First Evaluation of the Diagnostic Accuracy of an Automated 3D Ultrasound System in a Breast Screening Setting

FRANK STÖBLEN¹, SOLVEIG LANDT², RUTH STELKENS-GEBAHRDT¹, JALID SEHOULI³, MAHDI REZAI⁴ and SHERKO KÜMMEL⁵

¹Department of Radiology, Hyssensstift Kliniken-Essen-Mitte, Essen, Germany;
²Department of Gynecology and Obstetrics, University Hospital Düsseldorf, Düsseldorf, Germany;
³Department of Gynecology and Obstetrics, Charité University Hospital Berlin, Berlin, Germany;
⁴Breast Center, Luisenkrankenhaus, Düsseldorf, Germany;
⁵Breast Center, Huyssensstift Kliniken Essen-Mitte, Essen, Germany

Abstract. Background/Aim: Automated ultrasound examination of suspicious findings can reduce the physician’s workload in screening mammography. The present study examines the diagnostic accuracy of this method in comparison to mammography as the reference standard for the first time. Patients and Methods: A total of 304 patients underwent automated 3D ultrasound examination after screening mammography. Mammograms and ultrasound images were assessed by independent examiners, and sensitivity, specificity and the degree of agreement between both methods were calculated. Results: The degree of agreement was moderate (Cohen’s $\kappa = 0.130$ for all and 0.153 for positive/negative ratings), mainly owing to a high percentage of false-positive ultrasound results. However, the results of sonographical re-examination of suspicious mammograms were favorable. The only two undetected proven malignant lesions were microcalcified, and in three more cases with disagreement, the ultrasound diagnosis was correct. Conclusion: Automated 3D ultrasound imaging appears to be on a par with hand-held ultrasound in terms of diagnostic quality.

Ultrasound examination of the breast has evolved into an indispensable diagnostic tool for early detection of breast cancer since its introduction in the 1950s (1). According to current guidelines in Germany, ultrasound examination is recommended in the following situations (2): As a first-line method for the assessment of palpable lesions in women under 40 years of age; for the assessment of mammographically suspicious lesions (ACR 3-4, BI-RADS 0, III, IV and V) detected after clinical suspicion (40-49 and over 70 years) or routine screening mammography (50-69 years); and for interventional biopsy in BI-RADS IV/V lesions.

Presently, ultrasound examination is not recommended as the sole method of breast cancer screening.

Like ultrasound in general, sonographic breast diagnosis can be extremely accurate (3), but its quality depends not only on a highstandard of technical equipment (1, 2), but also on the proficiency, experience and diligence of the examiner (4). When performed thoroughly, the examination requires approximately 20 minutes of physician time according to the recently published results of the ACRIN 6666 study (5).

Another potential pitfall is the fact that sonography results are not automatically stored, limiting the possibilities for review under diagnostic, but also forensic aspects.

Since breast cancer screening is a setting with high economical impact, an optimized relationship between cost and benefit is of pivotal importance, and the reduction of physician time can be a means to this end, directing research focus on automated diagnostic systems that can be applied by medical technicians or assistants (6).

A relatively recent development in ultrasound diagnostics of the breast is the possibility for automated 3D real-time imaging that offers a number of potentially significant advantages over conventional ultrasound (1, 7, 8): Improved differentiation of architectural distortion, especially in the ‘bird’s eye’ view; better appreciation of tumour volume, indispensable for monitoring of patients under neo-adjuvant treatment regimens; storage of the complete imaging data for off-line expert review and image processing; securing of complete volume coverage; and reduction of physician time, i.e. cost.

Presently the rationale for automated 3D imaging in breast cancer screening in addition to or instead of conventional
ultrasound techniques is unclear, and clinical experience is rather limited (7-10). The issue of a possible application of automated ultrasound breast examination for the screening of dense breasts in addition to mammography is, however relevant, not the subject of the present study; rather this study examines the diagnostic accuracy of automated ultrasound breast examination in comparison to mammography as the reference standard.

Patients and Methods

Patients. Patients were recruited for the trial between August 26th and November 10th, 2008. All women attending the diagnostic centre for screening mammography (which is funded by compulsory health insurance (CHI) on a bi-annual basis for women of 50-69 years of age in Germany) during this period were eligible, and 310 consecutive patients were considered for participation.

Breast density was not a criterion for enrolment; patients with densities ACR 1-4 were enrolled, and the majority were classified as ACR 2 (‘fat with some fibroglandular tissue’).

Informed consent (oral and written) was obtained from all patients, and no patient refused participation after receiving the information regarding the study. Pre-menopausal patients were questioned about the possibility of pregnancy, which was denied by all.

The study met the criteria of ‘Good Clinical Practice’ and the principles of the Declaration of Helsinki.

Patients were between 50.1 and 69.8 years of age upon examination (mean 58.3±6.4 years). The diagnostic procedure was completed in 304 patients per protocol; the patient data was made anonymous and processed for statistical evaluation.

Data collection was planned before the breast scanner and mammographic examinations were performed, i.e. the study was prospective.

Diagnostic procedures. After declaring informed consent, patients first underwent mammography with the Mammomat Novation DR full-field digital system in combination with the syngo Acquisition Workstation (AWS) and MammoReport breast care workplace (Siemens Healthcare, Erlangen, Germany). This system is nationally and internationally approved for breast cancer screening.

According to the manufacturer’s recommendations, the W/Rh target/filter combination was employed for all breast types even though Mo/Mo and Mo/Rh target/filter combinations are also available. All examinations were performed under automatic exposure control (AEC).

After completion of mammography, automated 3D ultrasound examination was performed with a SomoVu device (U-Systems, San Jose, CA, USA) in technology and distribution collaboration with Siemens Health Care Inc., Ultrasound Division, Mountain View, CA, USA). This system allows automated acquisition of multi-plane ultrasound images of the breast by serial images in a volume of up to 14.5x17x5 cm. The images are interpreted at the BreastView workstation.

In contrast to standard screening procedure, mammographies were assessed by two independent, experienced radiologists who had no access to each other’s diagnoses or the ultrasound images in order to rule out inter-individual differences as a cause for diagnostic errors.

<table>
<thead>
<tr>
<th>Patients and Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients. Patients were recruited for the trial between August 26th and November 10th, 2008. All women attending the diagnostic centre for screening mammography (which is funded by compulsory health insurance (CHI) on a bi-annual basis for women of 50-69 years of age in Germany) during this period were eligible, and 310 consecutive patients were considered for participation.</td>
</tr>
<tr>
<td>Breast density was not a criterion for enrolment; patients with densities ACR 1-4 were enrolled, and the majority were classified as ACR 2 (‘fat with some fibroglandular tissue’).</td>
</tr>
<tr>
<td>Informed consent (oral and written) was obtained from all patients, and no patient refused participation after receiving the information regarding the study. Pre-menopausal patients were questioned about the possibility of pregnancy, which was denied by all.</td>
</tr>
<tr>
<td>The study met the criteria of ‘Good Clinical Practice’ and the principles of the Declaration of Helsinki.</td>
</tr>
<tr>
<td>Patients were between 50.1 and 69.8 years of age upon examination (mean 58.3±6.4 years). The diagnostic procedure was completed in 304 patients per protocol; the patient data was made anonymous and processed for statistical evaluation.</td>
</tr>
<tr>
<td>Data collection was planned before the breast scanner and mammographic examinations were performed, i.e. the study was prospective.</td>
</tr>
</tbody>
</table>

| Diagnostic procedures. After declaring informed consent, patients first underwent mammography with the Mammomat Novation DR full-field digital system in combination with the syngo Acquisition Workstation (AWS) and MammoReport breast care workplace (Siemens Healthcare, Erlangen, Germany). This system is nationally and internationally approved for breast cancer screening. |
| According to the manufacturer’s recommendations, the W/Rh target/filter combination was employed for all breast types even though Mo/Mo and Mo/Rh target/filter combinations are also available. All examinations were performed under automatic exposure control (AEC). |
| After completion of mammography, automated 3D ultrasound examination was performed with a SomoVu device (U-Systems, San Jose, CA, USA) in technology and distribution collaboration with Siemens Health Care Inc., Ultrasound Division, Mountain View, CA, USA). This system allows automated acquisition of multi-plane ultrasound images of the breast by serial images in a volume of up to 14.5x17x5 cm. The images are interpreted at the BreastView workstation. |
| In contrast to standard screening procedure, mammographies were assessed by two independent, experienced radiologists who had no access to each other’s diagnoses or the ultrasound images in order to rule out inter-individual differences as a cause for diagnostic errors. |
The results of the comparison of raw mammography and 3D-sonography data are displayed in Table I. In all instances where a consensus conference was held, because of contradictory results, those results were used; in all other cases, the higher mammography BI-RADS rating was employed for analysis. According to this table, the prevalence of BI-RADS IV was 4.6%, and the sensitivity and specificity of the 3D ultrasound scanning was 64.3% and 79.3%, respectively. Cohen’s $\kappa$ (weighted) was 0.130 for all ratings and 0.153 for positive/negative ratings, respectively, indicating only a moderate degree of agreement between both methods.

This was mainly due to a substantial percentage of false-positive ultrasound ratings: In 60 cases (20.4% of the entire sample and 20.7% of patients with mammography BI-RADS I or II), the ultrasound examination yielded a rating of BI-RADS IV/V, despite a non suspicious mammogram (I/II). Moreover, in those 5 cases where the sensitivity/specificity analysis yielded false-negative results of the breast scan (measured against the mammography as a reference), further analysis showed that only in two cases was the ultrasound rating indeed wrong, whereas in three cases, the mammography conference decision was wrong and the sonography rating was correct (Table I).

According to the clinical application of ultrasound in breast screening, however, false-positive results are irrelevant because only mammographically suspicious lesions are evaluated in the first place. Therefore, those cases in which the ultrasound examination would have been performed in a clinical screening setting (i.e. mammographies with BI-RADS-classification of III, IV, and V by one or both investigators) were reviewed in more detail (Table II).

In both cases where an actual tumor was not detected sonographically, the neoplasm was microcalcified, hence eluding ultrasound imaging. In all other instances, the conference decision or the histological diagnosis, respectively, were in accordance with the breast scanner result, so that...
false-negative results of the latter occurred exclusively when microcalcification was present. More importantly, the automated 3D ultrasound imaging was more accurate than the mammography in all instances where there was disagreement between the two and lesions were not microcalcified (Figure 1). The imaging findings of two patients are shown in Figure 2 (benign growth) and Figure 3 (cyst).

Furthermore, Table II underlines that the level of disagreement between both investigators rating the mammograms was substantially higher than that between the conference decision and the ultrasound result.

Discussion

From a practical point of view, automated 3D ultrasound has a number of compelling advantages that warrant its further examination in systematic studies. Most important is the reduction and adaptation of the time a physician spends on the diagnosis of a given patient. The actual examination of 6 scans per patient at the workstation only requires approximately 5 min. Moreover, off-line assessment at a time of the examiner’s choice greatly reduces organizational burden, and the complete storage of imaging data allows for the convenient consultation of colleagues. A spin-off effect is the complete documentation of the diagnostic process for later retrieval, be it for medical or forensic reasons. Further potential benefits are the opportunity to utilize volume measurement as a tool for the assessment of chemotherapy effect, differential viewing perspectives and whole-breast imaging.

The present study indicates that the choice of mammography as a reference standard for breast cancer imaging modalities may require critical review. In terms of detection of actual lesions, it is only superior to ultrasound when microcalcification is present, and this shortcoming of sonographical diagnosis is well known (3,12,13). On the other hand, the automated ultrasound scanning led to the correction of three BI-RADS IVa conference decisions, therefore contributing significantly to the crucial avoidance of overdiagnosis and over-therapy.

However, we also confirmed another shortcoming of ultrasound imaging of the breast, namely the relatively high percentage of false positive results (3,14,15) that currently disqualifies sonography from being a first-line screening method. Considering the results in their entirety, they suggest that the sequence of mammography and ultrasound imaging as per the current screening standard makes perfect sense, and that automated 3D ultrasound scanning is probably equivalent to conventional hand-held imaging in terms of diagnostic accuracy. The latter conclusion is shared by Wenkel et al. (10) based on a recently published comparative study of hand-held vs. automated sonography and confirms previous studies with similar results (16, 17).

Both the aforementioned pitfalls are also not exclusive to automated 3D ultrasound, but affect manual hand-held ultrasonography in equal measure (3, 12-15). Therefore there are presently no evidence-based reasons to favor one ultrasound modality over the other from a diagnostic point of view.

Obviously, the achievement of high diagnostic accuracy in ultrasound imaging requires a high degree of technical and medical proficiency from the examiner in charge. Whereas this generally applies to both methods, the automated 3D scan is less prone to examiner-related errors for two reasons: Firstly, the close adherence to a strict diagnostic protocol –
Figure 2. Mammographic and sonographic findings in patient #35.

Figure 3. Mammographic and sonographic findings in patient #165.
which also yields a high accuracy when applied to hand-held ultrasound (18) – is uncoupled from the examiner and ensured automatically; secondly, off-line examination allows for unlimited careful re-runs and the consultation of senior specialists independently of their geographical location.

To date, there is certainly a broader base of diagnostic breast centers with expertise and experience in hand-held ultrasound imaging of the breast than in automated 3D imaging since the latter has only been available for about three years (1, 3, 8-10), but this cannot justify an unquestioned preference for the former.

**Conclusion**

The impact of possible diagnostic advantages of automated ultrasound imaging such as the ‘bird’s eye’ view, better volume appreciation and whole breast coverage cannot be assessed at the moment, and certainly not based on the results of the present study. An evidence-based assessment of this issue would require a comparative study setting in which both methods being employed, either on the same patients or in a randomized controlled study, and positive imaging results are verified or falsified by histological or cytological examination.

In full appreciation of the effort involved in such a trial, we consider it well justified based on the – doubtlessly preliminary, but nevertheless suggestive – results of the present study.

**Acknowledgements**

Professor Per Skaane MD (Department of Radiology, Breast Imaging Center, Ullevaal University Hospital, Kirkeveien 166, N-0407 Oslo, Norway) reviewed each case of disagreement between two investigators with regard to BI-RADS classification. Hartmut Buhck, M.D., provided editorial advice and assistance in statistical methods being employed, either on the same patients or in a randomized controlled study, and positive imaging results are verified or falsified by histological or cytological examination.

**References**


Received March 16, 2011
Revised June 21, 2011
Accepted June 21, 2011