Author's response to reviews

Title: Diagnostic performance and inter-observer concordance in lesion detection with the Automated Breast Volume Scanner (ABVS)

Authors:

Sebastian Wojcinski (s@wojcinski.de)
Samuel Gyapong (samuel.gyapong@franziskus.de)
Andre Farrokh (afarrokh@online.de)
Philipp Soergel (soergel.philipp@mh-hannover.de)
Peter Hillemanns (hillemanns.peter@mh-hannover.de)
Friedrich Degenhardt (frideg@t-online.de)

Version: 3 Date: 24 April 2013

Author's response to reviews: see over
Re-Submission after major revision

“Diagnostic performance and inter-observer concordance in lesion detection with the Automated Breast Volume Scanner (ABVS)”

Dear Editors and Reviewers,

thank you very much for your review of the manuscript. I appreciate the energy and time that you have spent and your very helpful recommendations.

We have carried out a major revision of the manuscript. All of your detailed comments and suggestions have been incorporated into the revised draft. The new version has additional literature, and is depleted from redundant parts.

In order to facilitate the review process, relevant changes and new sections are highlighted using the “track changes” function.

Thank you again for your suggestions – we hope that our revision covers all of the necessary aspects.

With kind regards,

Dr Sebastian Wojcinski
Diagnostic performance and inter-observer concordance in lesion detection with the Automated Breast Volume Scanner (ABVS). The manuscript by S. Wojcinski et al. investigates “diagnostic performance and inter-observer concordance in lesion detection with the Automated Breast Volume Scanner (ABVS).” The subject is relevant and statistical methods used are familiar and well-known. Diagnostic performance of ABVS is rarely reported in the literature and this work is a good effort. However, the most important problem of this manuscript is that it is not well-written.

In addition, I have the following specific comments for the authors:

1) The authors claim in the “Materials and Methods” section that the responsible ethical committee did not require additional approval for this non-interventional study-design. But, it is unclear informed consent or waiver of informed consent for at the beginning of study. The authors must indicate that if there is any informed consent at the beginning of the study.

2) The authors were expressed this study is a cohort study in the “Materials and Methods, and Results” sections. But, the definition of cohort study is “In statistics and demography, a group of subjects who have shared a particular event together during a particular time span [1]. Cohorts may be tracked over extended periods in a cohort study. A cohort study is a form of longitudinal study (a type of observational study) used in medicine, social science, actuarial science, and ecology. A longitudinal study is a correlation research study that involves repeated observations of the same variables over long periods of time—often many decades. But this study was carried out between March 2010 and July 2011. So, this study can’t be considered as the cohort study.

3) In the “Statistical analysis” section, details of the used statistical methods are missing, and/or poorly justified/interpreted. In this section, the authors were specified three different statistical analyses. Firstly, the diagnostic sensitivity and specificity, as well as the accuracy of the ABVS, were calculated based on the Bayesian theorem using Fisher’s exact test. Secondly, the statistical analysis of the extent of agreement between the two raters was based on Cohen’s Kappa test.
the end, pair wise comparisons of proportions were performed using the Z-test. But it wasn’t explained that why these methods used for these statistical analyses. How the authors decided to use this tests? Why the authors didn’t used a receiver operating characteristic (ROC), or simply ROC curve instead of Fisher’s exact test? I suggest that, this section must be reevaluated by the authors.

In order to assess the diagnostic performance of the ABVS, we calculated sensitivity, specificity and accuracy for both examiners and used the Z-Test to compare the performance of examiner 1 with examiner 2. As our study population did not reflect the real prevalence of breast cancer, the positive and negative predictive values were estimated based on the Bayesian theorem using the reported prevalence of malignancies in screening collectives [19]. For the calculation of the 95% confidence levels, we used the Newcombe intervals with continuity correction [20].

The statistical analysis of the extent of agreement between the two raters was based on Cohen’s Kappa test. For the interpretation of \( \kappa \)-values we used the magnitude guidelines published by Landis and Koch, who characterized the values of \( \kappa < 0 \) as indicating no agreement, \( \kappa 0-0.20 \) slight, \( \kappa 0.21-0.40 \) fair, \( \kappa 0.41-0.60 \) moderate, \( \kappa 0.61-0.80 \) substantial, and \( \kappa 0.81-1 \) as almost perfect agreement [21].

Furthermore, we assessed the correlation between the expected and the observed rate of second-look ultrasounds using the Chi-square test.

Statistical significance was assumed as \( p<0.05 \) for all tests."

However, we decided against ROC curve, as we cannot define various thresholds in our analysis. We consider mere calculation of SE/SP to be more suitable for our clinical data.

4) In the “Statistical analysis” section, the authors were stated that, "Microsoft® Office Excel® 2007 (Microsoft Corporation) was used for data collection" and “statistical analysis was performed by the author SW and validated by the other authors”. But they didn’t mentioned, how they calculated the Fisher’s exact test values and, Cohen’s Kappa test values that presented in the results section. If any statistical program used the calculation of the statistical parameters in the study, this issue should be clarified.

We added the necessary information to the paper: "Microsoft® Office Excel® 2007 (Microsoft Corporation) was used for data collection. Statistical analysis was performed by the author SW using MedCalc® 7.6 statistical software (MedCalc Software bvba, Belgium) and validated by the other authors.”

- Minor Essential Revisions
  1. In the introduction section, the authors expressed that “Furthermore, not all of the images data from ultrasound examinations can be stored, and only subjectively chosen screenshots are archived.” At the present time, the new ultrasonography devices can be stored all the images data from ultrasound examinations. But the physician may archived only subjectively chosen screenshots. The authors could correct this matter.

We intended to express that “not the entire examination” can be stored. We agree that our expression is confusing. We revised this sentence: “Furthermore, only subjectively chosen screenshots from the ultrasound examination are printed and/or stored.”

2) The sample size of the study population was not authors.

We now give more detailed information about this
clearly expressed (how many women? or how many breast examination?) from the "Patient Database" subsection in "Materials & Methods" that the patients enrolled in this study have given informed consent. This issue should be clarified.

3) The authors stated that the discussion section many prior studies about the ABVS accuracy, specificity and the sensitivity. A table listing these studies and key characteristics would be helpful.

- Discretionary Revisions
1) In the introduction section, there is a "Mammography has demonstrated excellent sensitivity, specificity and inter-observer concordance." sentence. At the end of this sentence a reference may be added

Two more references have been added:


Thank you for this recommendation. A table containing these data has been added.

<table>
<thead>
<tr>
<th>Reviewer’s report:</th>
<th>Thank you for your evaluation of our paper. Ongoing studies will focus on these issues.</th>
</tr>
</thead>
<tbody>
<tr>
<td>An average tumor size is little bit larger. I hope to include more small cases at the next study. Level of interest: An article of importance in its field Quality of written English: Acceptable Statistical review: No, the manuscript does not need to be seen by a statistician.</td>
<td>Thank you for your evaluation of our paper. Ongoing studies will focus on these issues.</td>
</tr>
</tbody>
</table>

The manuscript by Wojcinski et al. describes the diagnostic performance and inter-observer concordance in lesion detection with the Automated Breast Volume Scanner. It is a quite good work. Here are my comments to the authors:

Minor Essential Revisions
Page 8: The inplane resolution of the images should be inserted after "300 high resolution slices".

Page 10: In the manuscript, it is told that first suspicious regions are marked with the systems default tool and all of the "potential" lesions selected are evaluated by the physicians. This would cause some some bias that would increase the number for false-positives. This issue should be clarified!

We added the additional information:

"...acquiring about 300 high-resolution slices for post-processing (resolution: axial=0.09 mm, lateral=0.16 mm, sagittal=0.44 mm)."

We revised the corresponding paragraph as it was misleading. There is no "automated lesion identification tool". However, the examiner can mark regions of interest by his own choice. This enables the examiner to relocate the lesions later.

"During this process, the examiner marked all mass lesions with the system’s default tool. Next, the examiner evaluated all of the selected lesions by displaying them in the sagittal and axial planes. Finally, the examiner assigned the lesions a
category according to the ACR Bi-RADS®-US system.”

We agree that this condition would increase the number for false-positives. However, we have already discussed this point:

“Actually, we had a relevant number of healthy women without breast lesions (i.e. Bi-RADS®-US 1) in our collective. This condition automatically increases the false-positive rate and actually reduces sensitivity. Hence, the main difference of our study design in comparison to the above-mentioned studies is that we did not focus on the evaluation of known lesions alone, but also on the detection. Therefore, our examiners did not know whether there was a lesion in the particular volume or not. In the studies from the literature, the examiners were well aware, that there is definitely a lesion in the volume that requires biopsy (i.e. Bi-RADS® 3 and above). This knowledge certainly has an influence on the evaluation of the lesion and the overall performance of the ABVS. Our interpretation: the almost perfect performance of the ABVS as described in the literature is only valid for collectives of women with already pre-diagnosed breast lesions. In realistic collectives that also include women without breast lesions (i.e. Bi-RADS®-US 1) or with clearly benign breast lesions (i.e. Bi-RADS®-US 2), our data may be more convincing.”

Page 11: It is told that the physicians know that there are no BIRADS-US 0,3 or 4 cases in the database. This would also lead to some bias that should be discussed in the manuscript.

We added a more detailed discussion to the “limitations”:

“Another limitation of our study is the selection of the study population and the resulting concentration on Bi-RADS®-US 1, 2 and 5 lesions which causes a certain bias. Therefore, the proportion of cases to controls is not representative of the whole population and Bi-RADS®-US 0, 3 and 4 lesions are missing in the study population. We attempted to overcome this limitation by conducting a model calculation. Although this approach is statistically correct, it must be considered that, due to the small sample size and the vague estimation of the prevalence, the results must be carefully interpreted.”

Page 13: Mean tumor size are reported for malignant lesions. A similar reporting should be done for benign lesions.

Actually, we did not collect these data and, consequently, did not analyze this point.

The focus of our study was on the detection of the malignant lesions. There was a maximum of 1 malignant lesion in each volume. Therefore, analysis of the mean tumor size was possible for the malignancies.

However, there may be multiple benign lesions in each volume (e.g. multiple uncomplicated cysts). Analysis of the numerous smaller or larger benign
lesions that may appear in the breast tissue was not feasible in our study design. Therefore, we cannot provide helpful data.

Page 23: It is told that, roughly, 95% of breast parenchyma is included in the scanned volume. I wonder how the authors reach this conclusion. This should be clarified in the manuscript.

We admit, that this is, as indicated, a very rough estimation based on our personal experience. Therefore, we changed the sentence:

“Therefore, a certain proportion of breast parenchyma may be lost in the volume data.”

The time required for the physician to complete his evaluation from the 3D images acquired should be mentioned and discussed in the manuscript.

Our study was not designed to answer this question. Nevertheless, we put some general information into the Methods section:

“Overall interpretation times usually range from 4 to 10 min per case.”

However, as we have no detailed data concerning this point, we did not add this issue to the discussion.

In evaluation of breast cancer, it is also very important to assess lymph node extent. But this is impossible with ABVS. This issue should be discussed in details in the manuscript.

We added a more detailed description the discussion:

“In addition, the ABVS is not capable of scanning the axillary region. Today, sentinel node biopsy is the standard therapy for women that were preoperatively staged with a negative nodal status, which requires ultrasound of the axilla [32]. Furthermore, lymph node alterations may be the first sign in mammographically and/or sonographically occult breast cancer or other malignant diseases [33, 34]. Therefore, additional conventional ultrasound of the axilla would be necessary after a suspicious ABVS scan.”

I wonder what happens when the automated lesion identification tool that comes with the ABVS’s software is turned off.

We revised the corresponding paragraph as it was misleading. There is no “automated lesion identification tool”. However, the examiner can mark regions of interest by his own choice. This enables the examiner to relocate the lesions later.

“During this process, the examiner marked all mass lesions with the system’s default tool. Next, the examiner evaluated all of the selected lesions by displaying them in the sagittal and axial planes. Finally, the examiner assigned the lesions a category according to the ACR BI-RADS®-US system.”

Major compulsory revisions

Page 19: Authors mention that “there is no data in the literature describing the interobserver concordance in lesion detection with the ABVS, and our study is the first to scrutinize this issue.”. Unfortunately this is not correct. I recommend the authors to read the following published papers and inspire them in the manuscript in details:

Automated ultrasound of the breast for diagnosis:

Thank you for this valuable comment. We revised this section of the discussion:

“There is only limited data in the literature describing the inter-observer concordance in lesion detection with the ABVS:

In 2011, Shin et al. reported on 55 women with 145 breast masses who were examined with handheld ultrasound and the ABVS [25]. Five radiologists reviewed the volume data and detected between 74% and 88% of the lesions. Substantial agreement was found for BI-RADS® final assessment category
Interobserver agreement on lesion detection and characterization.

Interobserver reliability of automated breast volume scanner (ABVS) interpretation and agreement of ABVS findings with hand held breast ultrasound (HHUS), mammography and pathology results.

(κ=0.63).
Recently, Golatta et al. published data on 84 single breast examinations in 42 women [26]. Six breast diagnostic specialists interpreted the 3D-images. Based on the BI-RADS® classification the multiple kappa coefficient was κ=0.35.

In our analysis, we found fair agreement between the two examiners, which correlates with the latter results (κ=0.27). However, more data is needed to evaluate the performance of the ABVS in breast imaging convincingly.”

Thank you again for your detailed and helpful comments!