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)	Assessment of the feasibility of an ultra-low power, wireless, digital patch for the continuous ambulatory monitoring of vital signs
AUTHORS	Hernandez-Silveira, Miguel; Ahmed, Kamran; Ang, Su-Shin; Zandari, Fahriya; Mehta, Tinaz; Weir, Rebecca; Burdett, Alison; Toumazou, Christofer; Brett, Stephen

VERSION 1 - REVIEW

REVIEWER	DOMENICO CIANFLONE, MD Director, Cardiovascular Prevention and Rehabilitation Unit, Ospedale San Raffaele & Università Vita-Salute San Raffaele Italy Consultant (2005-2011) Telbios-Milano-Italy Scientific Director AmicoMED-Quasarmed Milano-Italy
REVIEW RETURNED	23-Jan-2015

GENERAL COMMENTS	The paper reports the results of a well designed health tecnology assessment study regarding a novel and innovative approach to monitoring physiological parameters in an in-patients setting. The study was conducted appropriately and the results are interesting and promising. As a consequence of the presented data, the real world applicability of this novel technology needs further investigation and improvements: the inability to perform during occurrences of atrial fibrillation is a severe drawback in the post-surgical setting. Nevertheless the paper illustrates satisfactorily the move from the engineering-bench to bedside.

REVIEWER	Timothy Bonnici University of Oxford, UK Stephen Brett, the corresponding author, was the external examiner for my PhD transfer viva.
REVIEW RETURNED	25-Jan-2015

GENERAL COMMENTS	Overall the article provides a useful contribution to the field, both in describing the performance of a novel wearable sensor but also, most importantly, improving upon the testing methodology used in other papers describing wearable sensors by deliberately seeking out patients with comorbidities likely to pose problems for the sensor. This is not commonly done and provides a useful example for future researchers to follow. However, there are a number of issues which need to be addressed prior to publication. These are listed below.

	1. Is the research question or study objective clearly defined? No A minor point. The primary aim is said to be to assess the "performance" of the Sensium patch. This is a rather nebulous term. It would be more accurate to that the aim to assess the agreement of the Sensium patch with the Intellivue monitor. From reading the text I think the secondary endpoint has been wrongly described. See comments on the methods section below. 2. Is the abstract accurate, balanced and complete? Yes Would have liked to see a bit more detail in the methods section 3. Is the study design appropriate to answer the research question? Yes With caveats. See answer to 4. 4. Are the methods described sufficiently to allow the study to be repeated? No Inclusion and exclusion criteria not clearly stated. I imagine that there were more than simply age. In particular the inclusion criteria for the low voltage cohort need to be clearly explained. What criteria were made and who assessed whether they were met? It would be helpful to understand a bit more about what the patients were doing at the time of recording. The study has two aims: a) to assess the accuracy of the heart rate and respiratory rate measurements and b) secondly to evaluate the ability of the monitor to reject artifact corrupted data. Methods to meet the first aim are well described but for the second aim it is simply stated that the raw data were reviewed by three investigators. There is no mention of whether they were blinded to the results of the other assessors, whether they were blinded to the artifact rejection algorithm output, what criteria were used to decide whether quality was good or bad and what proportion of the recorded waveforms were assessed. If only a subsection of the waveforms were assessed then the sampling methodology should be described. These are all essential components of the methodology which should be described. However, reading the text suggests that the secondary aim is in fact reporting the proportion of data rejected by the artifact suppression algorithm. If this is the case then the secondary aim should be reworded. However, even if this change were made it would be ideal, but perhaps not absolutely essential, to have a description of the validation process as described above. Further comments below assume that the desire is to report on the proportion of data rejected. If this assumption is incorrect then substantial changes to the reporting of results would be required to describe whether the algorithm works well or not. 6. Are the outcomes clearly defined? No The outcome measures are described in the text but not as clearly as one might like. It would be nice to see them near the aims in order to make it easier to assess the limitations of the outcome measures in addressing the study aims. E.g. The primary aim was to measure agreement. The chosen outcome measures were limits of agreement and bias. The secondary aim was to measure rejection frequency. The outcome measure was proportion of data rejected. 7. If statistics are used are they appropriate and described fully? No The researchers use the Bland-Altman method described in their 1986 paper. However, this is not appropriate for repeated measures

	from multiple subjects because it does not account for the likely correlation between successive measurements nor that there is likely to be a high degree or within-subject correlation. Furthermore, because different recordings are different lengths and have a variable number of data points suppressed by the artifact rejection algorithm, each subject will contribute a different amount of data. This needs to be taken account of. Myles and Cui have suggested a method which addresses the majority of these issues. (Myles PS, Cui J. Using the Bland-Altman method to measure agreement with repeated measures. Br J Anaesth. 2007 Sep;99(3):309–11). The authors quote correlation coefficients. The Bland-Altman paper referenced by the authors explains why this statistic can be misleading, even when correlation is very high. This statistic is inappropriate and should be removed from the reported results. As I am not a statistician by training, I would recommend that a formal statistical evaluation carried out on any new statistical assessment to ensure that it is valid for use in this dataset. The statistics for validating instruments taking repeated measures of a variable whose true value is continually changing from multiple subjects is poorly described in the literature. 9. Do the results address the research question or objective? Yes See answer to 10. 10. Are they (results) presented clearly? No It would be good to see some data on the recruitment. The paper states that 61 subjects were recruited but the attached graphs show patient IDs in the 100s. Were some patients excluded from analysis perhaps? (Incidentally, if these are the actual Study IDs of patients then they should not be made publically available.) I would expect a table of relevant patient demographics. Some patient demographics are reported in the methods but should be in the results. Authors state that patients were monitored for approximately 2 hours. This statistic should be reported more precisely – e.g. median and IQR. Were there any significant discrepancies between subjects' contribution to the dataset? After rejection of artifact how many data points used for the analysis? The latter is stated in the figures but should be in the results text too or tabulated. It would be helpful to have the limits of agreement and bias data from the various groups tabulated for easy comparison. The caption for Table 1 suggests that there are reasons other than artifact which resulted in less than 100% available data. If this is the case these causes should be clearly explained. The dates of data collection start and finish are not reported. Therefore it is not possible to determine the veracity of the opening statement in the discussion. 12. Are the study limitations discussed adequately? No No mention of study limitations in within the discussion section, though I note a brief mention at on page 3 of the manuscript

VERSION 1 – AUTHOR RESPONSE

Reviewer Name DOMENICO CIANFLONE, MD

The paper reports the results of a well-designed health technology assessment study regarding a novel and innovative approach to monitoring physiological parameters in an in-patients setting. The study was conducted appropriately and the results are interesting and promising. As a consequence of the presented data, the real world applicability of this novel technology needs further investigation and improvements: the inability to perform during occurrences of atrial fibrillation is a severe drawback in the post-surgical setting. Nevertheless the paper illustrates satisfactorily the move from the engineering-bench to bedside.

Authors' Response: We thank the Reviewer for the comments and agree with his assessment that this early feasibility work is promising but also presents some drawbacks, which since the time of this study have undergone further development.

Reviewer Name TIMOTHY BONNICI

Overall the article provides a useful contribution to the field, both in describing the performance of a novel wearable sensor but also, most importantly, improving upon the testing methodology used in other papers describing wearable sensors by deliberately seeking out patients with comorbidities likely to pose problems for the sensor. This is not commonly done and provides a useful example for future researchers to follow.
However, there are a number of issues which need to be addressed prior to publication. These are listed below.

Q1: Is the research question or study objective clearly defined?
No. A minor point. The primary aim is said to be to assess the "performance" of the Sensium patch. This is a rather nebulous term. It would be more accurate to that the aim to assess the agreement of the Sensium patch with the Intellivue monitor. From reading the text I think the secondary endpoint has been wrongly described. See comments on the methods section below.

Authors' Response: We agree with the Reviewer that the primary aim of the study could be more clearly defined, and we have modified the Abstract text to reflect this suggestion. We have also modified the description of the secondary endpoint in the Abstract. We have further clarified that this is an early stage, feasibility assessment of a novel monitoring technology.

Q2: Is the abstract accurate, balanced and complete?
Yes, would have liked to see a bit more detail in the methods section

Authors' Response: We have included some further detail in the Methods section of the manuscript as we agree that this clarifies further the appropriateness of analysis methods selected. This detail outlines the sequential and non-continuous nature of the vital signs measurement, as well as describing the process by which signals deemed to be corrupted by motion or other artefact are rejected.

Q3: Is the study design appropriate to answer the research question?
Yes, with caveats. See answer to 4.

Authors' Response: The Reviewer's caveats are addressed in our answer to Q4.

Q4: Are the methods described sufficiently to allow the study to be repeated?

No. Inclusion and exclusion criteria not clearly stated. I imagine that there were more than simply age. In particular the inclusion criteria for the low voltage cohort need to be clearly explained. What criteria were made and who assessed whether they were met?

Author's Response: Main exclusion criteria are described in the section 'Methods: Study Procedure and Participants' (i.e. patients with pacemakers, implantable defibrillators or neurostimulators were excluded). Since this was an early technology assessment, the main inclusion criterion was to select available patients who gave consent during the study period. Group 1 was a group of ASA 1-2 patients presenting for elective surgery who did not meet the entry criteria for Group 2.

The four subgroups (diabetes, high BMI, AF and low voltage and abnormal QRS) were very clearly small exploratory comparisons and we believe obvious as such in the MS. As stated in the MS, the 'low voltage QRS' group was actually a somewhat heterogenous 'low voltage or abnormal ECG' group - again clearly exploratory in nature. Patients were screened and enrolled by the study nurse (FZ) with discussion with the clinical fellow KA or Chief Investigator (SJB).

It would be helpful to understand a bit more about what the patients were doing at the time of recording.

Authors' Response: The section titled 'Methods: Study Procedure and Participants' describes that '…group 1 patients were studied in the operating room while awaiting or recovering from routine surgery; group 2 patients were studied while resting in bed in ward or clinic areas…'

The study has two aims: a) to assess the accuracy of the heart rate and respiratory rate measurements and b) secondly to evaluate the ability of the monitor to reject artefact-corrupted data. Methods to meet the first aim are well described but for the second aim it is simply stated that the raw data were reviewed by three investigators. There is no mention of whether they were blinded to the results of the other assessors, whether they were blinded to the artifact rejection algorithm output, what criteria were used to decide whether quality was good or bad and what proportion of the recorded waveforms were assessed. If only a subsection of the waveforms were assessed then the sampling methodology should be described. These are all essential components of the methodology which should be described.
However, reading the text suggests that the secondary aim is in fact reporting the proportion of data rejected by the artifact suppression algorithm. If this is the case then the secondary aim should be reworded. However, even if this change were made it would be ideal, but perhaps not absolutely essential, to have a description of the validation process as described above.

Authors' Response: The investigators were blinded to the output of the artefact rejection algorithm, and this has now been clarified in the text under section 'Data handling and statistics'. The criteria by which the signals were judged to be of good or bad quality are now also better described (For ECG, a good signal is defined as that with at least 85% of data which is not corrupted by noise. For respiration, the signals must look like a 'quasi sine or triangular' periodic waveform that visually reflects inhalation/exhalation breathing patterns, and should exhibit noisy proportions no larger than 25% of the size of the respiration peaks and valleys seen in the signal).

Further comments below assume that the desire is to report on the proportion of data rejected. If this assumption is incorrect then substantial changes to the reporting of results would be required to describe whether the algorithm works well or not.

Authors' Response: The reviewer's assumption is correct, and this has been clarified in the study aims as mentioned in Q1.

Q5 (not listed)

Q6: Are the outcomes clearly defined?
No. The outcome measures are described in the text but not as clearly as one might like. It would be nice to see them near the aims in order to make it easier to assess the limitations of the outcome measures in addressing the study aims. E.g. The primary aim was to measure agreement. The chosen outcome measures were limits of agreement and bias. The secondary aim was to measure rejection frequency. The outcome measure was proportion of data rejected.

Authors' Response: The outcome measures have been more clearly stated in the Abstract as suggested.

Q7: If statistics are used are they appropriate and described fully?
No. The researchers use the Bland-Altman method described in their 1986 paper. However, this is not appropriate for repeated measures from multiple subjects because it does not account for the likely correlation between successive measurements nor that there is likely to be a high degree or within-subject correlation. Furthermore, because different recordings are different lengths and have a variable number of data points suppressed by the artifact rejection algorithm, each subject will contribute a different amount of data. This needs to be taken account of. Myles and Cui have suggested a method which addresses the majority of these issues. (Myles PS, Cui J. Using the Bland-Altman method to measure agreement with repeated measures. Br J Anaesth. 2007 Sep;99(3):309–11).
The authors quote correlation coefficients. The Bland-Altman paper referenced by the authors explains why this statistic can be misleading, even when correlation is very high. This statistic is inappropriate and should be removed from the reported results.
As I am not a statistician by training, I would recommend that a formal statistical evaluation carried out on any new statistical assessment to ensure that it is valid for use in this dataset. The statistics for validating instruments taking repeated measures of a variable whose true value is continually changing from multiple subjects is poorly described in the literature.

Authors' Response: We disagree with the reviewer that the Bland-Altman method is not appropriate for repeated measurements, since Bland and Altman themselves describe in 2007 that such plots can also be used for dependent data in the classic way without need for further adjustments provided that the true value varies – which is the case with patients' vital signs. Moreover, the assumption of statistical independence is strong since the vital signs readings are sequential and non-continuous as now highlighted in the paper. We have clarified in the data sampling in the MS; essentially, the technology cycles through sampling different variable- thus each reported value is an entirely independent quantum.

We share the Reviewer's opinion of the lack of appropriateness of correlation as a means of assessing performance, hence our reliance on the Bland-Altman plots; however, these data are commonly reported and we have included them for completeness.

Q8 (not listed)

Q9: Do the results address the research question or objective?
Yes See answer to 10.

Q10: Are they (results) presented clearly?
No. It would be good to see some data on the recruitment. The paper states that 61 subjects were recruited but the attached graphs show patient IDs in the 100s. Were some patients excluded from

analysis perhaps? (Incidentally, if these are the actual Study IDs of patients then they should not be made publically available.) I would expect a table of relevant patient demographics. Some patient demographics are reported in the methods but should be in the results. Authors state that patients were monitored for approximately 2 hours. This statistic should be reported more precisely – e.g. median and IQR. Were there any significant discrepancies between subjects' contribution to the dataset?

Authors' Response: The patient IDs are unrelated to the patient recruitment number in the trial and thus do not indicate exclusion of a large number of patients from the data analysis.
Patients were monitored for variable periods of time; this is because it was not possible to control how long the patients were monitored pre- and post-surgery as this was determined by the hospital workflow; this is explained in the study limitations. Further demographic data have been added to the text of the Results. The relative contribution of each patient's data to the overall analsyses can be seen in the electronic supplement.

After rejection of artifact how many data points used for the analysis? The latter is stated in the figures but should be in the results text too or tabulated. It would be helpful to have the limits of agreement and bias data from the various groups tabulated for easy comparison.

Authors' Response: These data have been added into the text as requested. We felt the limits of agreement and bias data were easier to follow if they were presented in the text with close reference to the relevant figure- rather than comparing a figure and data in a table. We recognise this is a matter of taste and journal style and could easily generate a table if the Editor so wished.

The caption for Table 1 suggests that there are reasons other than artifact which resulted in less than 100% available data. If this is the case these causes should be clearly explained.

Authors' Response: The other cause of data rejection apart from motion artefact is for patients with atrial fibrillation. This is explained in the discussion.

The dates of data collection start and finish are not reported. Therefore it is not possible to determine the veracity of the opening statement in the discussion.

Authors' Response: Start and finish dates have been added into the text

Q11 (not listed)

Q12: Are the study limitations discussed adequately?
No. No mention of study limitations in within the discussion section, though I note a brief mention at on page 3 of the manuscript.

Authors' Response: Study limitations have been added into the Discussion section. We have also further highlighted throughout the text that this study was an early assessment of the feasibility of this ambulatory monitoring technology, and thus is not presented as a complete and exhaustive study but rather a basis for further developments.