Author’s response to reviews

Title: Psychological wellbeing in a resource-limited work environment: Examining levels and determinants among health workers in rural Malawi.

Authors:

Julia Lohmann (julia.lohmann@lshtm.ac.uk)
Olzhas Shulenbayev (olzhasshulenbayev@gmail.com)
Danielle Wilhelm (danielle.wilhelm@uni-heidelberg.de)
Adamson S. Muula (amuula@medcol.mw)
Manuela De Allegri (manuela.deallegri@uni-heidelberg.de)

Version: 1 Date: 13 May 2019

Author’s response to reviews:

Dear editor and reviewers,

Thank you for reviewing our manuscript, we are grateful for the very helpful comment and questions. Before we respond to them one by one, however, we’d like to address one more fundamental issue.

Some of your comments have led us to make changes to the models. In doing so, however, we incurred certain issues with convergence for the dichotomous outcome variables. After some reflection and reading further into relevant methodological papers, we decided to build simpler linear/logistic regression models with standard errors clustered by facility rather than multilevel (random effects) models. This is also in light of the relatively small sample size and the fact that the focus of our paper is on individual-level factors. Moreover, results of these models might be easier for readers to understand and interpret. We also considered adding facility fixed effects, but discarded the idea as this leads to exclusion of a substantial number of observations due to perfect prediction for the dichotomized WHO-5 using the 40% and 60% thresholds.

This revised analytical approach has changed results slightly for the continuous WHO-5 measure. For the dichotomized WHO-5 measure, changes are more important. Whether the respondent is in a relationship (added, see reviewer #1’s first comment) is the only predictor variable that reaches statistical significance for the 50% cut-off, whereas being in a relationship does not, but a range of additional variables does predict wellbeing categorization if 40% or 60% cutoffs are used. Considering the overall picture including predictor variables where the
confidence interval only barely includes zero, however, the main message appears to remain the same in that clinical competence and satisfaction with interpersonal relationships at work are particularly important for wellbeing.

We believe that the advantages of this revised analytical approach outweigh the disadvantages, and hope that as editors and reviewers, you will agree with our decision. We have adjusted the description of the methods and results (including results for the 40% and 60% cut-off), and made the cut-off issue a more prominent discussion topic.

In the following, we will respond to your comments and questions one by one.

Reviewer 1:

Comment #1:

This manuscript is generally well-written. It investigates factors associated with the psychological wellbeing of mid-level maternal health workers in rural Malawi. The problem this research addresses is of importance as many countries face great difficulties in ensuring a stable and healthy health workforce, especially in rural areas. Understanding the levels of and predictors of psychological wellbeing of health workers is a first step towards remedying high levels of health worker turnover. Indeed, this study highlights the problem of workforce stability when it reports that only 10% of mid-level health workers are in the same job in 2015 as they were in 2013.

Unfortunately, the study appears to have been conceptualized after data collection was planned/undertaken (ie post-hoc), so as the authors acknowledge, not all work-related variables potentially relevant to the wellbeing of health workers were included in the questionnaire and in analysis. Further, the authors' conceptual model does not acknowledge the role of non-work-related factors in influencing wellbeing, nor does the discussion acknowledge the absence of non-work-related factors in the modelling as a limitation. This is important, since the outcome variable measures overall wellbeing, not just work-specific wellbeing. One way to overcome concerns about potentially important variables that may have been omitted from the modelling is to report how much of the variance in wellbeing is explained by the modelling and to demonstrate goodness of fit of the models.

Response: Thank you for underlining these issues. In response, we revised three elements. First, we have strengthened this point in the introduction and discussion, where we had indeed forgotten to include it, thank you for pointing this out. In this context, we have tried to better highlight that we approach the issue from a health system/human resource management perspective. From this angle, non-work factors (personality, stress and strain in personal life,
etc.) are interesting in that they influence behavior at work, but are at the same time difficult to address directly through interventions within the health system. For this reason and of course for lack of relevant data, we had excluded them entirely.

However, we have reintroduced partnership status and children/dependents into the model, as both variables have been found rather important in the wellbeing literature and might in fact be addressed through health systems interventions, for instance by providing particular support to health workers with children.

We have further added (pseudo) R2 to the results tables to give an idea of the amount of variance explained – between 20% and 36%, depending on the model. While this is actually a fairly substantial amount in psychological research, as you rightfully assume, a lot of variance remains unexplained, and we have discussed the issue more extensively:

“The study used data collected for a different primary purpose and questionnaires did not include all variables potentially relevant to PW. Results show that the included variables explain only between 20 and 36% of variance in PW, indicating the importance of other work-related and non-work-related factors at the individual and higher organizational levels in determining psychological wellbeing. More comprehensive, focused research would therefore be highly desirable for a more comprehensive picture of determinants of wellbeing of health workers.” (p. 16, l. 11 of the new clean version of the manuscript)

Comment #2:

Additionally, while the authors reported the percentage of health workers whose psychological wellbeing was less than 40%, 50% and 60%, their sensitivity analysis didn't extend to examining the results of logistic regression modelling if the different cut-off was used as the dichotomous outcome variable. This analysis could be provided as a supplementary file. Consistent results for the different cut-off levels would provide stronger evidence of the significance of each variable.

Response: Given the observed differences in results using simple logit models as explained above, we have included results for all three models in the main manuscript, and addressed the issue throughout the results and discussion sections.

Comment #3:

1. The continuous WHO-5 scores have a skewed distribution. In this instance, reporting median and interquartile range is preferable to mean and standard deviation.

Response: Since we do not present statistics, but only the actual distribution in Figure 1, we are not entirely sure what your comment refers to. However, we agree that normal distribution of the
data can be disputed from Figure 1. However, tests for normality (using swilk and sfrancia in Stata) both indicate that normal distribution can be assumed and presentation of means and standard deviations are therefore appropriate.

<table>
<thead>
<tr>
<th>test</th>
<th>Obs</th>
<th>W</th>
<th>V</th>
<th>z</th>
<th>Prob&gt;z</th>
</tr>
</thead>
<tbody>
<tr>
<td>swilk</td>
<td>174</td>
<td>0.98461</td>
<td>2.035</td>
<td>1.623</td>
<td>0.05228</td>
</tr>
<tr>
<td>sfrancia</td>
<td>174</td>
<td>0.99209</td>
<td>1.145</td>
<td>0.277</td>
<td>0.39096</td>
</tr>
</tbody>
</table>

On a side note, some skewness is the norm and often desired when using psychometric scales which unlike fully continuous variables are limited at both ends. In psychometric research, such type of data are usually treated as continuous and analyzed with methods normally used for continuous data, and a large number of methodological studies have shown that whether methods for continuous or rather for ordered-categorical data are used tends to be fairly irrelevant for results. See for instance


Comment #4:

2. Also, consistent reporting of 95% confidence intervals is preferred over p values.

Response: We now report confidence intervals instead of p values.

Comment #5:

3. I would expect that $\alpha=.05$ be used to test for statistical significance, however this is not clear with reporting of the results. Eg. Lines 12-14 page 14, refers to a statistically non-significant result ($p=.069$), so this should be made clear. Eg. P16 lines 8-10, although this difference was not statistically significant with the continuous outcome variable. Etc.
Response: Your assumption is correct. We have made this more explicit by reporting the 95% confidence interval instead of p values and adding one explicit sentence to the beginning of the results section: “In the following, we briefly summarize the results, using “statistically significant” to refer to coefficients for which the 95% confidence interval does not include zero.” (p. 11, l. 15)

Comment #6:

4. Also, perhaps having received professional training within the last year and having received any supervision within the last month should be consistently categorised as relating to 'organisational support' rather than to 'clinical competence' (eg. Table 3).

Response: We agree in regards to supervision and have moved it to organizational support, but left professional training in the clinical competence category as we feel it is better suited there.

Comment #7:

5. Table 3, line 20 p30, not sure what 'boni' refers to. Is this a typo?

Response: This was supposed to be ‘bonuses’. We’ve corrected.

Comment #8:

6. Table 2, line 34 p28. It took me a while to work out what 'PBF exposure' meant. The acronym should be used consistently through the paper.

Response: Thanks for pointing this out. We’ve clarified and applied the term “RBF4MNH exposure” consistently.

Comment #9:

7. Typo with first reference, should read 'Everybody's'

8. Type p3 line 21 emphatic should be emphatic

Response: Thank you for catching these typos. We have corrected accordingly.
Comment #10:

9. The authors refer to an impact analysis of intervention on PW that is unpublished. Can this be included as a supplementary file (or are there plans to publish separately?). Can further details be provided?

Response: We have deleted the paragraph related to RBF4MNH from the discussion sections for two reasons. First, in the revised version of the multivariate results, data collection year does not appear quite as important anymore. Second, upon reflection on your comment, we realized that the introduction of this context-specific strand of discussion without being able to explain in adequate detail is more confusing than helpful, particularly to an audience likely not very familiar with performance-based financing.

However, for your information, we provide some additional explanation. The data stem from an impact evaluation of RBF4MNH on a large range of outcomes, including quality and utilization of maternal care services, perceived quality of care, and health worker motivation. The study used a controlled before-and-after design. For details, please refer to the protocol [34] or the final report (https://www.klinikum.uni-heidelberg.de/fileadmin/inst_public_health/Dokumente/Final_Results_Report.pdf). The wellbeing measure was part of the health workers survey, but was not chosen as a key variable for the overall RBF4MNH impact evaluation. Results are therefore not published in the final report, but it is possible to use the data to estimate such impact. We have plans to publish results along with data from other countries in the near future and would therefore prefer not to include them here.

For your information, however, the difference-in-differences impact estimates for the continuous and dichotomized WHO-5 are in the table below. The estimation method was the same as in [33], with the only difference that we used all variables included in this study as predictors as covariates rather than only a handful of demographic characteristics.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Comparison</th>
<th>Difference-in-differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Endline</td>
</tr>
<tr>
<td></td>
<td>mean sd</td>
<td>mean sd</td>
</tr>
<tr>
<td>Continuous WHO-5</td>
<td>9.3 3.4</td>
<td>9.7 3.2</td>
</tr>
<tr>
<td>Dichotomous WHO-5*</td>
<td>66% 48%</td>
<td>78% 42%</td>
</tr>
</tbody>
</table>
* Percentages indicate proportion of the sample above the 50% threshold, i.e. with adequate wellbeing

In interpreting the estimates, please keep in mind two issues, which might also facilitate understanding of what we meant in the discussion section – where we have worked on the wording a little to make things clearer. First, the number of clusters is relatively limited (33), which is due to the small scope of RBF4MNH and the overall relatively small number of health facilities in Malawi. The trial was powered for certain outcomes, but not to detect small impact on mental health. Second, intervention and control facilities were situated within the same districts for a number of practical and methodological considerations. However, during qualitative study components accompanying the quantitative trial, there emerged substantial “contamination” in the sense that a) district health management teams (DHMTs) also received PBF and were incentivized to improve performance in the districts at large, not only at intervention facilities; and that b) whether linked to district-level incentives or independent of that, DHMTs made use of PBF to better deal with chronic resource shortages by redirecting their own resources to facilities which did not benefit from PBF. Given the impact evaluation design, we were not able to estimate the magnitude of these district-level effects, as we did not have controls outside the intervention districts. However, both qualitative study results and the positive trends over time that we observed on many indicators (including health worker wellbeing) strongly suggest that such spillover effects were present and powerful.

Comment #11:

10. Suggest remove subheadings in Discussion section

Response: We have removed subheadings as suggested.

Comment #12:

11. Method for scoring clinical vignettes should be explained

Response: We measured clinical competence with 2 case vignettes, one on postpartum hemorrhage and eclampsia. They were scored as follows, and the scores on the two then combined into an overall knowledge score.

PPH: Indicator met if health worker could give 4 or more correct answers for the treatment of post-partum hemorrhage: call for assistance, perform physical exam, give IV fluids, check patient’s vital signs, administer oxytocin, identify indication for oxytocin
Pre-eclampsia: Indicator met if health worker could give 3 or more correct answers for describing dangers signs of pre-eclampsia to a patient: vaginal bleeding, convulsions/seizures, headache, swelling of extremities and understand the indicator for giving magnesium.

Unfortunately, we do not have space to go into detail on this, but a paper with a detailed description of these variables is in the advanced review stage and expected to come out shortly. We have added a preliminary reference to be updated once the article is published (likely before our own).

Comment #13:

12. One possible explanation for negative association between having received training or actual clinical knowledge and lower wellbeing is that people with high levels of anxiety may be more likely to seek out training and seek out clinical knowledge. Have you tested for associations with scoring on the item of wellbeing that scores anxiety (being calm and relaxed)? This might have implications for initiatives to improve wellbeing by teaching ways to manage anxiety.

13. Another possible explanation of the negative association with training is that the training is poor and this contributes to decreased wellbeing/dissatisfaction, though this is entirely speculative.

Response: We have added both ideas to the discussion section as a potential starting point for further inquiry: “Possible alternative explanations include poor training quality or factors concurrently associated with PW and actively seeking out training, such as anxiety.” (p.15, l. 21)

We are somewhat reluctant to take the WHO-5 apart and attribute specific diagnoses to single items. For curiosity and your information, however, we regressed the “calm and relaxed” item individually on the predictor variables. Results more or less align with those for the combined WHO-5, but none of the predictor variables reach statistical significance.

Comment #14:

14. Basic demographic characteristics measured were limited. Not measured were factors such as ethnicity, marital status, having children, other measure of minority status etc. Statement page 18, lines 11-13 should be amended so that it is clear that while the research indicated no specific demographic group that could be targeted to improve their PW, there may be demographic factors that weren't measured in this study that are associated with PW or something similar.

Response: Thank you for underlining this. Please see our response to your first comment. In short, our approach to the topic is from a health system perspective and our primary aim is to identify factors which the health system can possibly target. As such, we had excluded most non-
work-related aspects. However, we agree that these are not only very important in general, but that some variables available to us but initially excluded might actually not be unimportant from a management perspective, such as family obligations. We have therefore added them to the model. Unfortunately, we do not have information on health worker ethnicity or any other non-work-related data. We have also added an explicit paragraph on this issue in the discussion as specified above. We have also entirely removed previous lines 11-13 on page 18, as results with the revised models do indicate a role of certain demographic factors.

Comment #15:

15. No explanations about missing data. Did all respondents answer all questions?

Response: We were lucky to have very few missing data points to deal with. There were no missings on the WHO-5 items. Missings for included predictor variables were all below 2%, except for age at 3.5%. None of the respondents had missings on more than 4 predictor variables. In light of this positive missingness situation and in order not to lose observations due to missings in predictor variables, we therefore decided to impute the few missing values. As imputation was done in the context of the PBF impact evaluation, we replaced missings using information from the respective PBF study arm x data collection time point subsample to which the respondent with a missing value belonged. For categorical data, we used the respective subsample mode. For continuous variables, we used the respective subsample mean (rounded to the next full integer). For variables with multi-item psychometric scales, we calculated the mean of the scale excluding the missings if less than 50% of the respective scale items were missing answers, which was the case for all.

We have added the following sentence to the methods section to clarify: “Data was complete for the WHO-5. For the predictor variables, data were missing for less than 2% of the sample for all variables except age (3.5%) and were imputed using modes/means in the respective RBF4MNH impact evaluation study arm*data collection year subsample.” (p.10, l. 9)

Comment #16:

16. A little more detail about sampling of participants from the secondary health facilities is needed. How many in total (approximately)? How were they 'randomly' sampled? Were adjustments made for survey sampling methods in analysis?

Response: In fact, this was mistakenly phrased, so thank you for inquiring and thereby reminding us of one aspect that we had indeed overlooked to discuss in the limitations sections.
Initially, the idea had been to interview a random sample of 5 out of all health workers in the maternity department on duty during the time the data collection teams were at the facility. However, this proved to be difficult to implement on the field, so that data collection teams were instructed to interview all maternity staff available in the study period in alignment with sampling procedures at primary-level facilities. In principal, the data collection strategy was therefore the same at primary and secondary level facilities. We have reason to assume, however, that in practice, interviewers were not quite as rigorous in finding and interviewing all eligible health workers in the much larger secondary-level facilities. Although we have little reason to believe that potential biases in sampling were related to health workers psychological wellbeing, we also cannot fully exclude the possibility. In the revised models, we included “level of care” as an additional covariate in the attempt to at least somewhat account for potential differences in sampling at the two levels.

Beyond this issue, your comment reminded us of one important limitation related to the sample not being fully representative of the health worker population. This was not a major concern for other objectives in the overall impact evaluation, but might in fact be important for mental health. Specifically, health workers who due to poor psychological wellbeing were unable to come to work were not captured by the survey. Results might therefore be somewhat positively biased. We have taken this up into the discussion section:

“In this context, it is also important to consider that the sample is not fully representative of the health worker population. Rather, only health workers present at the workplace were interviewed, thereby possibly excluding health workers unable to report to work due to particularly poor psychological wellbeing. Estimates of psychological wellbeing in our study are therefore likely positively biased.” (p. 14, l. 7)

Reviewer 2:

Comment #1:

Introduction: One of the factors that should also be reflected upon in the background is the wellbeing by the cadre of worker

Response: We would have liked to go into more detail on the various predictor variables in the introduction section, but are unfortunately unable to do so within the tight word limit of HRH. However, we have included cadre to the bracket where we give examples of individual-level work factors.
Comment #2:

Methods: The authors have referred the readers to the full manuscript; however, given that district*time is one of the random-effects parameters, it would be helpful to know the number of districts and the avg. no of facilities per district to contextualize that result.

Response: As described above, we have decided on a simpler model, where district is no longer included. However, we have included a bracket to indicate that there were 8 or 9 facilities per district, including 1-2 secondary-level hospitals each.

Comment #3:

Results: Not sure if I am interpreting the results in Table 5 correctly, but it appears that variables such as "Cadre" were modeled as continuous variables. If so, please explain. This variable should be modeled as a nominal variable. It also appears that in Table 1, there are a total of 31 clinical officers and medical assistants and 143 nurses. Perhaps, for the purpose of the multivariate analysis, the categories of clinical officers and medical assistants can be condensed?

Response: In all analyses, cadre was used as a binary/nominal variable where 0=clinical officer OR medical assistant and 1=nurse. So clinical officers and medical assistants had always been condensed into the same category, for the same reason you mention. We have replaced “0=clinical officer, medical assistant” by “0=clinical officer/medical assistant” in the Tables, in the hopes that this makes things clearer.

Comment #4:

Please present confidence intervals for all items in Table 5. Given the small sample size, and the high level of disaggregation, as well as random effects parameters, concerns about the robustness of the results should be addressed in the discussion.

Response: We have included confidence intervals in the revised tables. We agree that robustness of results might have been an issue with the more complex multilevel models, which we have now simplified as explained above. Using Greene’s (1991) rule of thumb of minimum n=104+(number of predictors), we should now have a sample size large enough to assume results are reasonably robust.

Comment #5:
Not certain I see the need for random effects by time, especially, given the relatively small sample size.

Response: As explained above, we have removed the random effects and performed simple regression models instead.

Comment #6:

Please clarify what the variable "health facility" in Table 5 is? Is that health facility in-charge (yes/no?)

Response: We are unsure to what you are referring. “Health facility in-charge” is the third predictor variable from the top, but was labelled completely. “Health facility” in the random effects section of the previous results tables referred to the health facility the health worker was working in at the time. The latter was removed in the revised version of the tables given the change in models. We hope that this clarifies.

Comment #7:

Have you considered minimizing the number of variables in the multivariate model based on importance and bivariate significance?

Response: We did consider this, but believe that in light of the objective of the paper – to investigate characteristics associated with poor psychological wellbeing that might be instrumental to managers and policy makers in addressing the issue – it is also important to show which characteristics are not associated with wellbeing. Indeed, for most of the included variables, this seems not to be the case. For all included variables, we had good reasons to believe that they might play a role based on the literature from other settings, which we unfortunately could not review in detail because of the tight word limit.

If your concern stemmed from a statistical point of view rather, the revised models are less complex so that the sample size is large enough to accommodate the number of predictors and lead to reasonably robust estimates, which might have indeed been a concern with the previous multi-level models.

Comment #8:

I would be curious to see these results for the cadre of nurses only given that they constitute most of the sample. Your results may have higher practical relevance if outlined by type of cadre.
Response: Thank you for this suggestion. Since “cadre” is fairly far from significance as a predictor of wellbeing, we do not expect that results differ dramatically for nurses and the other cadres. We performed the analyses separately on the nurses subsample, while the clinical officer/medical assistant category unfortunately does not have a sufficient sample size for a separate analysis. Not surprisingly, results are very similar to the overall results given that 80% of the overall sample are nurses. We therefore opted not to present them so as not to further overload the paper. We hope you agree with this decision.

Comment #9:

Discussion: If WHO-5 has not been validated in Malawi, I would highlight that as a limitation.

Response: We have made an explicit reference to Malawi in the first sentence of the limitations sections, which referred to lack of validation studies in LLMIC more generally.

Comment #10:

It might be worth commenting briefly on satisfaction with demands of the job, interpersonal relationships and supervisory support.

Response: We have done so in the revised version, also in light of slight changes in results.

Comment #11:

Are there any other estimates of psychological well-being (in any population) in Malawi? One thought is that while job satisfaction might be determined by factors such as demands of the job, interpersonal relationships and supervisory support; psychological well-being might be better determined by factors outside of the work environment in this context.

Response: We fully agree that non-work factors are certainly also very important. However, as we explained above, we have purposely taken a health system approach to looking at psychological wellbeing in this paper.

For your information, however, there is in fact quite some research on mental health in Malawi, but two aspects limit its comparability to our study. First, most of it takes a clinical approach to defining and measuring mental health and illness and a specific disease focus, usually depression, which is difficult to relate to our more holistic approach and measurement to wellbeing. Second, the vast majority of studies relate to the effect of specific health conditions
on mental health, most importantly HIV, maternity, and chronic diseases. Not surprisingly, physical illness has been found strongly associated with mental health issues.

The Global Burden of Disease study indicates that mental health issues are highly prevalent and problematic in Malawi, although many countries in Sub-Saharan Africa are even more affected (http://www.healthdata.org/sites/default/files/files/country_profiles/GBD/ihme_gbd_country_report_malawi.pdf). Recently, two studies quantifying mental health in the general population in a more detailed way were published, but both with data from older adults (45+) only (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6205235/, https://www.ncbi.nlm.nih.gov/pubmed/28752487). They show that depression and anxiety are higher among women and increase with age and decline in physical health. However, three quarters of our study sample is under the age of 45, and health workers are likely not well comparable to general population in many respects (background, educational level, socio-economic aspects, health awareness and behavior, etc.). We are not aware of any studies on mental health in a comparable population and setting in Malawi, and we are also not aware of any studies having looked at determinants beyond physical illness and basic demographic characteristics.

In light of this lack of comparable evidence, we prefer not to discuss our findings in relation to the existing literature, especially considering that within the word limit, we do not have the space to explain the matter in enough depth to do it justice and make it valuable to readers. We hope you agree with this decision.