PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Perceived bullying among Norwegian doctors in 1993, 2004 and 2014-15: A study based on cross-sectional and repeated surveys
AUTHORS	Rosta, Judith; Aasland, Olaf

VERSION 1 – REVIEW

REVIEWER	Bengt Arnetz Department of Family Medicine College of Human Medicine Michigan State University USA
REVIEW RETURNED	16-Jun-2017

GENERAL COMMENTS	This paper originates from the well-established research team affiliated with LEFO, see p 5, l 5. The question re the prevalence and risk factors for bullying among physicians are important. The dataset from Norway on are unique and has the potential to geerate unique data on the subject. Specific Comments. In general, use one term to describe how data was collected on bullying. It was self reports of perceived bullying. Do not use terms such as subjected to bullying. Title, rephrase to align with Journal standards. Abstract: The paper is largely based on cross sectional data. Do not use the term impact on or other words suggestion a cause-and-effect relationship. Make it clearer as to whether data is cross sectional or prospective data is being used. Especially in reporting risk factors for bullying in the 2014-2015 dataset. The text is incomplete in some places. e.g. Participants, p 2, line 9, what does the percentage within the parentheses represent? The conclusion section use the term detrimental, but, data is cross - sectional and thus a cause-and-effect relationship is not possible to establish. Strength and Limitation: The first dot does not relate to strength. It isa repeat of a main result only. Introduction: Expand on the challenge of the fact that different studies use different definitions of bullying. The authors should clearly state when data is prospective vs prospective.

	For example, p 4, l 22. The perception of being bullied is linked to personality trait. There is no supporting documentation for that sweeping statement. What is the cause-and-effect relationship? The next sentence use the term tend to be. Material and Method Reference to the original, large scale study makes the data content of the current study less clear. When n is provided, e g p 5, l 23, both numerator and denominator data should be listed. Use STROBE direction to present the various datasets and which ones were used to address the various hypotheses. It is not possible to delineate which results are based on cross sectional data vs prospective data. That is, risk indicator predicts future reports of bullying. Also, discuss why the question on bullying is different from other key Norwegian, non MD studies, and also teasing. Why was bullying in the last moth used as a outcome in some cases, and bullying intensity using other definitions in other instances? Results, see comments above. It is unclear when cross sectional vs prospective data from the cohorts were being used. If there is a cohort study but only cross sectional data that is being used to test the model, the fact it was sampled from a cohort study is irrelevant. This is especially important in order to interpret results presented on p 11, l 11 and forward and other places where the cohort is being used to test predictive models. Results would benefit using Table 2 as Table 1 and incorporating much of the text presented on p 9 in the table. The use of the term panel data suggest the same physicians were followed over time. Is that correct? Discussion section: The discussion section should be careful to draw causal conclusion if data is cross sectional in nature. Focus the discussion on risk factors that were studied. Attempt to relate it to some theory. If stress and working conditions have worsened, e g autonomy, among Norwegian MDs, why does the prevalence of bullying stay stable? the policy implication should be tempered since the next step would be an intervention to determine possible preventive actions.

REVIEWER	Jenny Firth-Cozens Retired. United Kingdom.
REVIEW RETURNED	04-Jul-2017

GENERAL COMMENTS	This longitudinal data set of Norwegian doctors must be one of the best/most important in the world as it covers most of Norway's doctors over a considerable timespan. As such it provides a useful way to consider how perceptions of bullying have changed or not over the years. The measures used are all valid and reliable and consistent over the time period, though I would have loved to see bullying by nurses and patients included (maybe next time). I don't know if I'm out of touch statistically, but I do think the authors could have used these data more excitingly. They have used them to consider levels of perceived bullying in different types of medicine and to compare this to other countries and to consider its cross-sectional influences.

	My principal problem with the paper is that it isn't using regression analysis (for example) to consider the relative influences of time 1 neuroticism and time 1 perceived bullying on time 3 bullying and health. I'm not sure why this hasn't been done. It would be such an valuable analysis. Have I missed something? Results: The findings that are the most unusual and significant (to me) are that perceived bullying is worse amongst senior doctors. Who bullies them? Do we know? Numbers in some groups (eg, general practice and private practice) are too small to consider. A longitudinal analysis, as suggested above, would greatly increase interest in this paper. Discussion: Something needs to be said about the much older age of this sample compared to others. Most bullying studies include mainly young doctors. For this reason, and because every study uses different measures, I don't think too much space should be given to cross-nation comparisons. Having said that, I do think the differences in Table 5 are so large to warrant some discussion. It's very interesting that the Scandinavian studies have the lowest levels – perhaps a follow-through of the fact that they have the best mental health in their young people too, so I do think it's good to note that. I would, however, omit comparison of the small groups within the sample with those in other studies. (p.12, l.5-11), apart from the consistency of bullying in surgery. It would be better to omit any discussion of findings where numbers are particularly small (l.18-19). P.13, line 10 – Substitute "On the other hand…' for 'Furthermore…". There are arguments that stress/job dissatisfaction are measuring the same thing which might be neuroticism, so it might not be too inappropriate to say that this might affect perceptions of bullying too or, on the other hand, people who are stressed and unhappy may make good victims for others. p.14, line 19-20. I don't understand the reference to 'Different tolerance levels…' and there isn't a reference. I would like more around this point (see above). I don't think there are strong policy implications from the study as it stands, except the new understanding that older and more senior doctors are seeing themselves as being bullied, along with the need for surgeons to appreciate the unnecessary levels of bullying in their specialty – consistently high round the world. There are a number of typos or errors in words which can be sorted by the editors. The language is remarkably good on the whole. I do think this remarkable data set adds a lot to our understanding but probably after a little more analysis. The either/or questions on the review sheet really don't work for the points I'm making here. For example, the statistics are adequate, but I think they could be taken much further, so I've said 'no'. I'm suggesting other things to do with the data statistically, but really it would be good to have a statistician look at that to make suggestions. I do think some of the discussion is based on too small numbers. In conclusion, bullying in healthcare is a very important issue and this is a remarkable dataset, so I hope a version of this paper will be published.

REVIEWER	Neill Thompson Northumbria University UK
REVIEW RETURNED	11-Jul-2017

GENERAL COMMENTS	Generally I think while the study has some methodological limitations around the measures used, the use and merits of the longitudinal design far outweighs this limitation. I dont think the authors make enough of the findings in their take home message or discuss the implications sufficiently. Rost and Aasland's manuscript is a very interesting study and certainly worthy of publication in BMJ Open in a revised format. It follows previous publications in BMJ Open in the area of workplace bullying in healthcare but is not simply a duplication of these previous studies as it offers a different insight. That said, I do feel that the manuscript needs work in content and overall message. The findings in relation to neuroticism confirms earlier research findings. However, I do not see this study adding anything that we haven't already gained using more comprehensive measures of neuroticism and more developed designs. Arguably this is the least noteworthy of the findings and I think the readership would be less interested in this. Also the authors may wish some consideration of (Page 14-Line 19-22)("neuroticism was a significant predictor of reporting bullying" this needs to be reviewed in light of the stronger evidence by Nielsen and Knardahl (2015) and Nielsen, Glaso et al (2016). While obviously this should not be removed from the study, as it was a planned and tested hypothesis, I am not convinced it adds much to earlier study findings. The strength of the paper is the data that has been collected across the four cohorts and the findings directly related to that. This is of course a great undertaking. It has some clear methodological issues in that it doesn't meet the gold standard we might consider in terms of measuring bullying (combining a behavioural checklist with a global measure – Nielsen et al (2011). I think this does need to be acknowledged as a limitation– however, alongside the caveat that these standard came in after the data collection commenced and would have been difficult to have been incorporated into the current wider survey design should also be included. Some of the literature is dated around personality and I would certainly recommend reviewing your findings in the context of the papers above on personality (e.g. the Neilsen et al 2016 paper and (Persson, Hogh et al. 2009). There seems to be two interesting messages – i) the cohort of doctors bullied in 1993 still report being bullied to the same prevalence in 2014. ii) There are consistent bullying levels reported across the cohorts over more than a decade. These findings are the core message of the study but I don't think this comes across as well as it could do in the discussion. I do feel these points could be explored further under the explanation of results: - Has the Norwegian healthcare services (or more broadly at Government level with legislation) not attempted any bullying interventions during this time, if they have then does this not reflect that these interventions aren't working if bullying levels are the same?

	- Environmental factors are discussed but healthcare and the workplace environment itself has moved on in the last 15 years, bullying it seems has not. What has not changed in the healthcare setting that means still bullying remains constant when all around it has improved? - Why does the 1993 cohort not report much lower levels of experience of bullying? – this can't be tested but could be considered in the explanation of findings. Is it because they are now the seniors but their leaders are still bullies , or is it the culture as such that facilitates the bullying BMJ Open Review 11/7/2017 - You would expect the 1993 cohort who were bullied to then do something about it. The constant levels seems to be stable suggesting that is not the case. Are the doctors in 1993 who are now in 2014 now the bullies or just staying quiet? Maintaining the bullying culture? For example - (Musselman, MacRae et al. 2005). Kind regards Neill Thompson Neill.thompson@northumbria.ac.uk References Musselman, L. J., H. M. MacRae, et al. (2005). "â€˜You learn better under the gunâ€™: intimidation and harassment in surgical education." Medical education 39(9): 926-934. Nielsen, M. B., L. GlasÃ¸, et al. (2016). "Exposure to workplace harassment and the Five Factor Model of personality: A meta-analysis." Personality and individual differences 104: 195-206. Nielsen, M. B. and S. Knardahl (2015). "Is workplace bullying related to the personality traits of victims? A two-year prospective study." Work & Stress 29(2): 128-149. Nielsen, M. B., G. Notelaers, et al. (2011). Measuring exposure to workplace bullying. Bullying and harassment in the workplace: Developments in theory, research, and practice. S. Einarsen, H. Hoel, D. Zapf and C. L. Cooper. 2: 149-174. Persson, R., A. Hogh, et al. (2009). "Personality trait scores among occupationally active bullied persons and witnesses to bullying." Motivation and Emotion 33(4): 387-399.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Bengt Arnetz
Comment
Specific Comments.
In general, use one term to describe how data was collected on bullying. It was self reports of perceived bullying. Do not use terms such as subjected to bullying.

Response: We have now consequently used the term "perceived bullying at work" in the whole manuscript, and clarified that the data are based on self-reports.

Page 3, line 5-6:
"Analyses are based on self-reported questionnaire data with the possibility of both over- and underestimation."

Page 16, line 23-24:
"A limitation is that we only have self-reported data, although this is considered a plausible methodology."

Comment: Title, rephrase to align with Journal standards.

Response: Page 1:
This has been revised.

Comment: Abstract:
The paper is largely based on cross sectional data. Do not use the term impact on or other words suggestion a cause-and-effect relationship.

Response: Page 2:
We have tried to follow this.

Comment: Make it clearer as to whether data is cross sectional or prospective data is being used. Especially in reporting risk factors for bullying in the 2014-2015 dataset.

Response: We have now described the data as either “cross-sectional” or “repeated”

Comment: The text is incomplete in some places. e.g. Participants, p 2, line 9, what does the percentage within the parentheses represent?

Response: This is now rectified.

Comment: The conclusion section use the term detrimental, but, data is cross -sectional and thus a cause-and-effect relationship is not possible to establish.

Response: This has now been revised.

Comment: Strengths and Limitations:
The first dot does not relate to strength. It is a repeat of a main result only.

Response: Page 3:
"Strengths and limitation" has now been revised.

Comment: Introduction:
Expand on the challenge of the fact that different studies use different definitions of bullying. The authors should clearly state when data is prospective vs prospective. For example, p 4, l 22. The perception of being bullied is linked to personality trait. There is no supporting documentation for that sweeping statement. What is the cause-and-effect relationship? The next sentence use the term tend to be.

Response: Page 4, line 16-26:
We have expanded on this point. And tried to remove all allusions to causality.
Comment: Material and Method
Reference to the original, large scale study makes the data content of the current study less clear.
When n is provided, e g p 5, l 23, both numerator and denominator data should be listed.
Response: Page 5, line 22-23.
Page 6, line 4-5.
Where natural, we have now included both numerator an denominator.

Comment:
Use STROBE direction to present the various datasets and which ones were used to address the various hypotheses. It is not possible to delineate which results are based on cross sectional data vs prospective data. That is, risk indicator predicts future reports of bullying.

Response: Page 5-6 and page 9, line 1-10:
More information on datasets and analyses are now added to the "Material and methods" section.

Comment: Also, discuss why the question on bullying is different from other key Norwegian, non MD studies, and also teasing.

Response: Page 6, line 15-24

More information (incl. references) about the questions related to bullying were added. In addition, revised the translation of the question on bullying.

Comment: Why was bullying in the last month used as a outcome in some cases, and bullying intensity using other definitions in other instances?

Response: We have revised the text (incl. tables) and consequently used the outcome as "at least a few times a month".

Comment: Results, see comments above.
It is unclear when cross sectional vs prospective data from the cohorts were being used. If there is a cohort study but only cross sectional data that is being used to test the model, the fact it was sampled from a cohort study is irrelevant. This is especially important in order to interpret results presented on p 11, l 11 and forward and other places where the cohort is being used to test predictive models.

Response: Page 9-12:
We have now clarified this better.

Comment: Results would benefit using Table 2 as Table 1 and incorporating much of the text presented on p 9 in the table.

Response: We have incorporated the sentence on perceived bullying at work at least few times a month into the table 2. However, we should like to keep the table 1 to show no changes in all 5 response categories of the response variable.

Comment: Discussion section:
The discussion section should be careful to draw causal conclusion if data is cross sectional in nature.

Response: Allusions to causality have been removed.

Comment: Focus the discussion on risk factors that were studied. Attempt to relate it to some theory. If stress and working conditions have worsened, e g autonomy, among Norwegian MDs, why does the prevalence of bullying stay stable.

Response: Page 13 line 20 to Page 16, line 15:
We have tried to make this clearer.

Comment: The policy implication should be tempered since the next step would be an intervention to determine possible preventive actions.

Response: Page 17, line 5-14:
This has been done.

Reviewer 2: Jenny Firth-Cozens

Comment: This longitudinal data set of Norwegian doctors must be one of the best/most important in the world as it covers most of Norway's doctors over a considerable timespan. As such it provides a useful way to consider how perceptions of bullying have changed or not over the years. The measures used are all valid and reliable and consistent over the time period, though I would have loved to see bullying by nurses and patients included (maybe next time).

Comment: I don't know if I'm out of touch statistically, but I do think the authors could have used these data more excitingly. They have used them to consider levels of perceived bullying in different types of medicine and to compare this to other countries and to consider its cross-sectional influences.

My principal problem with the paper is that it isn't using regression analysis (for example) to consider the relative influences of time 1 neuroticism and time 1 perceived bullying on time 3 bullying and health. I'm not sure why this hasn't been done. It would be such an valuable analysis. Have I missed something?

Response: Unfortunately, we cannot do this analysis due to the design. Only 10 % of our respondents (485 of 4,893) are "true repeaters", where we can follow changes over time on an individual level.

Page 5, line 18 to Page 6, line 9:
Hopefully this structure of the data set is now better described.

Comment: Results: The findings that are the most unusual and significant (to me) are that perceived bullying is worse amongst senior doctors. Who bullies them? Do we know? Numbers in some groups (eg, general practice and private practice) are too small to consider. A longitudinal analysis, as suggested above, would greatly increase interest in this paper.

Response: Unfortunately we have no data on who the perpetrators are, only that they could be both colleagues and superiors.
Page 6, line 17-24:
The question was: "Have you during the last year been subjected to vexation or uncomfortable teasing (bullying) from colleagues or superiors?"
Page 16, line 28 to page 17, line 3:
This is detailed in the "Limitation" section: "Other specific elements in workplace bullying like how it occurred (verbal or written by e-post or social media), who the perpetrators were (superiors, doctor colleagues, other personal, patients, relatives or friends of patients) or how long the bullying lasted might be also useful information, but was not obtainable in the present study."

Comment: Discussion: Something needs to be said about the much older age of this sample compared to others. Most bullying studies include mainly young doctors. For this reason, and because every study uses different measures, I don't think too much space should be given to cross-nation comparisons. Having said that, I do think the differences in Table 5 are so large to warrant some discussion. It's very interesting that the Scandinavian studies have the lowest levels – perhaps a follow-through of the fact that they have the best mental health in their young people too, so I do think it's good to note that. I would, however, omit comparison of the small groups within the sample with those in other studies. (p.12, l.5-11), apart from the consistency of bullying in surgery.

Response: We agree with you that less space should be given to cross-nation comparisons.

Therefore, we removed the Table 4 (Prevalence of being subjected to bullying or mobbing at work among doctors in selected countries).
Page 12, line 16 to Page 13, line 4:
In addition, we also revised and shortened the "Comparison with other studies" section.
Page 13, line 20 to Page 14, line 3:
We revised the whole "Explanation of results" section. We explain the lower prevalence of perceived workplace bullying at the population level, and explicitly in the Norwegian medical profession compared with some other countries, by variations in psychosocial working conditions.

Comment: It would be better to omit any discussion of findings where numbers are particularly small (l.18-19).

Response: Page 15, line 2-7:
Discussion of the findings was moved into the "Discussion" section.

Comment: P.13, line 10 – Substitute "On the other hand…' for 'Furthermore…". There are arguments that stress/job dissatisfaction are measuring the same thing which might be neuroticism, so it might not be too inappropriate to say that this might affect perceptions of bullying too or, on the other hand, people who are stressed and unhappy may make good victims for others.

Response: Page 13, line 21-28:
We have revised this paragraph.

Comment: p.14, line 19-20. I don't understand the reference to 'Different tolerance levels…' and there isn't a reference. I would like more around this point (see above).

Response: Page 16, line 10-15:
This paragraph has now been revised according to your suggestion.

Comment: I don't think there are strong policy implications from the study as it stands, except the new understanding that older and more senior doctors are seeing themselves as being bullied, along with the need for surgeons to appreciate the unnecessary levels of bullying in their specialty – consistently high round the world.

Response: Page 16, line 5-14:
The Policy implication section has been revised.

Comment: There are a number of typos or errors in words which can be sorted by the editors. The language is remarkably good on the whole. I do think this remarkable data set adds a lot to our understanding but probably after a little more analysis.

The either/or questions on the review sheet really don't work for the points I'm making here. For example, the statistics are adequate, but I think they could be taken much further, so I've said 'no'. I'm suggesting other things to do with the data statistically, but really it would be good to have a statistician look at that to make suggestions. I do think some of the discussion is based on too small numbers. In conclusion, bullying in healthcare is a very important issue and this is a remarkable dataset, so I hope a version of this paper will be published.

Reviewer 3: Neill Thompson

Comment: Rost and Aasland's manuscript is a very interesting study and certainly worthy of publication in BMJ Open in a revised format. It follows previous publications in BMJ Open in the area of workplace bullying in healthcare but is not simply a duplication of these previous studies as it offers a different insight.

That said, I do feel that the manuscript needs work in content and overall message

Comment: The findings in relation to neuroticism confirms earlier research findings. However, I do not see this study adding anything that we haven't already gained using more comprehensive measures of neuroticism and more developed designs. Arguably this is the least noteworthy of the findings and I think the readership would be less interested in this.

Also the authors may wish some consideration of (Page 14-Line 19-22)("neuroticism was a significant predictor of reporting bullying" this needs to be reviewed in light of the stronger evidence by Nielsen and Knardahl (2015) and Nielsen, Glaso et al (2016). While obviously this should not be removed from the study, as it was a planned and tested hypothesis, I am not convinced it adds much to earlier study findings.

Response: Page 16, line 10-15:
This paragraph has been now revised.

Comment: The strength of the paper is the data that has been collected across the four cohorts and the findings directly related to that. This is of course a great undertaking. It has some clear methodological issues in that it doesn't meet the gold standard we might consider in terms of measuring bullying (combining a behavioural checklist with a global measure – Nielsen et al (2011). I think this does need to be acknowledged as a limitation– however, alongside the caveat that these standard came in after the data collection commenced and would have been difficult to have been incorporated into the current wider survey design should also be included.

Response: Page 16, line 24-28:
The "Limitation" section has now been revised, partly according to your suggestion.

Comment: Some of the literature is dated around personality and I would certainly recommend reviewing your findings in the context of the papers above on personality (e.g. the Neilsen et al 2016 paper and (Persson, Hogh et al. 2009).

Response: We reviewed our findings on the associations between neuroticism and bullying in context of the suggested references.
Page 4, line 24-26 ("Introduction")
Page 16, line 10-15 ("Explanation of results")
Page 17, line 11-14 ("Policy implication")

Comment: There seems to be two interesting messages –
i) the cohort of doctors bullied in 1993 still report being bullied to the same prevalence in 2014.
ii) There are consistent bullying levels reported across the cohorts over more than a decade.

Response: Page 5, line 22 to Page 6, line 9:
Hopefully it is now made clearer that none of the respondents in 1993 were included in later surveys. However, the reported levels in the different groups and at different times are relatively consistent.

Comment: These findings are the core message of the study but I don't think this comes across as well as it could do in the discussion. I do feel these points could be explored further under the explanation of results:

Response: The "Explanation of results" section has been revised, partly according to your suggestion.

Comment: - Has the Norwegian healthcare services (or more broadly at Government level with legislation) not attempted any bullying interventions during this time, if they have then does this not reflect that these interventions aren't working if bullying levels are the same?

Response: Page 15, line 9-26:
Thank you for this important comment. We have now added information about main health care reforms, protective legislation and governmental anti-bullying attempts.

Comment: - Environmental factors are discussed but healthcare and the workplace environment itself has moved on in the last 15 years, bullying it seems has not. What has not changed in the healthcare setting that means still bullying remains constant when all around it has improved?

Response: Page 16, line 1-8:
We have also included some important environment factors that have not changed during the last 20 years.

Comment: - Why does the 1993 cohort not report much lower levels of experience of bullying? – this can't be tested but could be considered in the explanation of findings. Is it because they are now the seniors but their leaders are still bullies , or is it the culture as such that facilitates the bullying

- You would expect the 1993 cohort who were bullied to then do something about it. The constant levels seems to be stable suggesting that is not the case. Are the doctors in 1993 who are now in 2014 now the bullies or just staying quiet? Maintaining the bullying culture? For example - (Musselman, MacRae et al. 2005).

Response: Page 15, line 22-26:
We have attempted to make the role of the working culture clearer.

Page 17, line 11-14:
The "Policy implication" section has been revised, partly based on your input here.

References

Musselman, L. J., H. M. MacRae, et al. (2005). "You learn better under the gun": intimidation and harassment in surgical education." Medical education 39(9): 926-934.

Nielsen, M. B., L. GlasÃ¸, et al. (2016). Exposure to workplace harassment and the Five Factor Model of personality: A meta-analysis." Personality and individual differences 104: 195-206.

Nielsen, M. B. and S. Knardahl (2015). "Is workplace bullying related to the personality traits of victims? A two-year prospective study." Work & Stress 29(2): 128-149.

Nielsen, M. B., G. Notelaers, et al. (2011). Measuring exposure to workplace bullying. Bullying and harassment in the workplace: Developments in theory, research, and practice. S. Einarsen, H. Hoel, D. Zapf and C. L. Cooper. 2: 149-174.

Persson, R., A. Hogh, et al. (2009). "Personality trait scores among occupationally active bullied persons and witnesses to bullying." Motivation and Emotion 33(4): 387-399.

VERSION 2 – REVIEW

REVIEWER	Neill Thompson Northumbria University, UK
REVIEW RETURNED	26-Oct-2017

GENERAL COMMENTS	I have focused on the original comments that I made during thre review. Although I have also noted some of the comments from the other two reviewers I have focused my review on the areas I picked up on. Generally the authors have addressed all of the concerns I raised. I think when I first read this I intepreted the study as a partly longitudinal design and the revisions in title and other areas have helped clarify this. My concerns on the neuroticism treatment has also been addressed. Overall, I feel that the merit of the study fundamentally is comparing the group levels at different time points and seeing little fundamental difference and then placeing this in the healthcare context (which has been added from page 15 onwards) where a range of reforms have been performed. Sadly, the overwhelming message is - lots of effort but no change, which is worthy of reporting on in the publication. The authors have achieved this, so I have no problem in supporting this revision.