Reviewer's report

Title: Vibration response imaging evaluation

Version: 5 Date: 17 June 2009

Reviewer: Mordechai Yigla

Reviewer's report:

Comments for the Author

Title: Vibration Response Imaging Evaluation

The authors investigated agreement between different physicians in evaluating images derived from lung sound recordings utilizing a new, non-invasive technology. For each new technology, it is essential that the results are reliable and that they clearly differentiate between normal findings and findings from patients with pathologies. In this study, Bartziokas and co-authors described agreement among raters who evaluated the different aspects of VRI images that were used to determine healthy lung patterns from abnormal distribution of lung sound intensity in areas corresponding with consolidation on chest radiographs. The results are convincing; however, the study design is lacking a systematic study of the inter- and intra-observer agreement in interpretation of the images. If the authors can elaborate on the analysis, in a manner similar to Maher et al (as referenced in the current study), then this study would further validate this technology as an applicable tool with minimal training and validate the device’s role in disease states.

There are some crucial elements of the statistical analysis that should be clarified in order to strengthen the study design and the findings. In general, the authors should review the statistical design of Maher et al and consider utilizing the statistical tests in that published work (i.e. intra-class correlation).

1. The authors didn’t state the findings for intra-observer agreement analysis that was described in the methodology.

2. Need to describe the methodology for comparing agreement among the 6 different reviewers versus the gold standard physician diagnosis (based on physical examination, medical history and chest x-ray). State specifically that chest x-rays were performed on healthy controls, if this was the case.

3. pg. 7: VRI item #3 that was evaluated: presence of artefacts (yes/no) – should also report the number of artefacts, since a threshold of <3 or >3 was used for later analysis.

4. pg. 7: How were the findings from items #4-6 combined in order to report on localization of abnormality agreement in Table 2?

5. pg. 7, were all the images that were used for the blinded analysis of the 6
physicians also used for intra-observer evaluation? Were the images presented twice?

6. If the images were presented twice for the intra-observer evaluation, then which results were used for assessing inter-rater agreement afterwards once intra-observer agreement was established (the 1st or 2nd assessment)?

7. pg. 9, which test was used to establish a p value of 0.2 for testing significance between VRI recording time in patients and controls?

8. pg. 9: Add “Among the patients with pneumonia”...20 cases had COPD, 8 cases arterial hypertension...

9. Table 1: Are the numbers reversed for gender in the control group (i.e. distribution of F/M is different than for patients)?

10. Table 2: Since there is good agreement between rater assessments and chest x-ray findings, reported in Table 2, one can assume that there was good agreement between the raters; however, the authors should state the range of variability among the raters.

11. Table 2: Don’t need to include yes/no categorization for the presence of artefacts.

12. Table 2: Abnormally decreased signal –left lower zone: the value of 0.58 for overall agreement seems low, since agreement for controls is “1”...perhaps this is a typo?

13. Discussion: the authors report on the reproducibility statistics of Maher et al and then state that the current study further validates the reproducibility of the device. However, the statistical design does not support this claim; instead it tests the reliability of rater assessment. It would be more accurate to report the inter-rater/intra-rater agreement rates from Maher’s paper and then make a direct comparison including the intra-rater (not reported here) and inter-rater agreement rates from the current study.

14. Discussion: Were the results of total left distribution versus total right investigated? The results for left lower lobe might have been different than right lung due to the different number of lobes in each lung.

15. Discussion: can add to the discussion that in general, kappa underestimates the strength of agreement, so it is possible that the results could even be more robust than reported.

Minor comments:
1. The title should be more descriptive of the study (i.e. evaluation of rater agreement in healthy subjects and subjects with pneumonia).
2. In addition there are spelling and grammatical errors that should be corrected.
Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

Dr Yigla Serves has an consultation agreement with deepbreeze, ISRAEL, the company which has had invented the VRI machine. In the last 4 years he is paid hourly for interpreting VRI images recorded during clinical studies.

The answers for the above mentioned items are negative.