PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Sleepiness of day workers and watchkeepers on board at high seas: a cross-sectional study
AUTHORS	Oldenburg, Marcus; Jensen, Hans-Joachim

VERSION 1 - REVIEW

REVIEWER	Nicola Magnavita Università Cattolica del Sacro Cuore Roma e Fondazione Policlinico Gemelli IRCCS Roma
REVIEW RETURNED	17-Jan-2019

GENERAL COMMENTS	This interesting study tries to correlate the work activity of seafarers with sleepiness. The authors did not limit themselves to collecting subjective evaluations, but try to validate the reports with objective measures. A wearable device is used to estimate sleep; however, the used device is he device is probably more suitable for measuring bed rest than sleep [Lee et al. Comparison of Wearable Trackers' Ability to Estimate Sleep. Int J Environ Res Public Health. 2018 Jun 15;15(6). pii: E1265. doi: 10.3390/ijerph15061265]. The authors should discuss this point. Pupillometry has been used as a surrogate of more sophisticated tests for sleepiness. Recent studies show that this technique may be useful, even if it is not yet completely reliable as a screening test [Yamamoto K, et al. Association between pupillometric sleepiness measures and sleep latency derived by MSLT in clinically sleepy patients. Environ Health Prev Med. 2013 Sep;18(5):361-7. doi: 10.1007/s12199-013-0331-0]. Despite these limitations, the two tests may confirm the subjective statements of the seafarers. An unexpected but interesting result that should be further discussed is that even day-shift sailors suffer from drowsiness.

REVIEWER	Wessel M.A. van Leeuwen Stress Research Institute, Sweden
REVIEW RETURNED	28-Jan-2019

GENERAL COMMENTS	The authors have carried out a highly relevant area of research, namely sleepiness at the high seas using a variety of rather novel methods in order to assess sleepiness. The manuscript, however, could improve by considering my comments below. Furthermore, it is recommended to have the manuscript proof read and corrected by a native English speaker, as quite a few awkward sentences/words/formulations occur throughout the manuscript. General comments: 1. It does not become very clear as to whether the 18 voyages were at 18 different ships or at different time points on the same ship (or a limited number of ships). To rule out ship specific issues, this would need to be clarified. 2. Information on caffeine consumption of the participants is missing, but would form a very welcome addition to the methods section as well as Table 1. Caffeine intake usually is a good proxy for fatigue levels. 3. The authors use sometimes mean values, sometimes medians to describe the central tendency of variables, and sometimes it is not clear at all which one has been used (as for age for instance in Table 1). Consistency and clarity are highly appreciated. 4. The distribution of circadian preference (as can be seen in Table 1) with a great majority being morning types is not at all representing the general population. Rather than bringing up an ethnic and even racial argument in the last paragraph of the discussion, the authors are strongly encouraged to discuss more plausible reasons for this finding. 5. The authors use wearable technology only to assess crucial data on sleep. Although this may not harm the between group analyses, it may give a systematic over- or under-estimation of all the sleep related outcomes, such as sleep duration and sleep efficiency. Please discuss as to why no sleep diaries were accompanying the wearable technology, as is usually the case in this research field. 6. The result that both sleep duration was shorter and sleep efficient lower in the watchkeepers needs some discussion in the light as to how the split sleep problem was tackled by the authors. Dayworkers usually have one major sleep episode per 24 h period, whereas watchkeepers quite often have either a split sleep or one major sleep in combination with a nap. Did the authors looked at these sleeps separately? Lots of valuable information has left out this way and the authors are encouraged to include a more detailed description and analyses of the watchkeepers' sleep. As currently described on page 10, the observation of shorter duration while at the same time a lower efficiency does not make sense. The shorter a sleep gets the higher its efficiency, given it's obtained at the circadian nadir. Specific comments: 1. Page 2, line 45; "the subjective judgement of" can be deleted, as the ESS is by definition a subjective judgement. 2. Page 2, line 49; "could be objectified" should read "was observed". 3. Page 3, line 9; Please do specify the conclusion a bit more. Preventive to what? And how? 4. Page 3, line 32; I don't get the point as to why different ethnic groups should show different sleep/wake behaviour. It is

	recommended to leave the ethnic argument out of the manuscript (it even occurs later on in the manuscript) as it does not make sense. 5. Page 3, line 50; Fatigue is not the same as exhaustion and the authors are encouraged to remove “(fatigue)” here. 6. Page 4, line 20; When referring to van Leeuwen et al (2013) it might be of interest to state that this study also showed sleepiness increasing with time on watch, which is something the current manuscript seeks to confirm. 7. Page 4, line 47; “The available maritime fatigue studies…”. The accompanied reference number 13 is not a maritime study, but addresses pastry cookers and shop keepers ashore. 8. Page 4, line 55; “the relevance of” should read “the prevalence of” 9. Page 5, line 16; “no patients were involved”. How was this verified and did what kind of diseases did it include? What about, for instance, sleep disorders? 10. Page 5, line 23-25; once again, the continent of origin seems irrelevant when studying a purely biological phenomenon as sleep(iness). 11. Page 5, line 28; Mentioning “watch keepers”, “shift workers”, “day shift workers” makes it confusing. Isn’t the comparison mainly between day workers and shift workers? 12. Page 5, line 31; Instead of “chronotype” the authors may wish to write “circadian preference”, since only a subjective instrument is used and no melatonin and/or core body temperature to objectify chronotype. 13. Page 7, line 9/10; sleep seem to have only been assessed during a 72-hour period. This is a very brief period for a phenomenon that can vary greatly on a daily basis. It is recommended to discuss the rationale of this 72-hour period. Also in the light of a medical staff that was present all the time, carrying out tests which could have disturbed the seafarers greatly in their daily routines. 14. Page 7, line 21: “according to its validation requirements”. A reference would not do any harm here. 15. Page 7, line 27-29; Expressing sleep efficiency as the time spent asleep while lying down will result in a systematic under estimation of this variable. Most studies use “lights out” which tends to occur well after “lying down” as the start time point. This is why actigraphic devices tend to have a button for the participants to press and also why sleep diaries usually accompany the usage of wearables in field studies like this. Please discuss this limitation in the manuscript. 16. Page 7, line 35/46; “the armband has proved to be superior in comparison to other activity monitors”. Please do add the highly important point that this comparison only involved “commercially available” monitors. 17. Page 7, line 43; a reference as to the device “fit-for-duty” would be highly appreciated. 18. Page 8, line 21; “und” should read “and”. 19. Page 8, line 33/34; The Epworth Sleepiness Scale (ESS) does assess daytime sleepiness, not chronic fatigue. Please correct this. 20. Page 8, line 37/38; the rMEQ is a very short self-assessment tool for circadian preference. As such, it does not “identify the daily peak of alertness”. Alertness is regulated by several processes, the main ones being a circadian and a homeostatic process.

	21. Page 8, line 39; please replace “chronotypes” by “circadian preferences”. 22. Page 8, line 47/48; How was “the normal distribution investigated”? Which test? And why did the authors use non-parametric tests only? Does that imply that none of the variables displayed a normal distribution? 23. Page 10, line 9; “sleep…lasted for 5.6 hours”. Was this one main sleep episode or a more split sleep over 2 different off-watch periods? 24. Page 10, line 29-38; did the before-after comparison somehow take into account that work shifts of watchkeepers are much shorter (4h for instance for those working 4/8) than those of day workers? 25. Page 10, line 59-61; Severe sleepiness was most often displayed at the end of the 04-08 watch and considerably less at the end of the 00-04 watch. This once again raises doubts about the rMEQ distribution in Table 1. Morning types should feel more alert at 08:00 and certainly not at 04:00 after having worked 00:00 to 04:00. I suggest that the authors revisit their rMEQ results and check if the distribution was not the other way around compared to what arises from Table 1. 26. Page 10, line 53; Did the pupillometric examinations really take 30 minutes each? This would mean a 1 hour working time reduction on a 4-hour watch. Hence a 25% working time reduction. Please discuss in greater detail this examination and to what extend it interfered with their working time. 27. Table 3 (page 11/12) has a strong need for information on working times. 28. In Table 3 the authors describe rPUI as being assessed “after the shift” while in the text they state it took place “during the first and last 30 minutes of a shift” (page 10, line 53). Which statement is true? 29. Page 13, line 46/47; it is first here that the authors take up the problem of watch keepers sleeping more than once during a 24h period. It is strongly recommended to mention this from the beginning and to re-write the results section taking this into account. 30. Page 14, line 13; “war” should read “was”. 31. Page 14, line 27; the reference to van Leeuwen et al as if they had shown that most shipping disasters take place during the circadian nadir should be removed. Instead, please do refer to the Bridge Watchkeeping Safety Study by the Marine Accident Investigation Branch (MAIB) where the relevant figures originate from. 32. Page 14, line 40/41. “During the 11-min pupillary examination”. Earlier on, this examination was described as taking 30 minutes (page 10, line 53). Which one is true? 33. Page 14, line 47. “The period already spent on the vessel at the time of the examination seems to influence sleepiness on duty” is a rather strong statement that is not supported by the sleep/wake regulation literature nor by alertness regulation models. It is also questionable as to how the current design would justify such a conclusion. If keeping the statement, please elaborate further and in much greater detail. 34. It is recommended to delete the last paragraph of the discussion dealing about “Africans and fair-skinned people”. It is offensive, irrelevant and does not make any sense. The very last sentence, however, starting with “future studies should explore….” is a very good one and an idea that I fully support. 35. Page 16, line 6; “relevance” should read “prevalence”.

	36. Page 16, line 14; Reference number 33 is a review article on fishermen fatigue. Although a highly interesting piece of work, I do not see how this would support the sentence preceding it.

VERSION 1 – AUTHOR RESPONSE

Answers to the comments by “Reviewer 1”

“This interesting study tries to correlate the work activity of seafarers with sleepiness. The authors did not limit themselves to collecting subjective evaluations, but try to validate the reports with objective measures. A wearable device is used to estimate sleep; however, the used device is probably more suitable for measuring bed rest than sleep. The authors should discuss this point”: Yes, we agree that the armband monitor used has a limited informative value about sleep architecture, which is normally measured in sleep laboratories ashore, e.g. using polysomnographic techniques [Lee et al. Comparison of Wearable Trackers' Ability to Estimate Sleep. Int J Environ Res Public Health. 2018 Jun 15;15(6). pii: E1265. doi: 10.3390/ijerph15061265]. In maritime field studies, however, the use of such extensive examinations (only one measurement per night) does not appear to be very suitable on board. In comparison to other commercially available activity monitors, the armband has proved to be superior for determining sleeping time (Gruwez et al. 2017).

The frequently used sleep diary for sleep assessment is also only a subjective procedure and requires a survey period of at least 2 weeks, particularly as the shipboard measurements in the present study only took place during the sea passage. In the authors’ experience, the documentation of working und sleeping times on the high seas is often influenced by an underlying social desirability, which can lead to a considerable underestimation of the sleep deficiency (see Allen et al. 2007). Thus, we consider it important that an examiner who stays on board throughout this long period of time checks the seafarers’ statements. Therefore, an aim of this study was to test alternative devices to objectively assess the seafarers’ activity profiles and their sleep efficiency.
“Pupillometry has been used as a surrogate of more sophisticated tests for sleepiness. Recent studies show that this technique may be useful, even if it is not yet completely reliable as a screening test”: We agree that pupillometry has yet not been established as a “completely reliable screening test for sleepiness” [Yamamoto K, et al. Association between pupillometric sleepiness measures and sleep latency derived fromMSLT in clinically sleepy patients. Environ Health Prev Med. 2013 Sep;18(5):361-7. doi: 10.1007/s12199-013-0331-0]. Particularly sleep latency or sleep architecture are the domains of extensive examinations in sleep laboratories ashore and were not the focus of the present maritime field study. A recent study suggested the Pupillographic Sleepiness Test as a reliable measurement for detecting drowsiness-related impairment (Maccoraet al. 2018).

Despite the mentioned limitations, the armband monitor and pupillometry confirmed the seafarers’ subjective statements. Further studies are recommended to evaluate the validity of these devices for their use in field settings, and to check their suitability on board and their acceptance by the seafarers on the high seas.

The latter two paragraphs have now been included in the revised manuscript as limitations of the study.

“An unexpected but interesting result that should be further discussed is that even day-shift sailors suffer from drowsiness”: During the sea voyage, all seafarers are subjected to ship-specific psychophysical impacts. On the one hand, the crew is permanently exposed to ship movements, noise and vibration; on the other hand, the job profile is often less demanding in this voyage episode, which sometimes leads to monotony. These effects can explain why many seafarers, including day-shift workers, suffer from sleepiness on the high seas. This point has now been added in the Discussion.

Answers to the comments by “Reviewer 2”

“The authors have carried out a highly relevant area of research, namely sleepiness at the high seas using a variety of rather novel methods in order to assess sleepiness. However, it is recommended to have the manuscript proof read and corrected by a native English speaker, as quite a few awkward sentences/words/formulations occur throughout the manuscript”: Now a professional translation service with an English native speaker has proofread the whole manuscript (Norah Schmidt, elder’s - translations and more, Hamburg).

“It does not become very clear as to whether the 18 voyages were at 18 different ships or at different time points on the same ship (or a limited number of ships). To rule out ship specific issues, this would need to be clarified”: Now, we have clarified that the crews of 18 different ships were examined.

“Information on caffeine consumption of the participants is missing, but would form a very welcome addition to the methods section as well as Table 1. Caffeine intake usually is a good proxy for fatigue levels”: In this study, we have not asked about the amount of coffee consumed per day, but the coffee intake within the 4 hours before our pupillometric examination was requested. 64% of the participating seafarers stated that they had consumed coffee during that time, independent of their occupational group. This information is now included in the Results.

“The authors use sometimes mean values, sometimes medians to describe the central tendency of variables, and sometimes it is not clear at all which one has been used (as for age for instance in Table 1). Consistency and clarity are highly appreciated”: The information on the kind of descriptive statistic used (mean or median depending on whether the data were normally distributed or not) has been added.

“The distribution of circadian preference (as can be seen in Table 1) with a great majority being morning types is not at all representing the general population. Rather than bringing up an ethnic and even racial argument in the last paragraph of the discussion, the authors are strongly encouraged to discuss more plausible reasons for this finding”: As recommended later by the reviewer, these paragraphs about ethnic differences have been deleted.

“Please discuss as to why no sleep diaries were accompanying the wearable technology, as is usually the case in this research field”: see answer to the 1st question asked by Reviewer 1.

“The result that both sleep duration was shorter and sleep efficient lower in the watchkeepers needs some discussion in the light as to how the split sleep problem was tackled by the authors. Dayworkersusually have one major sleep episode per 24 h period, whereas watchkeepers quite often have either a split sleep or one major sleep in combination with a nap. Did the authors looked at these sleeps separately?”: No, this was not a focus of the present study. This point is now included as a limitation of the study.

“As currently described on page 10, the observation of shorter duration while at the same time a lower efficiency does not make sense. The shorter a sleep gets the higher its efficiency, given it’s obtained at the circadian nadir”: In this study, not only was the duration of sleep shorter among watchkeepers, but their sleep efficiency was also lower. This means that in spite of their reduced opportunities for sleep, the effective sleeping time during the lying time was also shorter. It is assumed that some seafarers have a problem falling asleep after a stressful working day (with scarcely any opportunities for sleep); this results in decreased sleep efficiency. Split sleep among watchkeepers can also not be excluded as the cause of this low sleep efficiency.

“Page 2, line 45; “the subjective judgement of” can be deleted, as the ESS is by definition a subjective judgement”: Done.
“Page 2, line 49; “could be objectified” should read “was observed””: Done.
“Page 3, line 9; Please do specify the conclusion a bit more. Preventive to what? And how?”: The conclusion has now been described more precisely as follows: “Preventive measures need to be taken by the shipping industry to counteract fatigue (e.g. through compliance with the obligatory rest and sleep times)”. Due to the restrictions of the maximum number of words in the Abstract no further examples of preventive measures could be added to this section.

“Page 3, line 32; I don’t get the point as to why different ethnic groups should show different sleep/wake behaviour. It is recommended to leave the ethnic argument out of the manuscript (it even occurs later on in the manuscript) as it does not make sense”: As recommended, this argument has been deleted.

“Page 3, line 50; Fatigue is not the same as exhaustion and the authors are encouraged to remove “(fatigue)” here”: Done.

“Page 4, line 20; When referring to van Leeuwen et al (2013) it might be of interest to state that this study also showed sleepiness increasing with time on watch, which is something the current manuscript seeks to confirm”: Done.

“Page 4, line 47; “The available maritime fatigue studies…”. The accompanied reference number 13 is not a maritime study, but addresses pastry cookers and shop keepers ashore”: The reference number 13 has been removed.

“Page 4, line 55; “the relevance of” should read “the prevalence of””: Done.

“Page 5, line 16; “no patients were involved”. How was this verified and did what kind of diseases did it include? What about, for instance, sleep disorders?”: Now, as requested by the editor, a paragraph about the Patient and Public Involvement is presented in the Methods.

“Page 5, line 23-25; once again, the continent of origin seems irrelevant when studying a purely biological phenomenon as sleep(iness)”: The differentiation concerning the continent of origin has been removed.

“Page 5, line 28; Mentioning “watch keepers”, “shift workers”, “day shift workers” makes it confusing. Isn’t the comparison mainly between day workers and shift workers?”: In the revised manuscript a distinction has been made between day workers and watchkeepers.

“Page 5, line 31; Instead of “chronotype” the authors may wish to write “circadian preference”, since only a subjective instrument is used and no melatonin and/or core body temperature to objectify chronotype”: Done.

“Page 7, line 9/10; sleep seem to have only been assessed during a 72-hour period. This is a very brief period for a phenomenon that can vary greatly on a daily basis. It is recommended to discuss the rationale of this 72-hour period. Also in the light of a medical staff that was present all the time, carrying out tests which could have disturbed the seafarers greatly in their daily routines”: Due to known variations of sleep quality on a daily basis, we decided to evaluate an observation time of at least 72 hours during the sea passage. As the sea passage only represents one part of the voyage, this period normally requires the presence of one examiner on board for at least 10 days.

To assess the seafarer’s sleep, only those devices were selected which did not considerably disturb the crew’s daily routines (low weight, no cable connection, easy use). While wearing the armband monitor, the seafarers could easily operate the device for themselves without the presence of a shipboard examiner. The pupillometric cross-shift examination took place 15 minutes before the respective shift started and directly after it ended so that that shift was neither shortened nor disturbed through this examination. Thus, it is not likely that the seafarers were distinctly disturbed by the medical staff on board. These arguments are included in the revised manuscript.

“Page 7, line 21: “according to its validation requirements”. A reference would not do any harm here”: A suitable reference for the validation of the SenseWear armband has been added: Lopez GA, Brønd JC, Andersen LB, Dencker M, Arvidsson D. Validation of SenseWear armband in children, adolescents, and adults. Scand J Med Sci Sports. 2018;28(2):487-495. doi: 10.1111/sms.12920. Epub 2017 Jun 28.

“Page 7, line 27-29; Expressing sleep efficiency as the time spent asleep while lying down will result in a systematic under estimation of this variable. Most studies use “lights out” which tends to occur well after “lying down” as the start time point. This is why actigraphic devices tend to have a button for the participants to press and also why sleep diaries usually accompany the usage of wearables in

field studies like this. Please discuss this limitation in the manuscript": These aspects are now presented as a limitation in the Discussion.

"Page 7, line 35/46; "the armband has proved to be superior in comparison to other activity monitors". Please do add the highly important point that this comparison only involved "commercially available" monitors": Done.

"Page 7, line 43; a reference as to the device "fit-for-duty" would be highly appreciated": A reference has been included: Różanowski K, Bernat M, Kamińska A. Estimation of operators' fatigue using optical methods for determination of pupil activity. Int J Occup Med Environ Health 2015;28(2):263-81. doi: 10.13075/ijomeh.1896.00274.

"Page 8, line 21; "und" should read "and"": Done.

"Page 8, line 33/34; The Epworth Sleepiness Scale (ESS) does assess daytime sleepiness, not chronic fatigue. Please correct this": Done.

"Page 8, line 37/38; the rMEQ is a very short self-assessment tool for circadian preference. As such, it does not "identify the daily peak of alertness". Alertness is regulated by several processes, the main ones being a circadian and a homeostatic process. Page 8, line 39; please replace "chronotypes" by "circadian preferences": Done.

"Page 8, line 47/48; How was "the normal distribution investigated"? Which test? And why did the authors use non-parametric tests only? Does that imply that none of the variables displayed a normal distribution?": The Shapiro-Wilk test was used to test for a normal distribution of data. This information is now added in the revised manuscript.
The means of two unrelated groups were only statistically compared in Table 2 and, indeed, both dependent parameters (stay on board at the time of examination and ESS score) were not normally distributed so that the Mann Whithney-U test was applied.

"Page 10, line 9; "sleep…lasted for 5.6 hours". Was this one main sleep episode or a more split sleep over 2 different off-watch periods?": This was the cumulative sleeping duration (including split sleep episodes) per 24-h period.

"Page 10, line 29-38; did the before-after comparison somehow take into account that work shifts of watchkeepers are much shorter (4h for instance for those working 4/8) than those of day workers?": The difference in the length of working time between day workers and watchkeepers has now been pointed out in several text passages. Now the Discussion highlights the fact that the watchkeepers' higher value for SSS and,by trend, for PUI after the examined shift is remarkable as they had worked a considerably shorter time than the day workers. Thus, these differences are surely underestimated.

"Page 10, line 59-61; Severe sleepiness was most often displayed at the end of the 04-08 watch and considerably less at the end of the 00-04 watch. This once again raises doubts about the rMEQ distribution in Table 1. Morning types should feel more alert at 08:00 and certainly not at 04:00 after having worked 00:00 to 04:00. I suggest that the authors revisit their rMEQ results and check if the distribution was not the other way around compared to what arises from Table 1": The data were checked and are now presented more clearly.

"Page 10, line 53; Did the pupillometric examinations really take 30 minutes each? This would mean a 1 hour working time reduction on a 4-hour watch. Hence a 25% working time reduction. Please discuss in greater detail this examination and to what extend it interfered with their working time": Done, see above.

"Table 3 (page 11/12) has a strong need for information on working times": Done. The differences in the lengths of the examined shifts between the two occupational groups are indicated and discussed in several passages.

“In Table 3 the authors describe rPUI as being assessed “after the shift” while in the text they state it took place “during the first and last 30 minutes of a shift” (page 10, line 53). Which statement is true?”:Corrected, see above.

“Page 13, line 46/47; it is first here that the authors take up the problem of watch keepers sleeping more than once during a 24h period. It is strongly recommended to mention this from the beginning and to re-write the results section taking this into account”: As watchkeepers have two work units per day - each of them about 4 h - split sleeping time is often observed in this occupational group. The watchkeepers were examined during a randomly selected shift period with the aim of achieving an equal representation of these periods. This aspect is now included in the Methods and Results.
“Page 14, line 13; “war” should read “was””: Done.

“Page 14, line 27; the reference to van Leeuwen et al as if they had shown that most shipping disasters take place during the circadian nadir should be removed. Instead, please do refer to the Bridge Watchkeeping Safety Study by the Marine Accident Investigation Branch (MAIB) where the relevant figures originate from”: Done.

“Page 14, line 40/41. “During the 11-min pupillary examination”. Earlier on, this examination was described as taking 30 minutes (page 10, line 53). Which one is true?”: Done, see above.

“Page 14, line 47. “The period already spent on the vessel at the time of the examination seems to influence sleepiness on duty” is a rather strong statement that is not supported by the sleep/wake regulation literature nor by alertness regulation models. It is also questionable as to how the current design would justify such a conclusion. If keeping the statement, please elaborate further and in much greater detail”: This text passage addresses the interesting result observed, namely that the duration already spent on the vessel at the time othe examination correlated with the PUI. This finding could indicate a cumulative effect on the seafarers’ sleepiness. We agree that this is only a hypothesis that needs to be confirmed in further studies.

“It is recommended to delete the last paragraph of the discussion dealing about “Africans and fair-skinned people”. It is offensive, irrelevant and does not make any sense. The very last sentence, however, starting with “future studies should explore….” is a very good one and an idea that I fully support”: As recommended, this paragraph has been deleted.

“Page 16, line 6; “relevance” should read “prevalence””: Done.

“Page 16, line 14; Reference number 33 is a review article on fishermen fatigue. Although a highly interesting piece of work, I do not see how this would support the sentence preceding it”: This reference has been removed now.

VERSION 2 – REVIEW

REVIEWER	Nicola Magnavita Università Cattolica del sacro Cuore, Roma, Italy
REVIEW RETURNED	08-Mar-2019

GENERAL COMMENTS	It is an interesting, well designed and conducted study of the drowsiness of seafarers. Some minor observations. Line211: The Stanford Sleepiness Scale, please give a reference and an explanation of the questionnaire. Also the other questionnaires (ESS and MEQ) should be briefly described. With reference to the methods used, authors should probably recognize that the wearable device may overestimate sleep, as it is more suitable for measuring bed rest [Lee et al. Comparison of

	Wearable Trackers' Ability to Estimate Sleep. Int J Environ Res Public Health. 2018 Jun 15;15(6). pii: E1265. doi: 10.3390/ijerph15061265], and that pupillometry is not yet completely reliable as a screening test [Yamamoto K, et al. Association between pupillometric sleepiness measures and sleep latency derived by MSLT in clinically sleepy patients. Environ Health Prev Med. 2013 Sep;18(5):361-7. doi: 10.1007/s12199-013-0331-0].

REVIEWER	Wessel M.A. van Leeuwen Stress Research Institute Stockholm University SWEDEN
REVIEW RETURNED	15-Mar-2019

GENERAL COMMENTS	The current revision has improved the manuscript considerably, and only a few minor points remain to be adjusted/clarified. Page and line numbers below refer to the manuscript version with the marked changes. 1. (page 3, line 52/53). "e.g. through compliance with the obligatory rest and sleep times". This statement suggests that the shipping industry currently violates these regulations, which is, at least on paper, not the case. Caution is needed with statements like this, and a revision is highly recommended. 2. (page 3, line 58). "provides" should read as "shows" 3. (page 3, line 66). "and ethnic" should be removed. Please refer to my review of the earlier version of the manuscript for more details. 4. (page 6, line 122). "..18 sea voyages on 18 different container ships". Earlier it is stated that 10 shipping companies participated. It is recommended to mention here how these 18 ships were divided over these 10 companies. 5. (page 6, line 148-151). "all seafarers..... were encouraged to participate". Please specify what percentage of those encouraged actually did participate in the end (response rate). 6. (page 7, Table 1). In response to my earlier review, the authors have still not discussed a plausible reason as to why such a large portion of participants were morning types. A skewed distribution towards morningness is very unusual and needs some discussion. 7. The statement of the authors in their rebuttal letter that "sleep diaries require a survey period of 2 weeks" is not correct. Sleep diaries can be used and are being used for periods from 2 days up to several months. A clear reason as to why no diaries are used to accompany the armband is therefore lacking and needs to be included in the manuscript. It is a clear and present weakness to use wearables without a diary. The vast majority of studies implementing for instance actigraphy also use diaries to check the start and end times of sleep. 8. (page 17, line 421). "analysis" should read as "study".

	9. (page 21, reference number 4). The last author is "van Leeuwen WMA" and not "van Leeuwen MW". 10. The revision made by the authors in the method section concerning the timing and length of the pupillometric examination is rather puzzling to me. In the first version of the manuscript, the authors wrote that the examination took 30 minutes and was carried out during their work shift and in the revision they write that it only took 15 minutes and now it was carried out before and after the work shift. How could such a considerable methodological error end up in the first draft (assuming that what is written in the revision is how the study was really carried out)?

VERSION 2 – AUTHOR RESPONSE

Answers to the comments by “Reviewer 1”

“It is an interesting, well designed and conducted study of the drowsiness of seafarers. Some minor observations.
Line211: The Stanford Sleepiness Scale, please give a reference and an explanation of the questionnaire. Also the other questionnaires (ESS and MEQ) should be briefly described”: in the revised manuscript, the descriptions of these questionnaires, including their respective references, have been added.

“With reference to the methods used, authors should probably recognize that the wearable device may overestimate sleep, as it is more suitable for measuring bed rest [Lee et al. Comparison of Wearable Trackers' Ability to Estimate Sleep. Int J Environ Res Public Health. 2018 Jun 15;15(6). pii: E1265. doi: 10.3390/ijerph15061265], and that pupillometry is not yet completely reliable as a screening test [Yamamoto K, et al. Association between pupillometric sleepiness measures and sleep latency derived by MSLT in clinically sleepy patients. Environ Health Prev Med. 2013 Sep;18(5):361-7. doi: 10.1007/s12199-013-0331-0]”: both aspects and the respective references are now included.

Answers to the comments by “Reviewer 2”
“The current revision has improved the manuscript considerably, and only a few minor points remain to be adjusted/ clarified.
1. (page 3, line 52/53). "e.g. through compliance with the obligatory rest and sleep times". This statement suggests that the shipping industry currently violates these regulations, which is, at least on paper, not the case”: this sentence has been revised and the assumption of possible violations of regulations by the shipping industry has been deleted.

“2. (page 3, line 58). "provides" should read as "shows"”: done.

“3. (page 3, line 66). "and ethnic" should be removed. Please refer to my review of the earlier version of the manuscript for more details”: done.

“4. (page 6, line 122). "..18 sea voyages on 18 different container ships". Earlier it is stated that 10 shipping companies participated. It is recommended to mention here how these 18 ships were divided over these 10 companies”: the proportion of ships was four times one, four times two and twice three vessels per shipping company.

“5. (page 6, line 148-151). "all seafarers..... were encouraged to participate". Please specify what

percentage of those encouraged actually did participate in the end (response rate)": we have stated the response and the participation rate in the manuscript: "206 out of 225 seafarers took part in the study (response rate 91.6%). Only the results of those 198 seafarers were included who could be interviewed in a cross-shift design (participation rate 88.0%)".

"6. (page 7, Table 1). In response to my earlier review, the authors have still not discussed a plausible reason as to why such a large portion of participants were morning types. A skewed distribution towards morningness is very unusual and needs some discussion": according to the results of the Morning- Evening-Questionnaire in our study, the morning type was indeed overrepresented in our study group. Due to the fact that their work shifts on board often begin early in the morning over several months in a stretch, many seafarers are surely adapted to this daily rhythm and subjectively feel particularly fit in the early morning hours. This is a hypothesis for the skewed distribution towards the morning type in our study that needs confirmation in further field studies on board.

"7. The statement of the authors in their rebuttal letter that "sleep diaries require a survey period of 2 weeks" is not correct. Sleep diaries can be used and are being used for periods from 2 days up to several months. A clear reason as to why no diaries are used to accompany the armband is therefore lacking and needs to be included in the manuscript. It is a clear and present weakness to use wearables without a diary. The vast majority of studies implementing for instance actigraphy also use diaries to check the start and end times of sleep": according to the authors' experience, sleep diaries covering a survey period of at least 2 weeks are recommended to assess a representative sleeping time and to minimize the risk of false assessment due to a short-time observation of perhaps only 2 days. However, they allow the checking of the start and end times of sleep so that in further studies the armband monitor examination should be accompanied by the use of sleep diaries. This aspect has been included in the revised manuscript.

"8. (page 17, line 421). "analysis" should read as "study": done.

"9. (page 21, reference number 4). The last author is "van Leeuwen WMA" and not "van Leeuwen MW"": done.

"10. The revision made by the authors in the method section concerning the timing and length of the pupillometric examination is rather puzzling to me. In the first version of the manuscript, the authors wrote that the examination took 30 minutes and was carried out during their work shift and in the revision they write that it only took 15 minutes and now it was carried out before and after the work shift ": this was indeed expressed somewhat misleadingly in the first version; each of the two pupillometric examinations (before and after one work shift) took 15 min; that means, in total 30 min examination time per seafarer.

VERSION 3 – REVIEW

REVIEWER	Wessel M.A. van Leeuwen Stress Research Institute Stockholm University Sweden
REVIEW RETURNED	06-May-2019

GENERAL COMMENTS	The manuscript has improved again since the latest revision. Only a couple of final minor points remain. If these points are dealt with, the manuscript would be suitable for publication.

	1. (page 5, line 141/142). This is a rather unclear way of stating the number of ships per participating company. Rather than writing 4*1, 4*2, 2*3, the authors may wish to consider writing more in terms like "1 ship of companies A, B, C, and D participated, 2 ships of companies E, F, G, and H participated etc. etc," 2. (page 9, line 229/230). After the normality testing, it is stated what test was used for those variables not normally distributed. However, information is lacking as to how variables were analysed that displayed a normal distribution.

VERSION 3 – AUTHOR RESPONSE

Answers to the comments by “Reviewer 2”

“The manuscript has improved again since the latest revision. Only a couple of final minor points remain. If these points are dealt with, the manuscript would be suitable for publication.
1. (page 5, line 141/142). This is a rather unclear way of stating the number of ships per participating company. Rather than writing 4*1, 4*2, 2*3, the authors may wish to consider writing more in terms like "1 ship of companies A, B, C, and D participated, 2 ships of companies E, F, G, and H participated etc. etc,": done as recommended.

“2. (page 9, line 229/230). After the normality testing, it is stated what test was used for those variables not normally distributed. However, information is lacking as to how variables were analysed that displayed a normal distribution”: in the revised manuscript the information has been added that the T-test was applied in the case of normal distribution.