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)	Using sensor-fusion and machine-learning algorithms to assess acute pain in nonverbal infants: a study protocol
AUTHORS	Roué, Jean-Michel; Morag, Iris; Haddad, Wassim H.; Gholami, Behnood; Anand, Kanwaljeet J. S.

VERSION 1 – REVIEW

REVIEWER	Raul Fernandez-Rojas University of Canberra, Australia
REVIEW RETURNED	08-Jun-2020

GENERAL COMMENTS	Overall, this is a clear, concise, and well-written protocol. However, I have some specific comments: 1. What are the "clinically-indicated procedures" (apart from heel stick) that will be used as noxious stimuli? 2. How the sample size in the training dataset was calculated? 3. There is no mention about any testing or validation procedure for the machine learning models. It is assumed that 20 subjects will be used for testing/validation. 4. What kind of event recorder is used to synchronise all the sensors? 5. It is not clear how sensor fusion will be used. It seems that feature fusion will be used instead (refer to page 16, line 6). In addition, there are a several useful features that can be obtained from each sensor modality, so if the sensors are fused immediately after cleaning, the sensitivity from each modality might be lost. 6. Is the audio recording used for analysis? or is there any reason to obtain audio too?

REVIEWER	Saab, Carl Brown University
REVIEW RETURNED	24-Jun-2020

GENERAL COMMENTS	Timely study protocol that is overall comprehensive, but missing a few critical elements. For e.g., the ground truth for pain is subjective, as the authors acknowledge, so the training of the algorithms will be extremely hard. Given low sample size, this is a critical caveat (estimation of the sample size is not adequate). This is more relevant because many confounding variables are not controlled for properly (age, gender, and type of painful interventions). Also, pain is multi-dimensional (sensory, cognitive, emotional, etc.) It is also acute, chronic, evoked and spontaneous. Which dimension / type is being measured? Most of the sensors chosen, especially EEG, are not hypothesis-driven, so it's hard to judge the merit of the proposal (what changes are expected and in

	what direction?) Movement artifacts will be the most significant challenge and a source for major noise. More detailed plans for remedy are needed.

REVIEWER	Rebecca Pillai Riddell York University, Canada
REVIEW RETURNED	26-Jul-2020

GENERAL COMMENTS	Using Sensor-fusion and machine-learning algorithms to assess acute pain in nonverbal infants: A study protocol (www.Clinicaltrials.gov #NCT03330496) Roué et al. **Preliminary Analysis of Protocol** ● Protocol papers should report planned or ongoing studies. The dates of the study should be included in the manuscript. **Dates need to be included for the study period** ● While some baseline data can be presented, there should be no results or conclusions present in the study protocol. **No Results presented** ● For studies that are ongoing, it is generally the case that very few changes can be made to the methodology. As such, requests for revisions are generally clarifications for the rationale or details relating to the methods. If there is a major flaw in the study that would prevent a sound interpretation of the data, we would expect the study protocol to be rejected. **Summary** This is a timely and interesting protocol to use sensor-fusion and machine-learning algorithms to assess acute pain in 60 preterm and term newborns (4 age groups from 34 weeks GA to full terms aged 6 months) undergoing painful procedures. The team plans to record five separate physiological measures (EMG, ECG, HR, HRV, SpO2 and EEG). It is a feasibility study setting out to finalize the approach to increase the accuracy of pain assessment and allow continuous pain monitoring at the bedside. **Strengths of the Protocol** • Continuous pain assessment monitoring at bedside is a fundamentally important and critical objective within the NICU and I am convinced that automated bedside multimodal assessment is the most promising methodology to reduce unnecessary pain and utilize pain medication more precisely

	• The team is to be commended in doing a feasibility study with a specific goal of examining recording artifacts to better understand data corrections and sensor variability to have an appropriate estimate of power for the larger study. **Challenges of the Protocol** • This is a feasibility study with strong multinational team. Our instructions were to in essence to determine if "there is a major flaw in the study that would prevent a sound interpretation of the data". As I am working in this area, I know there are very few studies being done out there, so the idea of the fatal flaw was not the focus of my review for a feasibility study. • There are a few areas where clarification/modification could be considered by the team to potentially increase the value of the study. o Are there funding sources or conflicts of interest to acknowledge? o Page 7, line 35: "We will record pain signals…". I would debate this language as none of the physiological measures being recorded are pain-*specific* signals. While the EEG has shown specificity it is within a limited window (less than 1 second). This brings up the larger point with language around "continuous monitoring of pain". Without specificity the multi-modal assessment, outside of a very circumscribed period of time (i.e. 1 sec post acute procedure), they are not continuously measuring pain. The language of 'estimating pain' used on page 8 seems more defensible o Has EMG been validated for you in infants in the acute pain context? This would seem to be very invasive for infants. As they note EMG records signals much more coarsely. Thus, the placement on the cheek and forehead may not map anatomically onto BB, ES and NLF in pre-terms. We are doing work now suggesting that the cluster of facial actions are different in preterm and full term infants in the post heel-lance. o Using a power analysis to justify 40 infants for a training dataset ignores the idea that they don't have just one group they are spanning across a large range of infant development (cortically, motorically, psychosocially). One would assume the logic would be 40 infants per group- thus a feasibility study of 160 not 60. From my understanding of machine learning principles, in any scenario, 40 infants for a training set would be challenging. o As HR and HRV are very closely related, there may be challenges in the assumptions of their modelling

	due to the dependency. They may want to consider one or the other. ○ There is not detail on the actual outcome variable of the EEG. What will they be measuring? In essence, the team reports the tool (Brain AMP cap with 32 active electrodes), but what will go into the machine learning algorithm? Moving away from a single electrode at vertex (which is an important direction) is great, but it is not clear how the immense data from 32-active electrodes will be used. ○ Page 15, line 45: What validated pain scale will be used to identify the pain state of the infant ○ The intersection of gestational age, postnatal age and days in hospital does not seem to properly be described as a factor in the proposal. Imagine a 1-month old (postnatal age) who has been born preterm vs full-term, juxtaposed with how many skin-breaking procedures that have done juxtaposed with varying lengths of hospital stay. Machine learning for discrimination without accounting for this type of variability using small sample size may cause challenge in understanding the data for a larger study.

REVIEWER	Gilmer Valdes University of California San Francisco, Radiation Oncology
REVIEW RETURNED	07-Sep-2020

GENERAL COMMENTS	Objective pain assessment in non-verbal populations is clinically challenging and its misdiagnosed can have long term consequences. The authors propose to use Machine Learning to provide a “patient-centered, context-dependent, observer independent, and objective pain measure.” To measure pain, the authors use Facial Electormyography (EMG) and electrocardiography (ECG) as surrogates. Statistical algorithms are proposed to find correlation between these signals and pain. The protocol is well written and it is very thorough. The authors acknowledge their limitations and make efforts to address them. For instance, inherent bias and noise from each sensor or channel the authors use a “calibration” period where they acquire images to establish a baseline for these multiple sensors. There is a major limitation to the study that cannot be addressed though. “The clinical staff at the bedside will identify the pain state of each neonate using validated pain scales and record the timing of pain-inducing clinically-indicated procedures such as a heel stick.” To me this is the main limitation of the study and what makes this problem particularly challenging. We still rely on a “subjective evaluation” of pain to find correlations between the sensors and the perceived pain. Please find below, general comments: 1) The power calculation resulting in 40 patients needed is not well explained and it is hard for this reviewer to establish its correctness.

	2) Since, the ML models will also be predicting subjective pain, the authors should examine as possible score to evaluate in the validation set a weighted majority vote between the algorithm predictions based on the sensors and the clinical classification of pain. 3) Efforts should be made to validate the algorithms on independent cohorts, possible from different hospital systems. 4) If the authors are planning to provide interventions to manage the pain, they should consider causal statistical frameworks in their statistical analysis. This is lacking.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Raul Fernandez-Rojas
Institution and Country: University of Canberra, Australia
Competing interests: None declared

Please leave your comments for the authors below
Overall, this is a clear, concise, and well-written protocol.

Response: Thank you for your comments.

However, I have some specific comments:
1. What are the "clinically-indicated procedures" (apart from heel stick) that will be used as noxious stimuli?

Response: All skin-breaking procedures required for routine clinical care will be considered. We will mainly enroll patients requiring heel sticks for blood sampling, although infants requiring subcutaneous or intramuscular injections (vaccine, drug shot) will also be considered for enrollment. We revised the manuscript accordingly (page 10 of the revised manuscript, Marked copy).

2. How the sample size in the training dataset was calculated?

Response: Sample size calculations are described on Page 9 of the revised manuscript. Briefly, we first reviewed the variability of physiological responses to acute pain in the published literature. For sample size calculations, we assumed an α-error =0.05, 1–β error =0.8, and a mean:SD ratio of 2:1 and applied these parameters to a very general but elegant sample-size calculator developed by Russel Lenth (University of Iowa, http://homepage.divms.uiowa.edu/~rlenth/Power/index.html).

3. There is no mention about any testing or validation procedure for the machine learning models. It is assumed that 20 subjects will be used for testing/validation.

Response: Data from patients will be divided to a training set and a test set. The training set is used for model training and optimization of model parameters. A leave-one-patient-out cross validation technique will be used, where the machine learning classifier is trained on data from all but one patient and the performance of the classifier is assessed on the remaining patient. Once the

appropriate machine learning classifier and its associated parameters are selected using the training set and the associated cross validation procedure, the performance of the machine learning classifier will be assessed on the test set. This discussion was added to the manuscript (page 15, Marked copy).

4. What kind of event recorder is used to synchronise all the sensors?

Response: All sensors will be monitored and displayed on the same laptop. We will use the Brain Vision software to display and record EEG, EMG and skin conductance responses and the MediCollector software for ECG and SpO2. The recording time of the 2 softwares will be synchronized based on the laptop digital clock. The event marker will be triggered by the researcher using a dedicated function of the Brain Vision software to time-lock and record times of noxious and non-noxious events for all sensors (Brain Vision and MediCollector).

5. It is not clear how sensor fusion will be used. It seems that feature fusion will be used instead (refer to page 16, line 6). In addition, there are a several useful features that can be obtained from each sensor modality, so if the sensors are fused immediately after cleaning, the sensitivity from each modality might be lost.

Response: We agree with the reviewer's comment. This research is exploratory in nature. However, the machine learning classifier has to be designed in a way that the information contained in features correlated with the pain event are not lost during sensor fusion.

6. Is the audio recording used for analysis? or is there any reason to obtain audio too?

Response: We audio recorded the patients in order to analyze cries' features as additional data that can also help identify behavioral responses to pain.

Reviewer: 2

Reviewer Name: carl saab
Institution and Country: Brown U, USA
Competing interests: None declared

Please leave your comments for the authors below
Timely study protocol that is overall comprehensive, but missing a few critical elements. For e.g., the ground truth for pain is subjective, as the authors acknowledge, so the training of the algorithms will be extremely hard. Given low sample size, this is a critical caveat (estimation of the sample size is not adequate).

Response: We recognize that the sample size may not be adequate but definitive evidence for that will require a quantitative assessment of the data variability. As a proof-of-principle pilot study, however, using a novel approach to extract features of the infant pain response from physiological recordings, this sample size should be enough.

This is more relevant because many confounding variables are not controlled for properly (age, gender, and type of painful interventions).

Response: The confounding variables will be more accurately defined and controlled in the next step of the project including a larger sample of patients.

Also, pain is multi-dimensional (sensory, cognitive, emotional, etc.) It is also acute, chronic, evoked and spontaneous. Which dimension / type is being measured?

Response: This pilot study is designed to exclusively assess acute pain responses during routine, clinically-required skin-breaking procedures – it measures the intensity of acute pain from the physiological responses of each subject. These physiological responses are arguably dependent on the sensory and emotional dimensions of pain. Preterm and term neonates, or infants of 1-3 or 3-6 months of age would not be expected to experience the cognitive dimensions of brief acute pain. The manuscript was revised accordingly to make it clearer including the primary objective page 7 (Marked copy).

Most of the sensors chosen, especially EEG, are not hypothesis-driven, so it's hard to judge the merit of the proposal (what changes are expected and in what direction?)

Response: Several papers have described the EEG patterns associated with neonatal pain in the published literature. The most frequent approach describing neonatal pain had used the EEG patterns in adults as a template of how they would identify the patterns of neonatal pain. However, the neonatal pain system is completely different – with different receptors, nerve pathways, neurotransmitters, as well as the cortical or subcortical areas participating in pain perception. In our study, we propose a more agnostic approach – with no pre-conceived patterns that we impose on to the newborn.

Movement artifacts will be the most significant challenge and a source for major noise. More detailed plans for remedy are needed.

Response: In order to address artifacts due to movement or suboptimal electrode-skin contact, we will initially use filtering techniques (e.g., to remove power line interference). In addition, we will identify channels exhibiting artifacts by considering the range of signal values, where signals showing extreme deviation from average values or channels showing virtually zero activity will be excluded. Specifically for EEG analysis, we will identify and remove EMG-related artifacts using well-established techniques such as filtering and independent component analysis. We revised the manuscript accordingly (page 11, Marked copy).

Reviewer: 3

Reviewer Name: Rebecca Pillai Riddell
Institution and Country: York University, Canada
Competing interests: None declared

Please leave your comments for the authors below
Please see report
* Please find this reviewer's report attached to this email *

Using Sensor-fusion and machine-learning algorithms to assess acute pain in nonverbal infants: A study protocol (www.Clinicaltrials.gov #NCT03330496)

Roué et al.

Preliminary Analysis of Protocol

● Protocol papers should report planned or ongoing studies. The dates of the study should be included in the manuscript. **Dates need to be included for the study period**

Response: Thank you for your comment. As listed in www.ClinicalTrials.gov, the study started on 30 October, 2017 and will be completed on 30 November, 2025. We included the dates in the revised manuscript (page 6, study design, Marked copy).

● While some baseline data can be presented, there should be no results or conclusions present in the study protocol. **No Results presented**

● For studies that are ongoing, it is generally the case that very few changes can be made to the methodology. As such, requests for revisions are generally clarifications for the rationale or details relating to the methods. If there is a major flaw in the study that would prevent a sound interpretation of the data, we would expect the study protocol to be rejected. **Understood**

Summary

This is a timely and interesting protocol to use sensor-fusion and machine-learning algorithms to assess acute pain in 60 preterm and term newborns (4 age groups from 34 weeks GA to full terms aged 6 months) undergoing painful procedures. The team plans to record five separate physiological measures (EMG, ECG, HR, HRV, SpO2 and EEG). It is a feasibility study setting out to finalize the approach to increase the accuracy of pain assessment and allow continuous pain monitoring at the bedside.

Strengths of the Protocol

- • Continuous pain assessment monitoring at bedside is a fundamentally important and critical objective within the NICU and I am convinced that automated bedside multimodal assessment is the most promising methodology to reduce unnecessary pain and utilize pain medication more precisely
- • The team is to be commended in doing a feasibility study with a specific goal of examining recording artifacts to better understand data corrections and sensor variability to have an appropriate estimate of power for the larger study.

  Response: Thank you for your comments.

  Challenges of the Protocol

• This is a feasibility study with strong multinational team. Our instructions were to in essence to determine if “there is a major flaw in the study that would prevent a sound interpretation of the data”. As I am working in this area, I know there are very few studies being done out there, so the idea of the fatal flaw was not the focus of my review for a feasibility study.

• There are a few areas where clarification/modification could be considered by the team to potentially increase the value of the study.

o Are there funding sources or conflicts of interest to acknowledge?

Response: Autonomous Healthcare, Inc. provided all the equipment for the study. The authors Haddad, Gholami and Anand have proprietary interests in the potential devices that may be developed from these studies. We have revised the competing interests section (Marked copy).

o Page 7, line 35: “We will record pain signals...”. I would debate this language as none of the physiological measures being recorded are pain-*specific* signals. While the EEG has shown specificity it is within a limited window (less than 1 second). This brings up the larger point with language around “continuous monitoring of pain”. Without specificity the multi-modal assessment, outside of a very circumscribed period of time (i.e. 1 sec post acute procedure), they are not continuously measuring pain. The language of ‘estimating pain’ used on page 8 seems more defensible

Response: We agree with the change in language to ‘estimating pain’. Previous studies have used EEG signals for a very brief period following acute pain (300-750 ms) from one single EEG lead (Cz), although that response may only represent cutaneous nociception from the initial pin-prick that is experienced, and not the ongoing pain from the consequent tissue injury. We believe that analysis of a more distributed EEG response over time will provide a more authentic picture of the ongoing infant pain. Thank you for your comments. We have revised the manuscript accordingly (Marked copy).

o Has EMG been validated for you in infants in the acute pain context? This would seem to be very invasive for infants. As they note EMG records signals much more coarsely. Thus, the placement on the cheek and forehead may not map anatomically onto BB, ES and NLF in pre-terms. We are doing work now suggesting that the cluster of facial actions are different in preterm and full term infants in the post heel-lance.

Response: EMG will be measured using leads with sticky pads and are therefore not considered as invasive. BB, ES and NLF furrow involve indeed several muscles layers that are impossible to measure individually. That is the reason why we decided to record facial EMGs signals from a facial region as opposed to any specific muscle. We therefore decided to place EMG sensors for NLF and mouth opening on the middle of the cheek in the axis of the NLF, and on the forehead over the periorbital area to measure the muscles involved in BB and ES. Based on the results of ongoing studies from Prof. Pillai Riddell’s group we will consider alternative locations for measuring the EMG response in preterm vs. term neonates. The manuscript was revised accordingly (Marked copy).

o Using a power analysis to justify 40 infants for a training dataset ignores the idea that they don’t have just one group they are spanning across a large range of infant development (cortically, motorically, psychosocially). One would assume the logic would be 40 infants per group- thus a feasibility study of 160 not 60. From my understanding of machine learning principles, in any scenario, 40 infants for a training set would be challenging.

Response: We recognize that the sample size may not be adequate but definitive evidence for its inadequacy will require a quantitative assessment of data variability. As a proof-of-principle pilot study, using a novel approach to extract features of the infant pain response from physiological recordings, this sample size should be enough.

o As HR and HRV are very closely related, there may be challenges in the assumptions of their modelling due to the dependency. They may want to consider one or the other.

Response: We agree the reviewer. We will investigate either HR or HRV to identify the metric showing a higher correlation with pain.

o There is not detail on the actual outcome variable of the EEG. What will they be measuring? In essence, the team reports the tool (Brain AMP cap with 32 active electrodes), but what will go into the machine learning algorithm? Moving away from a single electrode at vertex (which is an important direction) is great, but it is not clear how the immense data from 32-active electrodes will be used.

Response: This study is exploratory in nature, and hence, machine learning algorithm and feature extraction techniques showing desirable performance will be selected.

o Page 15, line 45: What validated pain scale will be used to identify the pain state of the infant

Response: Validated pain scales will be used including N-PASS, NFCS and PIPP-R for the neonates and FLACC scale or Visual Analogue Scale (VAS) for older infants. These pain scales are cited in section “Interventions/Experimental design” of the Methods with references. We specified them again at page 14 of the revised manuscript (Marked copy) which corresponds to the paragraph you mentioned.

o The intersection of gestational age, postnatal age and days in hospital does not seem to properly be described as a factor in the proposal. Imagine a 1-month old (postnatal age) who has been born

preterm vs full-term, juxtaposed with how many skin-breaking procedures that have done juxtaposed with varying lengths of hospital stay. Machine learning for discrimination without accounting for this type of variability using small sample size may cause challenge in understanding the data for a larger study.

Response: These factors will be taken into account in the analyses. Our initial studies will focus on term neonates within 1 week from birth and minimal exposures to prior painful events.

Reviewer: 4

Reviewer Name: Gilmer Valdes
Institution and Country: UCSF
Competing interests: None

Please leave your comments for the authors below
Objective pain assessment in non-verbal populations is clinically challenging and its misdiagnosed can have long term consequences. The authors propose to use Machine Learning to provide a "patient-centered, context-dependent, observer independent, and objective pain measure." To measure pain, the authors use Facial Electormyography (EMG) and electrocardiography (ECG) as surrogates. Statistical algorithms are proposed to find correlation between these signals and pain. The protocol is well written and it is very thorough.
The authors acknowledge their limitations and make efforts to address them. For instance, inherent bias and noise from each sensor or channel the authors use a "calibration" period where they acquire images to establish a baseline for these multiple sensors. There is a major limitation to the study that cannot be addressed though. "The clinical staff at the bedside will identify the pain state of each neonate using validated pain scales and record the timing of pain-inducing clinically-indicated procedures such as a heel stick." To me this is the main limitation of the study and what makes this problem particularly challenging. We still rely on a "subjective evaluation" of pain to find correlations between the sensors and the perceived pain.

Response: Thank you for your general comments.

In order to have a more objective evaluation, responses to nociceptive procedures will be compared to responses to non-painful stimulation (tactile stimulation before the recorded skin-breaking procedure) for every patient. Pain state using validated pain scales (NFCS, N-PASS, PIPP-R, FLACC, VAS) will be assessed as secondary outcomes.

Please find below, general comments:

1) The power calculation resulting in 40 patients needed is not well explained and it is hard for this reviewer to establish its correctness.
Response: Sample size calculations are described on Page 9 of the manuscript (Marked copy). Briefly, we first reviewed the variability of physiological responses to acute pain in the published literature. For sample size calculations, we assumed an α-error =0.05, 1–β error =0.8, and a mean:SD ratio of 2:1 and applied these parameters to a very general but elegant sample-size calculator developed by Russel Lenth (University of Iowa, http://homepage.divms.uiowa.edu/~rlenth/Power/index.html).

2) Since, the ML models will also be predicting subjective pain, the authors should examine as possible score to evaluate in the validation set a weighted majority vote between the algorithm predictions based on the sensors and the clinical classification of pain.
Response: We agree with your suggestion and we will incorporate it into the analysis plan.

3) Efforts should be made to validate the algorithms on independent cohorts, possible from different hospital systems.
Response: The observational prospective study proposed in this protocol is a feasibility study (cf methods in abstract and manuscript). We plan to validate the algorithms in larger samples including independent cohorts from different hospitals and countries.

4) If the authors are planning to provide interventions to manage the pain, they should consider causal statistical frameworks in their statistical analysis. This is lacking.

Response: All patients will receive non-pharmacological treatments (swaddling, non-nutritive sucking, sucrose) before and during the procedure. This will be considered for analysis.

VERSION 2 – REVIEW

REVIEWER	Raul Fernandez Rojas University of Canberra, Australia.
REVIEW RETURNED	26-Oct-2020

GENERAL COMMENTS	The authors did not mention the use of any windowing technique and length to analyze the data nor the type of machine learning algorithms to be used for the analysis.

REVIEWER	Saab, Carl Brown University
REVIEW RETURNED	26-Oct-2020

GENERAL COMMENTS	The EEG features to be incorporated into the algorithm(s) need to be specified, they cannot be open-ended. The authors could end up with hundreds of EEG variables in feature-space, which is unacceptable. There are several papers in the literature that could guide the authors.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Raul Fernandez Rojas

Institution and Country: University of Canberra, Australia.

Comments to the Author

The authors did not mention the use of any windowing technique and length to analyze the data nor the type of machine learning algorithms to be used for the analysis.

RESPONSE: We revised the manuscript (section “Data analysis/A. Signal-filtering and information extraction from sensors”, page 11) as follows: “Windows of different lengths will be used for the analysis. Specifically, for ECG, SpO2, EMG, and skin conductance signals an analysis window of 1-5

minutes will be used to extract appropriate features. The window lengths ranging from 400ms to 5 seconds will be used for EEG signals."

Details for the machine learning method were presented above and included in a new paragraph in section "Statistical approach- A. Sensor fusion" (page 15) as follows: "The extracted features from each modality will be used to train a machine learning algorithm. Specifically, we will train a binary classifier to assign "pain" and "no pain" class labels based on the extracted features. We will specifically investigate using the random forests classifier given their robustness to outliers and its classification performance when a large number of features are used for classification."

Reviewer: 2

Reviewer Name: carl saab

Institution and Country:

Comments to the Author

The EEG features to be incorporated into the algorithm(s) need to be specified, they cannot be open-ended. The authors could end up with hundreds of EEG variables in feature-space, which is unacceptable. There are several papers in the literature that could guide the authors.

RESPONSE: Please see the response above detailing the specific set of EEG features that will be investigated:

We revised the manuscript (section "Data analysis/A. Signal-filtering and information extraction from sensors", page 11) as follows: "Windows of different lengths will be used for the analysis. Specifically, for ECG, SpO2, EMG, and skin conductance signals an analysis window of 1-5 minutes will be used to extract appropriate features. The window lengths ranging from 400ms to 5 seconds will be used for EEG signals."

And we included new lines in the revised manuscript about EEG analyses (section "Data analysis/A. Signal filtering…/4. Electroencephalography", page 13)as follows: " In order to analyze EEG signals and extract appropriate features, we will first remove noise and artifacts using standard techniques such as independent component analysis (ICA) and wavelet denoising [Jiang X, Sensors (Basel). 2019 Mar; 19(5): 987, doi 10.3390/s19050987]. After artifacts are removed, we will investigate the correlation between features extracted from EEG data and pain. Specifically, we will use spectral decomposition and extract features such as mean power in different frequency bands (delta, theta, alpha, and beta) as well as asymmetry measures for each homologous pair and functional connectivity measures for further investigation."

Reviewer: 1

Competing interests 1: None

Reviewer: 2

Competing interests 1: none