Differences in the B-mode imaging quality of ultrasound devices in the mid-price segment

Daniel Merkel1,2, Ebaa Albarri1, Masuod Yousefzada1, Marko Böhm1, Michael Gottwald1, Christoph Schneider1

1Immanuel Klinik Ruedersdorf, University Hospital of the Brandenburg Medical School, Ruedersdorf b. Berlin, 2BIKUS – Institute for Clinical Ultrasound, Brandenburg Medical School Theodor Fontane (MHB), Neuruppin, Germany

Abstract

Aims: A meaningful sonographic examination is decisively dependent on the B-scan quality of the ultrasound device. When selecting a suitable ultrasound device, B-scan quality should be an important purchase criterion. Although there is no generally accepted method to measure B-scan quality, we tried to evaluate comparable sonography devices from different manufacturers regarding B-scan quality. Material and methods: We systematically assessed the B-scan quality in ultrasound devices of seven different manufacturers from the mid-price segment. All 7 ultrasound units tested had comparable equipment features and the purchase value of approximately $20,000. We recorded video sequences and compared B-mode image quality. We used both physiological sectional images and pathological findings from abdominal ultrasound. Results: We identified three ultrasound units that scored significantly better in measuring the B-scan quality than the other devices. The Canon Xario 200, the General Electric Logiq P7 and the Mindray DC70 (in alphabetical order) were the units that outperformed all others. The differences identified were found to be statistically significant. A subgroup analysis showed that the contrasts in quality were more pronounced in near-field examinations than in examinations with greater penetration depth. Conclusions: There are considerable qualitative discrepancies in B-scan ultrasound devices despite being similar in terms of equipment and price. Our findings show that these differences are statistically detectable and likely clinically relevant. Keywords: B-scan; sonography; quality; comparison; ultrasound; device

Introduction

In Europe, there is a large and complex market for ultrasound (US) medical equipment, with an extensive price range for stand-alone devices [1]. This assists greatly in a well-informed decision about which device best serves outpatient departments, hospitals, e.g., emergency rooms and ICUs difficult.

A previous study of high-end sonography devices by our research group [2] revealed significant differences in the B-scan quality between varying manufacturers. With increasing cost pressure for health care services [3] there is an increased demand for sonography devices from a lower price segment in hospitals as well as in practices (Market Research Reports, New York City, U.S.A., www.marketresearch.biz). If the B-scan quality is assumed to be a decisive criterion for the choice of a new device [4], the objective selection of a specific device manufacturer is very difficult due to a lack of comparative data. The aim of this investigation is to identify differences in the B-scan quality of devices from varying manufacturers. With this investigation we would like to create a ranking of comparable devices from the same price segment, which can be helpful in the choice of a suitable sonography device.

Sufficient quality of the B-scan image is decisive for meaningful ultrasound (US) abdominal examination. High technical demands are placed upon US devices,
particularly in cases where evidence of small parenchymal lesions (e.g., in the liver) is required. Many factors play a role in the selection of an appropriate US device (equipment, purpose of use, budget, availability, presence of company representatives, habits, etc.). The majority of sonographic questions in general and internal medicine are answered in B-scan mode, making the quality of the B-scan the utmost priority [4].

Objective acquisition of a B-mode image of high quality is difficult and not standardized. In this study, B-scan images of US devices from different manufacturers are evaluated in direct visual comparison. Our study is limited to the examination of the B-scan quality and does not cover other criteria such as the quality of different probes, duplex sonography, contrast enhanced US and others.

**Materials and methods**

Between May and August of 2022, we tested the B-mode image quality of ultrasound equipment from 7 competing equipment manufacturers (Table I). We used the latest mid-priced devices with a list price between 15 and 20 thousand dollars. Under consideration were stand-alone devices with at least 2 or more different sonographic probes (at least one convex and one linear transducer). The devices had to be equipped with tissue harmonic imaging (THI), duplex sonography and DICOM connection. A final qualifying factor was that the equipment be available in Germany.

The study protocol was reviewed and approved by the local Ethics Committee of the Medical School Theodor Fontane Brandenburg (reference number E-01-20220502). Study participants gave written informed consent as required by the study protocol.

Figure 1 illustrates the study design: three healthy subjects (mean age 48 years, mean body mass index 29 kg/m²) received US examinations in 2 defined abdominal standard sections using the 7 different sonography devices, in the same session. These sections included four sections with normal findings and two sections with pathological findings (details in Table II). Images with normal findings simulate a routine screening of the specified US sections. When recording sections with pathological findings, explicit attention is given to the display of lesions found.

All examinations were carried out by the same examiner, who held the highest level of qualifications (stage III) according to the DEGUM (German Society for Ultrasound in Medicine).
Differences in the B-mode imaging quality of US devices in the mid-price segment

Image recording was done in THI mode, image optimization was carried out in each examination regarding gain, dynamic range, penetration depth and focus. To ensure an optimal device setting, a technical assistant from the respective company was present during the examinations.

The documentation was carried out digitally by recording 10 second video sequences. The 42 video sequences (six sections times seven devices) produced were anonymized in digital post-processing, concealing the make of the manufacturer. Finally, two corresponding sequences were placed next to one another. The two recorded sequences compared the same standard section of the same subject, using the 7 different US devices (fig 2).

The 132 pairs of corresponding sequences thus created were compiled in random order as an online interview and evaluated by experienced US examiners regarding the image quality in a direct visual comparison. The quality criteria were the overall image quality, good spatial resolution, and the gray-scale contrast. The evaluation of the two images of the pathological findings was not about the detection of the findings, but the image quality of the 2 pathological findings. The evaluators were able to quantify the qualitative differences of the overall impression between the two corresponding videos with the aid of a slider (0 points: both videos are the same, shifting the slider to the side of the better-rated video with 3 gradations -1 point: slightly better, 2 points: reasonably better, 3 points: considerably better than the corresponding video). This way, a score can be determined for each device. If both videos are rated the same, both compared devices receive 0 points. If one of the videos performs better than the corresponding video, one to three plus points are awarded for the better device according to the above criteria, and one to three minus points are awarded for the worse device. By adding up the point score of each device, a ranking was generated.

The evaluation was carried out by a total of 16 physicians experienced in sonographic examinations. The inclusion criteria for the evaluators were a minimum five years of experience years of sonography experience and the performance of more than 40 abdominal sonographs per week. The evaluators were specialists in Internal Medicine, Gastroenterology, Nephrology and/or General Medicine. They worked either in one of 6 different hospitals (n=12) or in one of 3 different outpatient examination centers (n=4). Fourteen of the 16 evaluators were qualified according to the guidelines of DEGUM (stage I: n=7; stage II: n=4; stage III: n=3). The remaining evaluators (n=2) qualified by more than 5 years of prevailing work in the sonography laboratory.

By means of this evaluation, a list of manufacturer preference can be created for each examiner. If all points are pulled together, an overall evaluation of the sonography devices becomes possible. A subgroup analysis is carried out as a function of the 6 applied sonographic sections.

Statistical analysis

SPSS software version 22.0 (IBM SPSS Statistics®, New York, USA) was used to analyze the data. For the overall result and the subgroup analysis, the total points achieved by the respective devices were compared with each other. Statistical significance was evaluated using the exact Wilcoxon rank-sum test for continuous variables. All significance tests were bilateral and a p-value of less than 0.05 was considered statistically significant.

Results

The test was able to identify a group of the three best-rated devices (device A, B and C), which differed significantly from the group of the three worst devices. One of the 7 investigated devices was in the midfield (device D) (fig 3).

There were no significant differences in the quality scores within the group of the three best-rated devices. Among these best three devices were the Canon Xario 200, the General Electric Logiq P7 and the Mindray DC70 (in alphabetical order).

A subgroup analysis was carried out for each of the 6 individual sonographic sections by grouping near-field examination (sections 1, 2 and 3), examination with a greater penetration depth (sections 4, 5 and 6) and the two sections with presentation of pathological findings (sections 2 and 3, Table II)
The differences between the devices are more pronounced in near-field examinations than in far-field examinations - in our case, this included the two sections with the pathological findings. The ranking of the devices in the subgroup analysis was almost the same as when considering the entire cohort. Because of the smaller number of cases in the subgroup analysis, the level of significance was not reached in quantifying these differences (Table III).

In Figure 4, we exemplarily depicted the liver hemangioma (section 2) as imaged by 7 different devices. Significant differences in imaging quality are evident here. Detection of this hemangioma with devices D and G is likely to be difficult. We conclude that the differences in B-scan quality identified in our study are clinically relevant.

**Discussions**

The quality of an abdominal US examination depends on many different factors. In addition to examiner-dependent and patient-dependent factors, device quality plays a decisive role. However, the term device quality is generally defined in a vague manner; it can refer to technical imaging accuracy in standardized dummy models or to the subjective impression of the examiner during the clinical application of the US.

Phantom models have been used for many years to measure the imaging quality of B-mode image sonography. Phantom models of this type are commercially available and intended to minimize examiner-dependent and patient-dependent disturbance variables [5,6]. However, no generally accepted phantom could prevail as a standard [7]. While phantoms are accepted for functional control in the case of technical defects [8–10], the subjective overall impression plays a major role in the assessment of the image quality under clinical conditions and cannot be replaced by measurement data determined on phantoms [11]. Technical innovations in the development of B-mode image sonography have been repeatedly investigated on phantoms [12–14]. Irrespective of this, it has long been assumed that the combination of objective and subjective measurement criteria best reflects the image quality of US scanners [11,15].

The present work also relies on the subjective impression in the assessment of the individual videos. In this respect, we do not differ from many previous publications, which evaluate, for example, technical innovations such as tissue harmonic imaging (THI) [16–20], or photopic imaging [21,22]. Similarly, quality examinations of various US systems, such as handheld ultrasound (HHUS) [23] or endosonography devices, rely only on the subjective impression [24], whereby usually several experienced examiners carry out evaluations in parallel.

The literature regularly shows comparisons of sonography devices of different sizes and quality classes. In particular, several studies have been carried out to assess the portable devices known as HHUS systems. More than 40 years ago, HHUS has been described and compared in

<table>
<thead>
<tr>
<th>Device</th>
<th>Cumulative score in near field sections</th>
<th>Cumulative score in far field sections</th>
<th>Cumulative score in sections with pathological findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10.06 ± 14.56</td>
<td>3.17 ± 5.69</td>
<td>10.33 ± 8.7</td>
</tr>
<tr>
<td>B</td>
<td>0.94 ± 9.2</td>
<td>3.56 ± 7.12</td>
<td>11.17 ± 8.24</td>
</tr>
<tr>
<td>C</td>
<td>2.00 ± 9.22</td>
<td>4.94 ± 5.06</td>
<td>-1.33 ± 9.55</td>
</tr>
<tr>
<td>D</td>
<td>-6.39 ± 11.02</td>
<td>2.44 ± 7.08</td>
<td>-4.67 ± 11.34</td>
</tr>
<tr>
<td>E</td>
<td>-0.72 ± 9.25</td>
<td>-3.89 ± 7.74</td>
<td>-4.58 ± 10.65</td>
</tr>
<tr>
<td>F</td>
<td>-0.61 ± 8.47</td>
<td>-4.5 ± 4.64</td>
<td>2.42 ± 9.95</td>
</tr>
<tr>
<td>G</td>
<td>-6.69 ± 8.19</td>
<td>-6.17 ± 6.35</td>
<td>-7.96 ± 8.1</td>
</tr>
</tbody>
</table>
everyday clinical practice with the previously established sonography [25,26]. For comparison, either written findings are used [27], still images are compared [23,28], or the devices to be compared are compared directly at the patient’s bedside [29,30]. A meta-analysis has already appeared in the device constellation HHUS versus standing device, which, however, should not sufficiently cover the variety of available HHUS [31].

Comparative studies of US systems from different manufacturers have been carried out less frequently and relate almost exclusively to defined questions. Thus, various endosonography systems are described in the pancreatic assessment [24] and the imaging of contrast-enhanced US (CEUS) in Crohn’s disease patients [32]. There are comparisons with respect to the ergonomics of ultrasonic systems [33] and several comparisons of shear wave elastography (SWE) of different devices [13,34,35]. Since numerical values are determined, in particular, in SWE, the comparison of different systems will be easier to carry out, since the subjectivity in the evaluation of B-mode images is not significant. An extensive comparison of the B-mode image quality in abdominal sonography on high-end devices has been carried out by our own working group [2]. Only recently, a comprehensive comparison of 4 different HHUS appeared in terms of technical equipment, manageability and image quality [4].

The market for sonography equipment is extensive and highly complex [1]. The price for a medical sonography device varies greatly and depends on many factors. The simplest stand-alone devices are available for less than $1000 (e.g., Mindray DP-10). In the lower price range of up to $10,000, handheld US devices (HHUS) from more than 10 competing manufacturers could be predominantly identified. High-end devices more expensive than $60,000 are on the other side of the price scale. Such devices are likely to be difficult to afford for most outpatient facilities in general medicine. The present study compares sonography instruments in the middle price segment. We were advised on this by several commercial equipment dealers and were able to find out that $20,000 net is spent in Germany for a demanding mid-range device.

In our opinion, the B-scan generated by the convex transducer is the most decisive feature in determining the quality of a sonography device for Internal or General Medicine. Further assessment of the linear transducer and duplex sonography would have significantly increased the length of our study. For this reason, we chose to focus solely on the B-scan generated by the convex transducer in our device comparison.

All devices offered in this price segment were additionally equipped with at least one linear transducer and, in addition to a duplex module, also with various tools for B-mode image optimization (THI and others). One of the devices used even included contrast US software (CEUS) for the price mentioned above. All devices also had a so-called cineloop option, a DICOM connection and the option of recording sonography videos.

Devices can also differ in terms of ergonomics [33,36] and monitor quality [37]. The devices we used showed, without exception, a high and intuitive level of user-friendliness as well as appealing monitor quality. However, these criteria were not part of the present comparison and have therefore not been systematically recorded.

Each of the devices used has specific technical possibilities for image optimization, which can have a considerable influence on image quality [38]. The examinations were therefore carried out by an examiner with many years of experience with different US systems. In
addition, the relevant company representatives were pre-
sent during the examinations (application assistant), who
were supposed to ensure an optimal device setting. Image
optimization was carried out in each case with regard to
penetration depth, gain, focus placement, frame rate and
dynamic range. Therefore, disadvantages of individual
devices due to possible operating errors have been mini-
mized.

In addition to age, gender and constitution, there are
various other subject-dependent influences on the imaging
quality of an US examination. By selecting different
sectional planes on different subjects, attempts were
made to map a variety which simulates clinical reality. It
should be noted that our subjects were all normal-weight
and female. This therefore does not necessarily reflect
representatively the whole range of internal medicine
patients.

In contrast to the previous comparative examination
of high-end sonography devices from our working group
[2], we were also able to include pathological findings in
the current examination, which were, in each case, small,
isoechoic benign masses in the liver or in the spleen.
From our point of view, difficult to detect isoechoic parenchymal lesions are especially suitable when com-
paring imaging quality of varying devices. However, it
must be noted that the focus of our study was not on the
detection of pathological findings. The two pathological
findings investigated here have been assessed by contrast
US. They were benign in nature and were known to us
long before the start of our study.

The film sequences juxtaposed in pairs were evalu-
ated by 16 doctors from a total of 9 different inpatient
or outpatient facilities. All evaluators had many years of
experience in US and carry out at least 40 sonographic
examinations per week. Due to the comparatively high
number of evaluating physicians, attempts were made to
minimize the subjectivity in the visual assessment of the
videos.

Visual habits in relation to the sonography system
used in everyday life may play a role in the subjective
perception of image quality; we are not aware of any sys-
tematic investigations into this. Of the evaluating physi-
cians in the present study, an above-average number rou-
tinely work with Canon sonography systems (n=9), but
nobody used sonography devices from Mindray or Sam-
sung. The other 4 equipment manufacturers were repre-
sented 3 or 4 times. In fact, the evaluators who work on
Canon systems in everyday life rated the videos created
by the Canon tester somewhat better in terms of trend
than the other evaluators. However, this difference did
not reach any significance and should therefore have only
a slight influence on the overall result. In this context,
it was important to us to make a complete blinding of
the manufacturer’s information on the evaluation videos.
Thus, an attempt was made to achieve as independent and
thus objective an evaluation of the individual sequences
as possible.

All video sequences were recorded by the same ex-
mainer. Therefore, examiner-dependent factors in video
recording could be minimized. Blinding regarding the
device manufacturer was not possible for this examin-
er. Blinding took place only for the evaluating doctors.
Most of the previous studies used still images for qual-
ity assessment. Of course, these were available in paper
form in older studies [11,20,23]. Other studies compare
written findings with each other, which were collected
with different sonography devices [27,29]. Still other ex-
aminations use digital still images [17,21,39,40] or com-
pare different systems at the bedside on the same patient
(26,41–43) or a previously defined clinical outcome [42],
even if, in the case of the bedside examinations, a blind-
ing of the devices with respect to the manufacturer is not
possible.

We decided to record moving images to have them
assessed by the evaluators. From our point of view,
sonography videos simulate the reality of a sonogra-
phy examination better than still images. The literature
search revealed only a few studies which used sonogra-
phy videos to visually assess the B-mode image quality
[2,22,44,45].

The present study has some limitations. First, all the
investigations were carried out by the same examiner.
While this ensures consistent investigator quality, blind-
ing with respect to the device manufacturer was not pos-
sible. Furthermore, our study was limited to a total of 6
standard sections, including only two of those with path-
ological findings. Here, as well as with the limited num-
ber of subjects, a greater variety would render a more re-
alistic picture. As is known, the imaging quality depends
on a large extent on the device setting. Although the rep-
resentatives of the relevant manufacturers were present
during the investigations and endeavored to optimize the
device setting, it cannot be ruled out that all technical
possibilities were really exhausted during each investiga-
tion. Finally, the uniform purchase price of the devices as
a selection criterion must be mentioned as a limitation.
The selling price is not necessarily directly related to the
value of a device; several other factors, such as market
policy, service quality, market presence and much more,
play a role here.

Artificial intelligence (AI) has also long since found
its way into the field of US imaging. Examples are mod-
ules for assessing liver masses [46], thyroid nodules
[47], lymph nodes in breast cancer patients [48] and a
Differences in the B-mode imaging quality of US devices in the mid-price segment

meta-analysis with various AI systems for assessing liver fibrosis/cirrhosis [49], which also includes abdominal sonography. In our literature search, we have not yet been able to find any publications on AI-based systems for the comparative assessment of B-mode image quality in sonography devices.

Conclusions

There are considerable qualitative discrepancies in B-scan US devices despite being similar in terms of equipment and price. We were able to identify a group of the three best-rated devices which differed significantly from the group of the three worst devices. Our findings show that these differences are statistically detectable and very likely relevant for clinical practice.

Conflict of interest: none

References

27. Seitz K, Vasilakis D, Ziegler M. Efficiency of a Portable B-Scan Ultrasound Device in Comparison to a High-End Ma-


