PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	How do doctors in the Netherlands perceive the impact of disciplinary procedures and disclosure of disciplinary measures on their professional practice, health, and career opportunities? A questionnaire among medical doctors who received a disciplinary measure
AUTHORS	Laarman, Berber; Bouwman, Renée; de Veer, Anke J. E.; Hendriks, Michelle; Friele, Roland

VERSION 1 – REVIEW

REVIEWER	Cynthia Krom Franklin & Marshall College USA
REVIEW RETURNED	24-Jul-2018

GENERAL COMMENTS	I found this paper to be interesting, and the research questions to be intriguing. I am impressed that you were able to gain the full cooperation of the disciplinary board, particularly to send out the letters, and that they seemed interested in the results. I think this is an important area to study. Your literature review is very fine, and Box 1 is helpful. I do have several minor, and several major, suggestions to enhance quality of the paper. I will refer to line numbers whenever possible. **Minor changes:** Abstract, line 25: Since this is the first thing I think it is very important that you clarify that

	these are disciplinary **Netherlands** (or for was the doctors disciplined, licensed disciplines know about. clarified on “all doctors measure **Medical** the organization). all as the implied, Netherlands since some elsewhere that you [note page that 4 who received a **from the Disciplinary Board**…” correct wording would be I assume it disciplined doctors in or even all the who were those may be and received wouldn’t even this should also be line 32]
	Abstract, line 33: I do not find it at all remarkable that doctors did not agree with the judges’ judgments. I think most convicted criminals also do not agree with the judgment. If anything is remarkable, I am surprised that 22.6% of those warned thought the judgment was correct. I think you need to clarify what you find remarkable, and why.
	Abstract, line 53: True, the study population was not large, but you made an attempt to study 100% of that population (worth mentioning). Your study sample, more importantly, was not large.
	Methods, p 4, line 48: The envelope cannot both be blank and have the word “confidential” printed on it. Perhaps you mean a plain white envelope?
	“Results” lines 39 – 43: you present an interesting comparison of the respondents to the Dutch doctors as a whole. It would be more useful to know how the group compares to the population that was disciplined. If the people disciplined are 78.7% male, then your sample is truly representative of the population you are trying to study. Also, how does the breakdown between warning and reprimand compare to the study group?
	Page 7, line 52: is having a quarter of people absent from work in a 12-month period unusual? We need a point of comparison. It seems to me that just

about everyone day of personal I work reasons. know each misses year at for least health a or

Major changes:

Abstract, not Really, functioning address action. not, section lines do your the The as needs 49 what questions are impact other I to – you the of questions will be 51: say about only the explain re-written. You you professional ones disciplinary you later. actually did did. that ask This do

Introduction, above. lines 30-31: Same issue as #1

Aim and You test think in they of how which Table and but comparison you research did this. back the felt the someone could 2. published you have not You to past then. disciplinary fully The simply of serious questions, do asked something and Surely, feels explain total together) cannot the threats line this. You individuals that describe the action in your percentages do two to 21: did not to happened how outcome would influence retrospect, results are statistical groups validity. (warned interesting, because

Aim and 25: what tested. the action so a worse. the health. health entirely depressed, mistakes action, made harsher indicate research Again, you Your doctors' (self-report) you change Therefore, impact This questions possible burned that those and bigger punishment. otherwise. you are measurement health cannot in of is you didn't claiming due is make health the true as that out, lead who questions, really of to not for you for well. doctors etc., to were that have lines test to the the even assumption better did not disciplinary the It who made the more resulted no 22 – some of have change in the disciplinary presented, or measure action on mental is were the disciplinary depressed in data to

	Similarly, you have some confounding in the constructs of "colleagues who don't want to collaborate…" since this was not determined prior to the disciplinary action. Maybe the doctors are just not good to work with or respected, and no one ever wanted to work with them. And "consequences for career opportunities" is so vague I don't have any idea what it means – do they have improved or damaged opportunity? If I can't tell from the question, how could the people being surveyed?
	Page 8, I am totally clueless about what you mean by negative impact on professional functioning. Do you mean as they are no longer as competent physicians, and they are functioning less well in that position? I don't think so, since your Table 4 indicates that they are taking better notes, doing more research, working according to protocol, etc., all of which sound like better professional functioning. What professional functioning was negatively impacted? You need to clarify what you mean.
	Table 5: Why would respondents be able to choose multiple options? If a question can be answered with Yes, No, I don't know, or N/A, it seems to me those are mutually exclusive answers. Could someone really answer both Yes and No? I find this whole thing very confusing.
	Other: In light of the above major suggestions, I think you need to simply re-write the paper through the lens of accuracy and with consideration of what you really tested and are trying to determine. Your whole discussion section would need to then reflect those changes.
	I am intrigued by the idea that the Dutch system aims to improve quality of care rather than removing incompetent doctors. I think having a reference to the mission statement (or something like

	that) of the Dutch medical review board would be very useful to support this and help non-Dutch understand the position.

REVIEWER	Lynley Anderson University of Otago New Zealand
REVIEW RETURNED	12-Aug-2018

GENERAL COMMENTS	This is an interesting manuscript. I did wonder whether those doctors who had received a complaint might be more likely to be suffering from burnout, be somehow unwell, or perhaps lacking in insight thereby making errors more likely leading to complaints. This issue wasn't raised in the discussion or the limitations section - could this have been an issue? This does need to be addresssed. I wondered about the value of reporting whether the participants thought the judge had got it right - I felt that most people would think that the judge had made a mistake, especially if they are going to experience disciplinary action and negative consequences. I agree that the gender split was interesting - and deserves more attention. Future study might include getting the perspectives of patients. What do they want from a complaints structure? What do they think about the outcomes of their complaint?

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Cynthia Krom

Minor changes:

1) Abstract, line 25: Since this is the first thing the reader sees, I think it is very important that you clarify thatthese are "all doctors who received a disciplinary measure from the Netherlands Medical Disciplinary Board…"
(or whatever the correct wording would be for the organization). I assume it was not all the disciplined doctors in the world, as implied, or even all the doctors in the Netherlands who were disciplined, since some of those may be licensed elsewhere and received disciplines that you wouldn't even know about. [note that this should also be clarified on page 4 line 32]

We've added the sentence 'from the Dutch disciplinary board' to specify our study sample.

2) Abstract, line 33: I do not find it at all remarkable that doctors did not agree with the judges' judgments. I think most convicted criminals also do not agree with the judgment. If anything is remarkable, I am surprised that 22.6% of those warned thought the judgment was correct. I think you need to clarify what you find remarkable, and why.

We agree with both reviewers that it might not be a remarkable result, but it is relevant, as doctors are supposed to learn from complaints. If healthcare professionals don't agree with the measure, they can hardly be expected to learn from it. We added the sentence (p.7): This might not be surprising, but it is relevant as disciplinary procedures are supposed to be a learning experience.

3) Abstract, line 53: True, the study population was not large, but you made an attempt to study 100% of that population (worth mentioning). Your study sample, more importantly, was not large.

We altered ‘population’ into ‘sample’.

4) Methods, p 4, line 48: The envelope cannot both be blank and have the word “confidential” printed on it. Perhaps you mean a plain white envelope?

We changed ‘blank’ into ‘plain white envelope without sender address’.

5) “Results” lines 39 – 43: you present an interesting comparison of the respondents to the Dutch doctors as a whole. It would be more useful to know how the group compares to the population that was disciplined. If the
people disciplined are 78.7% male, then your sample is truly representative of the population you are trying to study. Also, how does the breakdown between warning and reprimand compare to the study group?

Unfortunately, due to the privacy considerations as explained before concerning the non-response analysis, we have no information about the study population as a whole (i.e., all disciplined healthcare professionals). Similarly, as we lack information about non-responders, we can’t compare the break down between warning and reprimand in our study sample to the study group.

6) Page 7, line 52: is having a quarter of people absent from work in a 12-month period unusual? We need a point of comparison. It seems to me that just about everyone I know misses at least a day of work each year for health or personal reasons.

In the original manuscript, we compared our results to The National Survey of Working Conditions (NEA) of 2015. In the NEA benchmark, 50% of respondents working in healthcare indicated that they had been absent from work at least once in the past 12 months. In our study, this was only a quarter. However, as absenteeism was low in our study we decided to leave this out of the revised manuscript.

Major changes:
1) Abstract, lines 49 – 51: You actually did not do what you say you did. Really, your questions about professional functioning are the only ones that address the impact of the disciplinary action. The other questions you ask do not, as I will explain later. This section needs to be re-written.

2) Introduction, lines 30-31: Same issue as #1 above.

Introduction (p.3): We realize that differences between the groups can be both resultant of the outcome (heavier measure) as of the disclosure of the measure, or can be mutually reinforcing as a reprimand is experienced as a heavier measure precisely because it was disclosed. Throughout the manuscript, we made a nuance by emphasizing it is about the *perceived* impact of doctors, and the *perceived* health, professional practice, etc.

3) Aim and research questions, line 21: You did not do this. You did not test this. You asked individuals to think back to something that happened in the past and describe how they felt then. Surely, the actual outcome of the disciplinary action would influence how someone feels in retrospect, which could fully explain your results in Table 2. The total percentages (warned and published together) are interesting, but you simply cannot do statistical comparison of the two groups because you have serious threats to validity.

As the reviewer rightly states, comparison of doctors with warnings and reprimands is difficult. However, as we wanted to gain insight in the experiences of doctors whose disciplinary measure was disclosed, this was the best feasible study design, as the alternative was choosing doctors receiving a reprimand before July 2012 as a comparison group. Besides, it could also be said that doctors experience a reprimand as a heavier sentence – and thus more impact – *because* it is disclosed. Our study design does, however, present some limitations which we included as follows:

Limitations (p.12) […] two groups of professionals with disciplinary measures (warning and reprimand) may not be comparable because of the context and nature of the complaint and the related culpability and judgement of the disciplinary court. Reported (mental) health issues could have been a result, or an underlying cause of complaints. The bigger the health issues, the heavier the measure and hence

the disclosure of the measure, one might reason. Respondents also might experience the measure as heavier precisely because it is publicly disclosed. Furthermore, the relationship between the measure and the outcome variables has not been analysed, but the results are self-reported by the respondents. This may be rather subjective. Therefore, a causal relationship between the disciplinary procedure and the outcome variables, or publication of the measure and the outcome variables cannot be proven.

4) Aim and research questions, lines 22– 25: Again, you didn't really test some of what you are claiming to have tested. Your measurement of the change in the doctors' health due to the disciplinary action (self-report) is not even presented, so you cannot make any assumption of a change in health for better or worse. Therefore, you did not measure the impact of the disciplinary action on health. This is true for the mental health questions as well. It is entirely possible that doctors who were depressed, burned out, etc., made the mistakes that lead to the disciplinary action, and those who were more depressed made bigger mistakes that resulted in harsher punishment. You have no data to indicate otherwise. Similarly, you have some confounding in the constructs of "colleagues who don't want to collaborate…" since this was not determined prior to the disciplinary action. Maybe the doctors are just not good to work with or respected, and no one ever wanted to work with them. And "consequences for career opportunities" is so vague I don't have any idea what it means – do they have improved or damaged opportunity? If I can't tell from the question, how could the people being surveyed?

- Concerning health: We understand the reviewers comment as follows that we did not *measure* the impact on health, but we did gain insight in the *perceived* impact as *experienced* by doctors *themselves.* We agree with the reviewer we cannot draw conclusions on causality, but we did gain insight in the experience of doctors concerning their health after the procedure. We revised as follows:

  Results (p. 7): As time passed, the perceived effect of the procedure on health diminished (a mean of 1.7 for the whole group at moment of filling out the questionnaire). The difference between the doctors receiving reprimands (2.1) and warnings (1.6) continued to exist. Differences in the impact between respondents whose judgement was issued up to one year ago and more than one year ago were not significant (not in table). As we have no information on the health of professionals prior to the procedure, the perceived change in health directly after the procedure and after the passing of time can be due to other circumstances.

- Concerning the causal connection between the procedure and (mental) health: The reviewer is right to state it is difficult to ascribe (mental) health problems to disciplinary procedures/complaints, since it is indeed entirely possible these problems led to the complaint in the first place. As Balch *et al* state quite simply: it is difficult to determine the direction of effect. We revised our manuscript to include this nuance:
  Discussion (p. 11): We are careful to jump to conclusions regarding the impact of disciplinary procedures on (mental) health, as Balch et al state in a study regarding the consequences of malpractice lawsuits, it is difficult to determine the 'direction of effect'. I.e., our data can also be explained such that mental issues led to suboptimal healthcare, leading to a complaint to a disciplinary board, with more severe mental issues resulting in a reprimand instead of a warning.
  Limitations (p.13): Reported (mental) health issues could have been a result, or an underlying cause of complaints. The bigger the health issues, the heavier the measure and hence the disclosure of the measure, one might reason. Respondents also might experience the measure as heavier precisely because it is publicly disclosed.

- Concerning the confounding of 'colleagues who don't want to collaborate': We agree with the reviewer there is some confounding in this question, which is emphasized by the manner of formulation in our manuscript. In our questionnaire, we specifically asked if doctors experienced colleagues unwilling to collaborate with them or refer to them anymore *since* the disciplinary procedure. We expressed this nuance as follows:
  p. 9: These differences were significant for loss of patients (p=0.000), fewer new patients (p=0.002), colleagues who no longer want to work with them or refer patients to them (p=0.036), and consequences for career opportunities (p=0.000) since the disciplinary procedure.

p. 11: Public disclosure also clearly led to consequences for practice, such as losing patients, getting fewer new patients and obstruction of career opportunities since the disciplinary procedure.

- Concerning career opportunities: We agree with the reviewer this question seems vague. Because 'career opportunities' can refer to an array of things we asked respondents who confirmed with 'yes' to explain their answer. In the manuscript, we added a sentence on p.11: For the latter category, examples given were not being able to get a new job or getting questioned about the reprimand by the health insurer.

5) Page 8, I am totally clueless about what you mean by negative impact on professional functioning. Do you mean they are no longer as competent as physicians, and they are functioning less well in that position? I don't think so, since your Table 4 indicates that they are taking better notes, doing more research, working according to protocol, etc., all of which sound like better professional functioning. What professional functioning was negatively impacted? You really need to clarify what you mean.
We've revised 'functioning' to 'practice', hoping this is a better translation. We asked doctors whether they made changes in the way they perform their tasks as medical doctors, giving multiple options that can be perceived as both positive and negative. We've made a clearer distinction between positive and negative as follows:

Impact on professional practice (p.8): The majority of doctors reported the disciplinary process had a negative impact on their professional practice. 71.1% of doctors given a reprimand indicated that the procedure only had a negative impact. Among doctors receiving warnings, this was significantly less, at 40.8% (p=0.004, chi2=13.19). 4.4% of doctors given a reprimand and 8.5% of doctors given a warning indicated that the procedure only had a positive impact (not in table).

Respondents reported various changes in their professional practice that are obviously negative (see Table 4): avoiding high-risk patients (47.5% with a reprimand versus 38.2% with a warning), seeing each patient as a new complainant (41.4% vs. 35.2%) and avoiding similar patients as the complainant (41.4% vs. 29%). Some changes can be perceived as positive, such as making more accurate notes in patients' files (64.2%) and discussing improvement measures with their colleagues and/or supervisor (60.8%) more often since the disciplinary process. Some reported changes can be either positive or negative according to context, but are commonly associated with defensive medicine, such as complying to patients wishes more and doing more supplementary research.

Discussion (p.12): Besides negative effects, the responding doctors also reported positive changes, such as making more accurate notes in patient records and discussing improvement measures with colleagues.

6) Table 5: Why would respondents be able to choose multiple options? If a question can be answered with Yes, No, I don't know, or N/A, it seems to me those are mutually exclusive answers. Could someone really answer both Yes and No? I find this whole thing very confusing.

This was a mistake, the answers were indeed mutually exclusive, we corrected it in the manuscript.

Other:

In light of the above major suggestions, I think you need to simply re-write the paper through the lens of accuracy and with consideration of what you really tested and are trying to determine. Your whole discussion section would need to then reflect those changes. I am intrigued by the idea that the Dutch system aims to improve quality of care rather than removing incompetent doctors. I think having a reference to the mission
statement (or something like that) of the Dutch medical review board would be very useful to support this and help non-Dutch understand the position.

The Dutch Disciplinary Board does not have a missionary statement, but aims to fulfill the two disciplinary norms as set out in Box 1. We hope this is a clear enough description of what disciplinary law in the Netherlands seeks to achieve.

Reviewer 2: Lynley Anderson

Institution and Country: University of Otago, New Zealand

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below This is an interesting manuscript.
I did wonder whether those doctors who had received a complaint might be more likely to be suffering from burnout, be somehow unwell, or perhaps lacking in insight thereby making errors more likely leading to complaints. This issue wasn't raised in the discussion or the limitations section - could this have been an issue? This does need to be addressed.

We agree with the reviewer it is difficult to determine, as the Dutch say, 'which was there first: the chicken or the egg', in this case: the reported mental issues or the complaint and following procedure, including measure? We hope to have explained this satisfactorily by answering reviewer 1 on p. 4.

I wondered about the value of reporting whether the participants thought the judge had got it right - I felt that most people would think that the judge had made a mistake, especially if they are going to experience disciplinary action and negative consequences.

We agree with both reviewer 1 and 2 and left this out of our revised manuscript.

I agree that the gender split was interesting - and deserves more attention.
Future study might include getting the perspectives of patients. What do they want from a complaints structure? What do they think about the outcomes of their complaint?

We agree with the reviewer the perspective of patients is needed to evaluate the functioning of complaints procedures. In the Netherlands, disciplinary law is explicitly aimed at quality improvement, *not* patients satisfaction, reparation or compensation. However, disclosure of disciplinary measures is supposed to be helping patients by providing them with quality information. Part of our research concerned the perspective of patients on disclosure of disciplinary measures. For the sake of the focus of the manuscript and the limited space we have, we didn't include these results, but we are happy to inform you through this letter.

We studied the patient perspective by sending a questionnaire to a patient panel. The questionnaire concerned whether 1) patients visited a healthcare professional for the first time that year; 2) patients searched for information about this healthcare professional online; 3) patients think it is important that information about disciplinary measures can be found online; 4) if, and how, they would act when they heard about their family physician got a disciplinary measure and if 5) they were a healthcare professional themselves. A majority of patients attached importance to online transparency about disciplinary measures (42%). A minority of patients that visited a new healthcare professional searched for information online (14.7%). Most patients state they would do nothing when they found out about a disciplinary measure (29%). Those who would act, indicated they would talk about the disciplinary measure with their family physician (26.8%), look up more information online (23.9%) or watch their physicians behavior more closely (22.5%). In short: if asked, patients do think it is important to be able to look up the information online, but only few of them actually search for the information online, and when they do, most patients state they don't act on it.

Also, we do have data on the opinions of patients on disclosure of disciplinary measures and on the satisfaction of patients with the disciplinary procedure. We added this in the discussion on p. 12.

VERSION 2 – REVIEW

REVIEWER	Lynley Anderson University of Otago, New Zealand
REVIEW RETURNED	27-Oct-2018

GENERAL COMMENTS	This article has benefitted from the review process. The authors have improved the article and made it much clearer for the reader.

	My ony comment is that in the discussion, the authors have said 'jump to conclusions' when they should say 'not jump to...'