PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Relationship between Physician Burnout and Quality of Healthcare in terms of Safety and Acceptability – A Systematic Review
AUTHORS	Dewa, C; Loong, Desmond; Bonato, Sarah; Trojanowski, Lucy

VERSION 1 - REVIEW

REVIEWER	Renee Scheepers Academic Medical Center, Amsterdam, the Netherlands
REVIEW RETURNED	30-Nov-2016

GENERAL COMMENTS	The authors have thoroughly revised their review. Overall, this is an interesting and informative review on a timely topic. They explained how their review specifically added to the current body of knowledge and the existing reviews on this topic. This is clearly described in the introduction. In the method section, I would like to have some more information on the exclusion criteria. The authors report that studies involving only residents or students were excluded. However, how did the authors cope with studies on a sample consisting of both physicians and residents? Were the results analyzed together or were studies then only included when results were available per sub group (physicians apart from residents)? Furthermore, the authors state in their method section that they choose a time limit from 2002 on, with the argument that this time frame fits in the current health care environment. However, this argument seems somewhat arbitrary. For the argument on the current health care environment, different time frames could have been chosen. Why for example not 2000, 2005 or 2010? What year reflects the transition to modern health care? More clarity and foundation of the chosen time frame with references on societal and/or research develepments is appropriate here. For their risk of bias assessment, the authors write in the rebuttal letter that they accounted for generalizability of the sample by describing the study sample and a discussion of the representativeness of the sample with regard to the population. Could the authors please clarify where in the manuscript they reported their disucssion on the representativeness? It does not appear to be present in the risk of bias assessment results. Furthermore, would the authors clarify how they actually determine whether a sample is representative for the population? At the least, the criterium on the number of institutions used in the eligible studies should be taken into account when assessing the generalizability of the samples. Furthermore, in the risk of bias assessment results, the authors report that all outcome measures are clearly defined. However, they

	do not report on the psychometric quality of the measures, the studied validity and reliability of the scales used and whether they are self-reported (such as in the case of self-reported medical errors). Although the authors do reflect on the potential bias of self-reported errors in the Discussion section, these and other considerations on the measures used would strengthen the risk of bias assessment. In addition, it would provide clear overview when table 1 and table would be merged into one table. Then the study findings can be clearly interpreted by readers in the light of the study methods used. The authors indicate that moderate evidence is present for the association on physician burnout and safety aspects of provided health care. This statement would be strenthened by also reflecting on the size of the associations (such as the regression coefficients or size of odds ratios), not merely statistical significance. In the Discussion section the authors suggest that their review can help to inform decisions about how to improve patient care by adressing burnout. It would be interesting to clarify this sentence by explaining exactly how this would help. In what way exactly can the current conclusions help in informing decisions?

REVIEWER	Jef Adriaenssens Leiden University, The Netherlands KCE, Brussels, Belgium
REVIEW RETURNED	03-Dec-2016

GENERAL COMMENTS	Thanks for giving met the opportunity to review the manuscript. The study is interesting and the quality of the paper is good. I only have a few minor comments (see under) - Abstract: Please add a bit of information regarding the time of the search - page 4 line 19: ...and they define it as... ==> ...and defined as ... - page 5 line 8: the two dimensions ==> remove 'the' - page 5 line 39: please add a line to explain the differences between the reviews. - page 5 line 43: from ours ==> from our review - page 5 line 43: articles from ours ==> remove 'from ours" - page 8 line 3 to 17: the list of specialists is to my opinion redundant (not an added value for the paper). Please consider to remove this. - page 8 to 11: regarding the table. Is it possible to remove this to an appendix? - page 18 line 13: not all authors used the MBI 22-item scale. One author used the 17 item. Is this the MBI-HSS or the GS version? - page 25 line 27 with ==> had

REVIEWER	Louise Hall University of Leeds, UK
REVIEW RETURNED	08-Dec-2016

GENERAL COMMENTS	I do not have much to add to the previous reviewers comments, which I can see have been adequately addressed in the revised version of the manuscript. I agree that the focus on physician-patient

	interactions and the acceptability of these is an important area that provides a novel contribution to the literature. However I am struggling to see the novel aspect of the safety articles included, other than one paper (Weigl et al, 2015) that hasn't been previously included in a review. There is an additional recent review on these topics that may warrant inclusion in your introduction, although I understand it was published around the time yours was under review: Salyers, M.P., Bonfils, K.A., Luther, L. et al. J GEN INTERN MED (2016). doi:10.1007/s11606-016-3886-9 1. Throughout the introduction you state that you are concerned with quality of care and physician-patient interactions. However you are also including studies related only to safety, which although related, differs to quality of care and isn't necessarily related to patient-physician interaction e.g. prescription errors. I would suggest rewording of the last sentence of the Introduction section to clarify that you are reviewing both quality and safety. 2. For quality of care you have included a variety of measures that reflect different aspects of quality of care, however for safety you focus solely on medical errors despite safety being a broad term with a variety of measures employed within the literature. Did you consider widening your inclusion criteria to include other measures of safety such as SSIs, UTIs, bed sores, safety grade/culture?

REVIEWER	James Beguah Odei The Ohio State University, College of Public Health, Division of Biostatistics, 241 Cunz Hall, 1841 Neil Avenue, Columbus, OH-43210 USA
REVIEW RETURNED	16-Jan-2017

GENERAL COMMENTS	Look forward to your future research in evaluating burnout interventions for physicians and continued work at the relationship between dimensions of acceptability-related quality of measures and burnout.

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1: RENEE SCHEEPERS

IN THE METHOD SECTION, I WOULD LIKE TO HAVE SOME MORE INFORMATION ON THE EXCLUSION CRITERIA. THE AUTHORS REPORT THAT STUDIES INVOLVING ONLY RESIDENTS OR STUDENTS WERE EXCLUDED. HOWEVER, HOW DID THE AUTHORS COPE WITH STUDIES ON A SAMPLE CONSISTING OF BOTH PHYSICIANS AND RESIDENTS? WERE THE RESULTS ANALYZED TOGETHER OR WERE STUDIES THEN ONLY INCLUDED WHEN RESULTS WERE AVAILABLE PER SUB GROUP (PHYSICIANS APART FROM RESIDENTS)?

The reviewer is concerned about the treatment of studies that included both physicians and residents. We handled this with one of our inclusion criteria. One of the inclusion criteria was that the results

were reported in a way that separated practicing physicians from residents. Any paper that did not do that was not included. To address any remaining confusion, Criterion 2 was clarified to read:

2. The sample population was comprised of practicing physicians regardless of specialty who worked in civilian settings. That is, the results were reported such that the practicing physician (as opposed to resident) outcomes were reported separately.

FURTHERMORE, THE AUTHORS STATE IN THEIR METHOD SECTION THAT THEY CHOOSE A TIME LIMIT FROM 2002 ON, WITH THE ARGUMENT THAT THIS TIME FRAME FITS IN THE CURRENT HEALTH CARE ENVIRONMENT. HOWEVER, THIS ARGUMENT SEEMS SOMEWHAT ARBITRARY. FOR THE ARGUMENT ON THE CURRENT HEALTH CARE ENVIRONMENT, DIFFERENT TIME FRAMES COULD HAVE BEEN CHOSEN. WHY FOR EXAMPLE NOT 2000, 2005 OR 2010? WHAT YEAR REFLECTS THE TRANSITION TO MODERN HEALTH CARE? MORE CLARITY AND FOUNDATION OF THE CHOSEN TIME FRAME WITH REFERENCES ON SOCIETAL AND/OR RESEARCH DEVELEPMENTS IS APPROPRIATE HERE.

The reviewer is concerned that the time limits were arbitrary. It might be interesting to note that although our search could have included papers published in 2002, the first relevant paper that our search identified was published in 2005. To address any remaining concern, we also include an explanation for our choice of time period. The rationale is more clearly stated in the text:

The time frame was chosen to represent the current healthcare environments in which physicians are practicing. For example, the year 2002 was the year after the Institute of Medicine's report14 on the quality of healthcare that discussed the six dimensions of quality of care. By beginning in 2002, we have allowed for a one year lag after publication of this report during which healthcare settings and researchers could have incorporated the Institute of Medicine's quality of healthcare framework into their work.

FOR THEIR RISK OF BIAS ASSESSMENT, THE AUTHORS WRITE IN THE REBUTTAL LETTER THAT THEY ACCOUNTED FOR GENERALIZABILITY OF THE SAMPLE BY DESCRIBING THE STUDY SAMPLE AND A DISCUSSION OF THE REPRESENTATIVENESS OF THE SAMPLE WITH REGARD TO THE POPULATION. COULD THE AUTHORS PLEASE CLARIFY WHERE IN THE MANUSCRIPT THEY REPORTED THEIR DISUCSSION ON THE REPRESENTATIVENESS? IT DOES NOT APPEAR TO BE PRESENT IN THE RISK OF BIAS ASSESSMENT RESULTS. FURTHERMORE, WOULD THE AUTHORS CLARIFY HOW THEY ACTUALLY DETERMINE WHETHER A SAMPLE IS REPRESENTATIVE FOR THE POPULATION? AT THE LEAST, THE CRITERIUM ON THE NUMBER OF INSTITUTIONS USED IN THE ELIGIBLE STUDIES SHOULD BE TAKEN INTO ACCOUNT WHEN ASSESSING THE GENERALIZABILITY OF THE SAMPLES.

The reviewer seeks to confirm that we discuss the representativeness of the studies. This concern was addressed in two parts of the paper.

First, we discuss the representativeness within the Results section with the text:

Only one study comprehensively5 described the study population from which the study sample was drawn. Only one study used longitudinal data.28 Other limitations involved not reporting the response rate29-32…

Second, the first item that we used to assess the risk of bias looks to see whether the authors describe the study population. This item is important to understanding the representativeness of the population. We note that only one study did this. In the Strengths and Limitations of Interpreting the

Literature section, there is the text:

Finally, only one study5 described the population from which the study sample was drawn. Thus, it is difficult to determine whether there was a difference between the study participants and non-participants. To aid in the interpretation of the results (i.e., the generalizability), it would be useful for future studies to report this type of information.

The reviewer also seems to suggest that generalizability is related the number of institutions participating in the study. In and of itself, seeking to have a great number of institutions seems to miss the mark. Quantity does not guarantee generalizability. If only one institution is selected and that one institution selected represents all the institutions in the population of interest, its results would be generalizable. In the interest of understanding the generalizability of results, it is more informative to have information about the population from which the study population is sampled. This was addressed when we took that into account in our risk of bias assessment.

FURTHERMORE, IN THE RISK OF BIAS ASSESSMENT RESULTS, THE AUTHORS REPORT THAT ALL OUTCOME MEASURES ARE CLEARLY DEFINED. HOWEVER, THEY DO NOT REPORT ON THE PSYCHOMETRIC QUALITY OF THE MEASURES, THE STUDIED VALIDITY AND RELIABILITY OF THE SCALES USED AND WHETHER THEY ARE SELF-REPORTED (SUCH AS IN THE CASE OF SELF-REPORTED MEDICAL ERRORS). ALTHOUGH THE AUTHORS DO REFLECT ON THE POTENTIAL BIAS OF SELF-REPORTED ERRORS IN THE DISCUSSION SECTION, THESE AND OTHER CONSIDERATIONS ON THE MEASURES USED WOULD STRENGTHEN THE RISK OF BIAS ASSESSMENT.

The reviewer is concerned that the data source for the outcomes data was not reported.
This was addressed in the reviewed version of the manuscript where we noted in the “Description of Studies” the source of information for the Quality of Care Measure. For example, it notes, “Patient report:” or “Physician report”.

In addition, the reviewer would like to see that the risk of bias criteria takes into account the psychometric quality of the measures. To address this, we revised Item 4 from in our Risk of Bias Assessment to focus on the psychometric properties of the outcomes studied.

The reviewer also requested more information about the psychometric properties of the scale. To address this, we included information about the psychometric properties of the scales used in the “Description of Studies” table. In addition to the Discussion section to which the Reviewer refers, we also addressed this by focusing a great deal of attention to discussing the outcome measures and the variability in the measures used. The text reads:

Measuring Quality of Care related to Acceptability and Patient Safety
Four types of quality of care measures related to acceptability and safety were used in these studies. In terms of patient safety, medical errors were measured. Acceptability related measures included patient satisfaction, perceived general quality of care, and physician communication/attitudes.

Patient Safety Measures: Medical errors
Patient safety was examined with medical errors. This outcome was assessed in five studies.5,29,30,42,46 Wen et al.46 asked respondents whether they had made any medical errors including one that resulted in a patient being harmed, a medication error, delay in treatment, or incomplete or incorrect item being added to the patient record. Hayashino et al.30 and Shanafelt et al.42 used similar questions about whether the respondent made major medical errors. However, the studies differed in the time frame that the respondent was asked to consider. Hayashino et al.30 asked about the past year while Shanafelt et al.42 inquired about the past three months. In contrast to

these studies, Klein et al.5 asked about frequency of diagnostic mistakes and treatment without specifying a time frame. The studies differ in the types of errors that they asked about (i.e., major errors rather than any errors). In addition, they depend on recall and self-report. Shanafelt et al.42 note that studies have used this type of question to gather information about medical errors. However, there are also studies that have found that physicians under-report medical errors.53 Furthermore, there is evidence that physicians have a limited ability to self-assess their practice patterns.54
In addition to questions about frequency of diagnostic mistakes and treatment, Klein et al.5 included a questionnaire based on the Canadian Physician Achievement Review to evaluate physician self-perceived quality of psychosocial care, diagnosis/therapy, and quality assurance.55 However, the authors note that additional work regarding its validity is warranted.5
There was only one study that did not rely on self-report to gather information about medical errors. Rabatin et al.29 used a chart audit to assess medical errors characterized by adherence to guidelines, responsiveness to "recurrent abnormalities" and missed drug interactions.

Acceptability Measures: Patient satisfaction/Perceived Quality of Care
With regard to acceptability measures, patient satisfaction was assessed in four studies.31,32,35,43 In two of these studies, the SERVQUAL was used to measure patient satisfaction/quality of care.32,43 The SERVQUAL was developed to measure service quality along five dimensions: (1) tangibles (i.e., physical facilities), (2) reliability (i.e., performs dependably and accurately), (3) responsiveness (i.e., willingness to help), (4) assurance (i.e., ability to inspire trust), and (5) empathy (i.e., caring).56 Halbesleben and Rathert32 used a healthcare specific version of the SERVQUAL. The psychometric properties of the scale were examined.38 However, Asubonteng et al.39 have raised questions about the strength of the scale's psychometric properties.
Shirom and colleagues43 adapted the SERVQUAL by eliminating seven items and revising the language for physicians to rate their own quality of care using the remaining 15 items. The validity of this modified measure was not examined.
Weigl et al.45 looked at physician-perceived quality of care by asking physicians to rate two statements on a 5-point scale, "My workload frequently leads to reduced quality of work," and "Adverse work conditions frequently lead to a loss of quality." The authors reference the German version of the MBI as the source for these questions. However, they do not provide information about the psychometric properties of the individual use of these items.
One study31 used the Consultation Satisfaction Questionnaire (CSQ) scale that was created and validated to assess patient satisfaction with general practitioners.36 It is comprised of 18 items and measures satisfaction along four dimensions: general satisfaction, professional care, depth of relationship, and perceived time.
Finally, in their study, Weng et al.35 used two questions to indicate patient satisfaction, "I am satisfied with the care provided by my doctor," and "I would recommend this doctor to my friends and family." The first of Weng et al.'s35 question is similar to one of the CSQ's36 general satisfaction items, "I am totally satisfied with my visit to the doctor." However, the use of this single-item has not been validated. A version of the second question has been used to measure satisfaction and was correlated with the EUROPEP patient satisfaction questionnaire.47

Acceptability Measures: Communication/Attitudes
Two studies focused on physician communication/attitudes.33,34 Using audiotapes of physician/patient interactions, Ratanawongsa et al.33 assessed the interactions by employing the Roter Interaction Analysis System (RIAS).57 RIAS is a validated method of categorizing these interactions into three categories related either to content, affection, or process.58 There is evidence that there is an association between the content and the socioemotional nature of the interactions as categorized using the RIAS and patient satisfaction.57,58
Travado et al.34 examined the association between burnout and communication using two measures: the Self-Confidence in Communications Skills (SCSS) and the Expected Outcomes of Communication (EOC).44 In their article, Parle and colleagues44 note that exploration of the psychometric properties

of both measures were being conducted but were not yet completed. Both were developed to understand the communication skills of physicians working with cancer patients.

IN ADDITION, IT WOULD PROVIDE CLEAR OVERVIEW WHEN TABLE 1 AND TABLE WOULD BE MERGED INTO ONE TABLE. THEN THE STUDY FINDINGS CAN BE CLEARLY INTERPRETED BY READERS IN THE LIGHT OF THE STUDY METHODS USED.

It is not clear what tables the Reviewer refers to in this comment. In the version of the manuscript that was reviewed, Table 1 contained the search terms. If this comment referred to the search terms, the table has been moved to the Appendix.

THE AUTHORS INDICATE THAT MODERATE EVIDENCE IS PRESENT FOR THE ASSOCIATION ON PHYSICIAN BURNOUT AND SAFETY ASPECTS OF PROVIDED HEALTH CARE. THIS STATEMENT WOULD BE STRENTHENED BY ALSO REFLECTING ON THE SIZE OF THE ASSOCIATIONS (SUCH AS THE REGRESSION COEFFICIENTS OR SIZE OF ODDS RATIOS), NOT MERELY STATISTICAL SIGNIFICANCE.

The reviewer is concerned about the strength of the association between burnout and the outcome measure. Our assessment of “moderate evidence” is based on the strength of the study design. Given the risk of bias scores were mediocre, the size of the association would not have influenced our assessment. The values of the statistical tests can be found in Table 2.

IN THE DISCUSSION SECTION THE AUTHORS SUGGEST THAT THEIR REVIEW CAN HELP TO INFORM DECISIONS ABOUT HOW TO IMPROVE PATIENT CARE BY ADRESSING BURNOUT. IT WOULD BE INTERESTING TO CLARIFY THIS SENTENCE BY EXPLAINING EXACTLY HOW THIS WOULD HELP. IN WHAT WAY EXACTLY CAN THE CURRENT CONCLUSIONS HELP IN INFORMING DECISIONS?

The reviewer would like greater elaboration about how our results could inform decision-making. As requested, additional text was added to clarify:

The focus on quality related to direct care can highlight additional ways that physician burnout affects the healthcare system. These results contribute evidence about whether the effects of physician burnout are limited to physicians or whether consequences of physician burnout are more extensive. They also can help to inform decisions about how to improve patient care by addressing physician burnout. That is, decisions can be informed when confronting a question of how to improve quality of patient care. There are a number of ways in which this may be done through investment in capital such as new technologies. The results of this systematic review suggest that an alternative investment could be in human resources as represented by physician staff.

REVIEWER 2: JEF ADRIANENSSENS

ABSTRACT: PLEASE ADD A BIT OF INFORMATION REGARDING THE TIME OF THE SEARCH

As requested, more information was added:

Using a multi-phase screening process, this systematic literature review was based on publically available peer-reviewed studies published between 2002- 2017.

PAGE 4 LINE 19:… AND THEY DEFINE IT AS… \ … AND DEFINED AS…

The requested change was made.

PAGE 5 LINE 8: THE TWO DIMENSIONS \ REMOVE “THE”

As requested, the change was made.

PAGE 5 LINE 39: PLEASE ADD A LINE TO EXPLAIN THE DIFFERENCES BETWEEN THE REVIEWS.

As requested, the explanation was added:

For example, Hall et al.18 consider healthcare staff wellbeing and Salyers et al. 20 examine staff burnout as opposed to specifically examining physician burnout as our review does. de Jong et al.17 examine common mental disorders as opposed to burnout. Williams and Skinner19 look at physician satisfaction rather than burnout.

PAGE 5 LINE 43: FROM OURS \ OUR REVIEW

As requested, the change was made.

PAGE 5 LINE 43: ARTICLES FROM OURS \ REMOVE ‘FROM OURS”

As requested, the change was made.

PAGE 8 LINE 3 TO 17: THE LIST OF SPECIALISTS IS TO MY OPINION REDUNDANT (NOT AN ADDED VALUE FOR THE PAPER). PLEASE CONSIDER TO REMOVE THIS.

The reviewer suggested that the list of specialties be deleted from the manuscript. We have not removed this list because it was added at the request of a reviewer of the previous review.

PAGE 8 TO 11: REGARDING THE TABLE. IS IT POSSIBLE TO REMOVE THIS TO AN APPENDIX?

As requested, the search strategy was moved to the Appendix.

PAGE 18 LINE 13: NOT ALL AUTHORS USED THE MBI 22-ITEM SCALE. ONE AUTHOR USED THE 17 ITEM. IS THIS THE MBI-HSS OR THE GS VERSION?

The reviewer requested more information about the Japanese scale. The 17-item scale was a Japanese translation based on the MBI-HSS. This information was added to the “Description of the Studies” table.

PAGE 25 LINE 27 WITH \ HAD

As requested, the change was made.

REVIEWER 3: LOUISE HALL

I DO NOT HAVE MUCH TO ADD TO THE PREVIOUS REVIEWERS COMMENT, WHICH I CAN SEE HAVE BEEN ADEQUATELY ADDRESSED IN THE REVISED VERSION OF THE MANUSCRIPT. I AGREE THAT THE FOCUS ON PHYSICIAN-PATIENT INTERACTIONS AND THE ACCEPTABILITY OF THESE IS AN IMPORTANT AREA THAT PROVIDES A NOVEL CONTRIBUTION TO THE LITERATURE. HOWEVER, I AM STRUGGLING TO SEE THE NOVEL ASPECT OF THE SAFETY ARTICLES INCLUDED, OTHER THAN ONE PAPER (WEIGL ET AL. 2015) THAT HASN'T BEEN PREVIOUSLY INCLUDED IN A REVIEW.

The reviewer is concerned about the contribution of this paper. First, we would like to point out that our question is different from previously systematic reviews. Thus, we would contend that the contribution is based on the fact that we asked a different question. Because of the difference, we developed a different search strategy that yielded a set of papers that are unique from the other reviews and a different discussion of the literature as it relates to the question that we posed. It is not just the difference in the number of papers reviewed. However, it should be noted that we updated the search to include 2016. This resulted in the addition of two papers and a more current search.

Second, we used a comprehensive search strategy guided by best practices as suggested by McGowen (2016) and Cochrane (2011). In comparison to existing reviews, ours is more comprehensive incorporating the recommended adjacency commands and synonyms for keywords. As such, the search strategy used in our review is a contribution to the literature and will help those who seek to replicate the search in the future as the literature in this area grows.

THERE IS AN ADDITIONAL RECENT REVIEW ON THESE TOPICS THAT MAY WARRANT INCLUSION IN OUR INTRODUCTION, ALTHOUGH I UNDERSTAND THAT IT WAS PUBLISHED AROUND THE TIME YOURS WAS UNDER REVIEW: SALYERS, M.P., BONFILS, K.A., LUTHER, L. ET AL. J GEN INTERN MED (2016). DOI: 10.1007/S11606-016-3886-9

The reviewer has suggested including an additional systematic review. Thank you for bringing this to our attention. Reference to it has been added:

For example, Hall et al.18 consider healthcare staff wellbeing and Salyers et al. 20 examine staff burnout as opposed to specifically examining physician burnout as our review does. de Jong et al.17 examine common mental disorders as opposed to burnout. Williams and Skinner19 look at physician satisfaction rather than burnout.

1. THROUGHOUT THE INTRODUCTION YOU STATE THAT YOU ARE CONCERNED WITH QUALITY OF CARE AND PHYSICIAN-PATIENT INTERACTIONS. HOWEVER YOU ARE ALSO INCLUDING STUDIES RELATED ONLY TO SAFETY, WHICH ALTHOUGH RELATED, DIFFERS FROM QUALITY OF CARE AND ISN'T NECESSARILY RELATED TO PATIENT-PHYSICIAN INTERACTION E.G., PRESCRIPTION ERRORS. I WOULD SUGGEST REWORDING OF THE LAST SENTENCE OF THE INTRODUCTION SECTION TO CLARIFY THAT YOU ARE REVIEWING BOTH QUALITY AND SAFETY.

The reviewer would like us to separate quality from safety. We would respectfully disagree with this suggestion. We are using the WHO and Institute of Medicine definitions in which quality is divided into six dimensions. One of these dimensions of quality is safety.

2. FOR QUALITY OF CARE YOU HAVE INCLUDED A VARIETY OF MEASURES THAT REFLECT DIFFERENT ASPECTS OF QUALITY OF CARE, HOWEVER FOR SAFETY YOU FOCUS SOLELY ON MEDICAL ERRORS DESPITE SAFETY BEING A BROAD TERM WITH A VARIETY OF MEASURES EMPLOYED WITHIN THE LITERATURE. DID YOU CONSIDER WIDENING YOUR INCLUSION CRITERIA TO INCLUDE OTHER MEASURES OF SAFETY SUCH AS SSIS, UTIS, BED

SORES, SAFETY GRADE/CULTURE?

The reviewer is concerned that our search strategy was too narrow and that we focused on medical errors. We deliberately did not focus on medical errors. Our search was designed to identify papers that had measureable outcomes for the quality of care. This was done by incorporating search terms that would have identified adverse effects (ae.fs. in the search strategy) associated with treatment or interventions, including diagnostic, therapeutic, prophylactic, anesthetic, surgical, or other procedures. For example, our search would have captured content on seizures because they would have been identified as adverse effects from medical inaction or inappropriate action. The same could be said of bedsores (MeSH heading pressure ulcers) - our search would capture bedsores as an adverse effect from poor care. None of these types of outcomes were used in the studies that came up in the searches. The absence of these types of measures is not a result of our search strategy but of the existing studies.

To clarify, the text was added:

In addition, safety was identified by measures that reflected risks or harm to patients such as adverse events or medical errors.

REVIEWER 4: JAMES BEGUAH ODEI

LOOK FORWARD TO YOUR FUTURE RESEARCH IN EVALUATING BURNOUT INTERVENTIONS FOR PHYSICIANS AND CONTINUED WORK AT THE RELATIONSHIP BETWEEN DIMENSIONS OF ACCEPTABILITY-RELATED QUALITY OF MEASURES AND BURNOUT.

Thank you for your interest in our work.

VERSION 2 – REVIEW

REVIEWER	Renée Scheepers Academic Medical Center, the Netherlands
REVIEW RETURNED	02-Mar-2017

GENERAL COMMENTS	The authors have adressed the previous comments. It would provide clear overview when table 1 and table 2 of the current draft would be merged into one table. Then the study findings can be clearly interpreted in the light of the study methods used.

REVIEWER	Jef Adriaenssens Leiden University, The Netherlands Federal Knowledge Center For Healthcare (KCE), Belgium
REVIEW RETURNED	26-Feb-2017

GENERAL COMMENTS	Thank you very much for giving me the opportunity to tereview your reworked manuscript. You addressed well all the remarks made regarding the previous version. I have no further remarks. The paper is very interesting and has added value for the research field of occupational stress. To my opinion this paper is ready for publication.

VERSION 2 – AUTHOR RESPONSE

We have combined what were formally Tables 1 and 2 into one table (Table 1) as Ms. Scheepers wanted.

VERSION 3 – REVIEW

REVIEWER	Jef Adriaenssens Leiden University, The Netherlands Federal Knowledge Center for Health Care (KCE), Brussels, Belgium
REVIEW RETURNED	25-Mar-2017

GENERAL COMMENTS	Thank you very much for the revised version of your manuscript. I have no further remarks on the paper.