the case of self-reported physical health, prior research shows that it is a strong and reliable predictor of subsequent mortality⁴, indicating that individuals’ subjective assessments are useful indicators for evaluating objective statuses. We have added this information on page 2.

20. Appropriate use of statistics and treatment of uncertainties:

a. I believe the approach the authors used is appropriate. However, I believe the authors should report as their main results the analysis controlling for time period. I do not believe the paired sample t-test is sufficient as it is missing a lot of information. I would like to see the authors test

different comparisons (i.e., higher vs. lower reduction in work time) rather than rely on the comparison with no change + increase in work hours.

Response: Following the comments, we have now controlled for time period in all our regression analyses (page 23). When conducting paired *t*-tests, we present results for both trial companies and control companies, as well as for a subset of trial companies surveyed around the same time as the control companies to control for period effects (pages 10–11). We have also performed pairwise comparisons, allowing different reference groups to be used for comparisons across groups with varying changes in work hours (page 12).

b. The authors measured a small subset of well-being measures after three months that I did not see analyzed (unless I missed this). It would be interesting to see how baseline compares to three months and three months to six months.

Response: As mentioned in our response to your comment #17a, we did not initially analyse the midpoint measures for simplicity and because the midpoint survey is intentionally brief, lacking coverage of all the measures needed for our analyses. For the available measures, we have now included in Supplementary Table 6 the changes from baseline to midpoint (three months) and from midpoint to endpoint (six months).

21. Conclusions:

It would help to have a discussion of what types of jobs are most likely to benefit from a reduction in work time. The authors focus on benefits to well-being in a company where everyone in the organization reduces their work hours. How might this play out when individuals are striving for a promotion (e.g., jobs where people are climbing the ladder; trying to make partner at a law firm). How does change in job (e.g., increased enjoyment or intrinsic motivation at work) relate to these findings?

Response: We have examined whether the well-being findings vary across occupations, and few interaction terms between reductions in hours and occupation are significant (Supplementary Table 3, Panel 9), suggesting that the benefits of four-day weeks are widespread and not confined to workers in specific roles.

One goal of the trial is to combat the ideal worker norm that equates long hours with hard work and strong work ethics. Therefore, instead of using input (work hours) as a metric to evaluate performance, the trial encourages employees to shift their focus to output. In other words, the cultural change should incentivise those who want to climb the ladder to focus on results and identify more effective ways to do their jobs. Although we do not have direct measures of career drive to test this hypothesis, given that our findings do not differ across major socio-demographic groups (age, gender, race-ethnicity, educational attainment, tenure), the well-being benefits do seem to be broad. Our qualitative interviews with management and employees also indicate evidence of cultural change in the reward structure (not reported in the paper). Conversely, if the trial is not implemented at a collective level but applies only to a few employees, this might parallel the accommodation model, where the use of flexibility precipitates career panics and stigma. We have emphasised this point on page 18.

Changes in the job, such as increased enjoyment or intrinsic motivation, are relevant to our findings. Enjoyment is arguably an aspect of subjective well-being. Intrinsic motivation could be a psychological mediator that explains the relationship between a reduction in work time and improved well-being, as the references you generously provided suggest. Unfortunately, we do not have these measures, which we now note on page 18 as a promising direction for future research.

22. *Three references that are relevant here:*

- Shin, J., & Grant, A. M. (2019). *Bored by interest: How intrinsic motivation in one task can reduce performance on other tasks*. *Academy of Management Journal*, 62(2), 415-436. >>> *This paper found that people have certain parts of their job that they like more than others. In the current research, might people be cutting out the parts of their job they don't like, increasing enjoyment of work and well-being as a result.*

- Giurge, L. M., & Woolley, K. (2022). *Working during non-standard work time undermines intrinsic motivation*. *Organizational Behavior and Human Decision Processes*, 170, 104134. >>> *This work demonstrates the importance of working standard (vs. non-standard) work time and relates to the question about days off vs. hours worked. If people are able to take time off and coordinate with others taking a standard work week, this research suggests that could have benefits for work performance and motivation*

- Sharif, M. A., Mogilner, C., & Hershfield, H. E. (2021). *Having too little or too much time is linked to lower subjective well-being*. *Journal of Personality and Social Psychology*, 121(4), 933. >>> *this paper finds that the effect of free time on wellbeing is curvilinear – having too little free time is bad, but having too much free time is also bad (e.g., being too busy vs. too bored). Presumably, reduced work hours benefit well-being in the current study because they do not give people too much free time.*

Response: We thank the reviewer for directing us to the relevant references and their connections to our paper. We have added these works to our paper (page 7; page 8; pages 4 and 15).

23. *Clarity and context:*

a. *The abstract was clear and engaging. The abstract, introduction, and conclusions were appropriate. One thing that could use clarifying in the front end is # of employees and organizations. The abstract stated 2,134 employees and 123 organizations. The paper mentions that trials have involved 211 companies and 7,200 employees, which could be confusing to readers as this is not reflective of the dataset the authors analyze. Similarly, the authors later write that “four out of five” companies gave one full day off, which is confusing as it makes it seem like they only had data for five companies.*

Response: We have now mentioned only the number of companies and employees used in our analysis (page 1). Additionally, we have replaced expressions like “four out of five” with percentages to avoid any misunderstanding (page 6).

b. *Is this a paper on the effects of a four-day work week or on the effects of work time reduction? That is, what is the primary IV of interest? The front end and abstract make this seem like it is about a four day work week trial, but the data and methods are more about hours worked per week, not necessarily days worked per week. The authors have information on whether people had a three day weekend that may be able to address this point about days off vs. hours worked. This is*

important as some professions people work fewer days but very long hours (e.g., nurses in the medical profession may work three 12-hour shifts a week).

Response: We acknowledge the reviewer's comments regarding the distinction between working four days and reducing working hours. The primary reason for our previous focus on working hours is the lack of a control group to identify the effect of four-day weeks. With the inclusion of the control companies, we now first use the intention-to-treat (ITT) analysis to identify the effect of working four days by comparing trial and control companies. ITT analyses participants according to their original group assignments, regardless of whether they complied with the treatment protocol. This is the most commonly used method to evaluate trials, as it maintains the integrity of the randomisation and reflects real-world conditions where noncompliance is likely to occur¹. However, we have also conducted an as-treated analysis based on the actual changes in work hours at both the employee and organisation levels. We have clarified our identification strategy on pages 9–10.

In a supplementary analysis, we examined the effect of working a four-day week. We found that consistent with our main findings, employees on a four-day schedule experience significantly greater improvements across the four well-being outcomes than those not on the schedule (Supplementary Table 7). We do not include this analysis in our main models because having a four-day week is highly correlated with changes in work hours, which could lead to multicollinearity issues. Additionally, changes in work hours provide greater variation, allowing us to capture more nuanced relationships between work time and well-being than the dichotomous measure of having a four-day workweek.

24. Other comments

- How does intensity drop but productivity increase or stay the same when working fewer hours?*
- What about the role of hedonic adaptation – will people adapt to the 32 hour work week, erasing these benefits? How much are these results due to reflecting on the comparison with the 40 hour work week? How much is this about what is happening at work (e.g., change in job demands) vs. outside of work (e.g., having more free time to take care of oneself)?*
- I believe the authors may be able to use more sophisticated models to control for time effects that they are not currently doing; however I am not an expert on those models.*

Response: We have included more discussion on the different ways organisations can benefit from four-day workweeks in terms of productivity. White-collar firms are more likely to improve productivity by reducing inefficiencies such as cutting out low-value tasks, while human service occupations such as health care or restaurants experience productivity gains and cost reductions from reduced attrition—a prevalent issue in these fields that negatively affects productivity. We have added discussions on this point on page 6.

The sustainability of the well-being benefits was also a concern for us. Accordingly, we have continued to collect data after the end of the trial from the more than 90% of organisations who decided to maintain the schedules. Examining changes at 12 months following baseline, the results are presented in Supplementary Table 4 and discussed on pages 14–15. We observe slight evidence of hedonic adaptation only in the case of job satisfaction, which shows some sign of a return to baseline levels. However, all outcomes at 12 months are still significantly higher than at baseline.

We have provided additional results to disentangle the mediation effects occurring at work versus outside of work (page 14). Additionally, all our models control for period indicators (i.e., when the trial started) to account for time effects (page 23).

[Reviewer 4]

The paper reports the impact of a shorter workweek on workers' well-being. It uses unique, large scale data of workers over many different organisations in different countries. This is an important study in showing the large-scale impact of a working time reduction with wage retention.

I do have some remarks that I would like to see addressed:

25. I would like to see a little more context on the setting of the trials. They took place in Anglo-Saxon countries known for long work hours and strong neoliberal cultures. After COVID-19, it might be no surprise that it is exactly in these countries that most resistance to work as they knew it grew?

Response: We appreciate this observation. While our study consists of primarily Anglophone countries, efforts to reduce working hours have been most common in the Nordic countries, and the shift to a four-day week began in Spain. 4 Day Week Global's initial focus on Anglophone countries was partly a coincidence due to it being a New Zealand organisation. The trials are now occurring in many countries. Additionally, while the Anglophone countries in our study are characterized by neoliberal approaches to social and economic life, all the Anglophone countries in our sample, apart from the U.S., work fewer hours per week than the OECD average⁵. We added this discussion on page 5.

26. On page 2 is stated that also companies from Continental Europe and South Africa participated. However, later on it seems that only Ireland, US, Canada, Australia, NZ and UK are included in the data analysis. This is not explained well in the text.

Response: Thank you for pointing out the inconsistency. Data collection from Continental Europe and South Africa was in progress when we submitted the manuscript; therefore, these countries were not included in the analyses. In this submission, we have included more data from South Africa and a few companies from Continental Europe, now grouped under the "other" category due to the small sample size.

27. The paper is very ambitious in that it includes many measures and different types of analyses, yet this is also a pitfall. There is little room for nuance and I would like to see a little more of this:

o Make more of the difference between the groups of decreased hours (≥ 8 hours, 5-7 hours, 1-4 hours). I think this is an important aspect that the paper should address more: do all decreases matter to well-being or is it only the large decreases. The paper points out that there is a difference, but interesting might be to test the significant differences between groups as well. Also include the number of respondents included in the decreased work hours groups.

Response: In response to the comment, we have now conducted pairwise comparisons, allowing different reference groups to be used for comparisons across groups with varying changes in work hours (page 12). We have also included the number of respondents in each of the groups on page 22.

o The mediating variables are interesting to use and shine a light on how a shorter workweek impacts well-being. However, the organizational changes that were made in many of the companies might also contribute to those mediating factors and these effects cannot be separated. I would like to somewhat more reflection on this.

Response: We agree that organisational changes—as opposed to individual accommodations—likely contributed to the changes in the mediating factors we have observed. Since submitting our initial manuscript, we have collected data from 12 control companies that attended an information session about the trials but did not join a trial, allowing us to directly compare companies that undergo a collective change with those that do not.

Accordingly, our analyses now include both organisational-level and individual-level measures. While control companies did not experience changes in work hours, mediators, or well-being outcomes, companies that implemented the four-day week trials showed a significant reduction in work hours, accompanied by notable improvements in both mediating factors and employee well-being (pages 10–13). We believe these results provide evidence that the organisational changes in the trial companies led to the observed pre-post changes in the mediators and well-being.

We acknowledge that questions remain about whether similar changes in mediators and well-being could occur through other channels without an organisational change. On page 18, we draw on prior research on the drawbacks of individual accommodations to suggest that organisational changes are likely necessary to avoid unintended consequences such as flexibility stigma.

o Make more of the difference between well-being indicators. 5 indicators are included and there is some difference between these, some are job related, others are more general. I think there is room for nuance between these.

Response: We agree that the well-being indicators we examine are distinct. We have added further discussion on our rationale for selecting these outcomes (page 2) and the different results across them (page 15).

o I wonder if it might be interesting to include the start point values of the mediating variables in the analyses of Table 3? In addition to the change over time.

Response: We have now included a supplementary analysis that includes the start point values of the mediating variables as covariates (page 15). Our main results remain unchanged.

28. The limitations section can include some more reflection on the Hawthorne effect and hedonic adaption, especially since the trial was only six months it is hard to know how this would change in the long run.

Response: We appreciate the reviewer's comments on the Hawthorne effect and hedonic adaptation. The durability of the well-being benefits was also a concern for us. Accordingly, we have continued to collect data after the end of the trial from the more than 90% of organisations who maintained the new schedules. The results, examining changes at 12 months following baseline, are presented in Supplementary Table 4 and discussed on pages 14–15. As shown, the well-being benefits are largely sustainable, suggesting that the findings are not merely due to the Hawthorne effect. We observe slight evidence of hedonic adaptation only in the case of job satisfaction, which shows some sign of a gradual return to baseline levels, although all outcomes at 12 months are still significantly higher than at baseline.

29. *Smaller comments:*

- p 5, line 131: 'low- and zero-productivity activities': can you give more examples of these?
- p 13, line 354: work relatedness also seems unaltered by reductions (not really significant)
- Table 2: include N

Response: We have now added examples on page 6 and included the sample size in Table 2. The relationship between changes in hours and work relatedness shifts somewhat after the inclusion of control companies and the examination of both company- and individual-level changes in hours, which we now discuss on page 17.

30. *Concluding:*

The paper is an interesting broad study on the impact of an organisational working time reduction on workers' well-being. Although it might need some more nuancing here and there, it does give a good first general impression of the impact of these trials. In the future, the inclusion of the qualitative findings and more in-depth focus on one aspect will be very valuable as well. I am looking forward to reading more on these trials.

Response: We greatly appreciate the reviewer's positive assessment and encouragement of our study! We have collected in-depth interview data on workers' and managers' experiences of the trials, which will be the focus of future papers. We also mention the need for more qualitative research on page 18 as part of future directions.

References

1. Hernan, M. A. & Robins, J. M. *Causal Inference: What If*. (Taylor and Francis, Boca Raton, 2023).
2. Deutskens, E., de Ruyter, K., Wetzels, M. & Oosterveld, P. Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Mark. Lett.* **15**, 21–36 (2004).
3. Kristensen, T. S., Borritz, M., Villadsen, E. & Christensen, K. B. The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. *Work Stress* **19**, 192–207 (2005).
4. Idler, E. L. & Benyamini, Y. Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *J. Health Soc. Behav.* **38**, 21 (1997).
5. OECD. Average annual hours actually worked per worker. *OECD* <https://www.oecd.org/en/data/indicators/hours-worked.html> (2024).

February 12, 2025

We thank the editor for the opportunity to revise and resubmit our paper, “Work time reduction via a four-day week finds improvements in workers’ well-being.” The reviewers’ comments provided very helpful guidance. All the comments and suggestions have been seriously considered and addressed. We believe these revisions have resulted in an improved manuscript that is suitable for publication in *Nature Human Behaviour*. To facilitate reading, we underline our changes in the manuscript.

[Editor]

*Your manuscript has now been evaluated by Reviewers 1, 2, and 4 from the original round of review. All reviewer feedback is included at the end of this letter. Although the reviewers found your manuscript to have improved during revision, they also raise some important outstanding concerns. We remain interested in the possibility of publishing your study in *Nature Human Behaviour*, but would like to consider your response to these outstanding concerns in the form of a revised manuscript before we make a decision on publication.*

1. In particular, Reviewers 1 and 4 both raise questions regarding the statistical approach. We would ask you to perform the recommended analysis, including adding an omnibus test rather than pairwise comparisons only.

Response: As we respond below to comments #4 and #16, we have revised some analyses as suggested (page 9; pages 10–11) and added an omnibus test in Table 2.

2. Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

Response: We thank the editor for the checklist. We have ensured that all editorial policies and formatting requirements were followed when preparing the resubmission materials.

[Reviewer 1]

I would like to commend the authors for their thorough and thoughtful responses to my previous comments, as well as those of the other reviewers.

3. I have one comment regarding the nature of the control group. In the current version, the authors used a sample of companies that were willing to implement working time reductions. It is unlikely that these companies are fully representative of the broader population of firms, as they may possess unique characteristics (e.g., a specific workforce composition).

I recognize that the authors have likely made the best possible use of the available data, and I do not expect them to include random firms that might be unwilling to adopt working time reductions. Nonetheless, the paper would benefit from a more in-depth discussion of the extent to which the control firms are representative of the wider business landscape. Such a discussion would strengthen the manuscript and provide valuable insights for policymakers.

Response: We greatly appreciate the reviewer for this important point. The control companies are indeed not a representative sample of the boarder population of firms. However, as you noted, they represent a more comparable group than randomly selected firms that are unlikely to adopt working time reductions. We have added a discussion along this line on page 18.

[Reviewer 2]

The quality of the manuscript has improved, in particular the interpretation of the results is more balanced. The manuscript contains a solid presentation of results (5 tables, 7 tables in supplement and 3 figures) and the editor will have to decide if some tables can be moved to supplement. I have three minor comments, and one major comment. I start with the major comment.

4. I think that the statistical analysis including the control group (table 2) should be based on a statistical model including group (intervention group vs control group) and time (baseline vs endpoint). The key statistical test for a significant effect of the intervention is the interaction between group and time. An interaction analysis means that only one test per outcome variable is needed, which reduces the risk of type 1 errors (compared with the t-test analysis). Thus, I recommend that the analysis based on t-tests are replaced with a “real” interaction-based analysis.

Response: We appreciate the reviewer for this great suggestion. Accordingly, we have added two columns in Table 2 to present the significance of the interaction term between treatment/control group and time for all the measures. We have also added the corresponding findings on pages 10–11.

5. It is a strength that the study design has included a control group. However, the control group is very small and represents about 10% of the participants in the study. In addition, the control group is collected later than most other trail groups in the study. These shortcomings should be mentioned as a limitation in the discussion.

Response: We acknowledge the limitations of the control group used in our analysis and have now mentioned this in the discussion on page 18.

6. The third comment is about conflict of interest. Do the authors have any formal connection or link with 4DWG group? This connection should be described (or have I read carelessly and missed this paragraph).

Response: We thank the reviewer for raising this important question. We have no financial or legal connections or agreements with 4DWG, as indicated in the Competing Interests Statement

(page 25) and also noted on page 19. We are independent researchers invited by 4DWG to collect and analyse data independently, and our ongoing research is funded by grants from public and private foundations—which have no influence on our research findings as well—as disclosed in the Acknowledgements section. While we have occasionally participated in recruitment and onboarding webinars organised by 4DWG to introduce the research process, and 4DWG sometimes publishes reports written by us, they have no influence on the design, results, or outputs.

7. Finally, the abstract should end with a general conclusion, preferably related to the question related to the title.

Response: Following the suggestion, we have revised the end of the abstract to reference the title (page 1).

[Reviewer 4]

As mentioned in the previous review, the paper's strength is its the unique, large scale data of workers over many different organisations in different countries. This is an important study in showing the large-scale impact of a working time reduction with wage retention.

The paper has been adapted quite heavily although the original aim remains. The challenge with a paper like this, including many analyses, variables and levels is keeping it orderly and clear. I think the paper may want to show too much and loses focus here and there. However, I do think the authors did a great job trying to summarize the findings in the discussion section.

I still believe the paper adds important insights to the literature on working time reduction and do think it is fit for publication if the following points of feedback are considered:

Introduction/literature review

8. P6 line 114-121: *This part seems to fit better in the result section as it does discuss empirical findings of non-published research.*

Response: Thank you for this suggestion. We have now removed this part of the paper because it is only tangentially related to our focus on well-being, allowing us to include other requested discussions.

9. P. 6-8: *consider moving 'Hypothesis 2' to right below line 130*

Response: As suggested, we have moved our previous Hypothesis 2 (now Hypothesis 3) to right below line 130 in the previous version (page 7).

10. P. 6-8: *I think the structure of this text could be improved and made more clear to clearly state what you are talking about in each paragraph and outlining this before the start of the three paragraphs on (1) company level changes, (2) work experiences and (3) health behaviors.*

I also think Hypothesis 3 is too broad. Could you be more specific in what work experiences and what health behaviors you think might be enhanced?

Response: Thank you for helping us improve the flow and clarity of the paper. We have added two sentences on page 7 to summarise the upcoming discussion on company-level changes, work experiences and health behaviours to better present our argument. We have also revised Hypotheses 4 and 5 to specify the types of work experiences and health behaviours we expect to change with four-day workweeks (pages 8–9).

11. I know the space is limited, but I miss some differentiation between wellbeing indicators. What does wellbeing entail exactly in this paper and why were these indicators chosen? How do you expect shorter working hours to impact these indicators – do you expect the same improvement for all – why not?

Response: We greatly appreciate your understanding of the limited space available. In response to your comment, we have added a hypothesis to differentiate the well-being indicators we examine (page 6) and have revisited the hypothesis in the discussion section (page 15).

Results:

12. Can we have some more information on the control companies – what country are these based in?

Response: The control companies are from the U.S., which we have now clarified on page 9.

13. Can we have some more information on the data collected at 12 months? Were the employees still working less at that point?

Response: We have added a discussion on work hours at 12 months on page 14—yes, they were still working fewer hours at that point. The corresponding results are shown in Supplementary Table 4.

14. Line 267 should be ‘less likely’?

Response: We have corrected the typo on page 12—thank you for catching that!

15. Be consistent in the vocabulary that you use. What does health behaviors entail and why is it called like that?

Response: We have now clarified what we mean by health behaviours on page 8.

16. How big are the different work hours groups? And how big is the reference group of control companies/no reduction? Are there many employees part of trial companies but had no reduction in hours? Maybe I missed this information, but it seems strange to me to include both groups in the reference group, as these are very different.

Response: We have provided the sizes of the different work hours groups and control companies on page 22. The reference group includes employees in trial companies who had no reduction in hours; they are grouped with control employees due to perfect collinearity when both company-

level and employee-level changes in hours are included. Specifically, the control group cannot serve as the reference level for both the company-level and employee-level changes in hours variables. The software automatically removes employees with no reductions in work hours (see the highlighted lines in the screenshot below). To ensure the robustness of our results, we have conducted a sensitivity analysis in which company-level changes in hours are removed and found similar results (see the table below).

```
. mixed burnoutc i.chourmc ib5.hours c4 burnoutb $hours || Companyb:
note: 4.hours c4 omitted because of collinearity.

Performing EM optimization ...

Performing gradient-based optimization:
Iteration 0: log likelihood = -2994.2783
Iteration 1: log likelihood = -2994.1571
Iteration 2: log likelihood = -2994.157

Computing standard errors ...

Mixed-effects ML regression
Group variable: Companyb
Number of obs = 3,181
Number of groups = 153
Obs per group:
    min = 1
    avg = 20.8
    max = 392
Wald chi2(33) = 1449.13
Prob > chi2 = 0.0000

Log likelihood = -2994.157
```

burnoutc	Coefficient	Std. err.	z	P> z	[95% conf. interval]
chourmc					
-8 or more	-.5212702	.1115884	-4.67	0.000	-.7399794 -.3025609
(4,8)	-.4381215	.0948655	-4.62	0.000	-.6240545 -.2521885
(0,4]	-.4443286	.0969341	-4.58	0.000	-.6343158 -.2543413
hours c4					
8+	-.2713527	.0326914	-8.30	0.000	-.3354267 -.2072787
5-7	-.2026829	.0335066	-6.05	0.000	-.2683546 -.1370112
1-4	-.098054	.0348922	-2.81	0.005	-.1664415 -.0296665
0/+	0 (omitted)				

Table A for Reviewers: Coefficients for individual-level change in hours predicting changes in well-being using a two-level mixed-effects model (control variables omitted)

	Burnout	Job satisfaction	Mental health	Physical health
Individual-level change in hours (reference: control group)				
Reduced >=8 hours	-0.723*** (0.092)	1.393*** (0.210)	0.630*** (0.110)	0.348*** (0.093)
Reduced 5-7 hours	-0.650*** (0.093)	1.282*** (0.215)	0.478*** (0.112)	0.260** (0.095)
Reduced 1-4 hours	-0.545*** (0.095)	1.090*** (0.218)	0.428*** (0.114)	0.255** (0.097)
No change or increase in hours	-0.447*** (0.094)	0.846*** (0.215)	0.324** (0.112)	0.183+ (0.095)

17. I am not an expert in multilevel analyses, but I do wonder if the two level model is necessary here?

Response: We thank the reviewer for this question. We believe a two-level model is necessary. Traditional single-level regression assumes that observations are independent, which is not the case here, as employees from the same organisation tend to be more similar due to their shared exposure to the organisational context. We now clarify our model choice on page 9.

18. I appreciate the attention paid to differences between work hours groups in the analyses, yet I would like to read more about this in the discussion. Maybe it is more interesting to compare these, as I have doubts about the references group used now (see above).

Response: We have now added more discussion on the comparison between work hours groups on page 15.

19. A paper that might be interesting and similar, about a longitudinal working time reduction and its impact on three domains of wellbeing, is the following: Mullens, F., & Laurijssen, I. (2024). An organizational working time reduction and its impact on three domains of mental well-being of employees: a panel study. BMC public health, 24(1), 1727. Although with its limited numbers and only one company differs in various aspects from this study.

Response: We thank the reviewer for pointing us to this research, and we have now included this paper on page 4.