Quality Assurance of Ultrasound Imagers: Procedures, Expectations, and Philosophies

James M. Kofler, Jr.

INTRODUCTION

This handout discusses several crucial issues regarding quality assurance (QA) in diagnostic ultrasound imaging. First, common misperceptions for not establishing a quality assurance program are addressed, and justification for a QA program is presented. Some of the complicating factors surrounding ultrasound QA are discussed. Second, representative QA measurements are presented, followed by conclusions regarding the clinical usefulness of each QA test, based on the measured data. Lastly, some ideas for future tests are described.

I. A PERSPECTIVE ON THE STATE OF ULTRASOUND QUALITY ASSURANCE

The purpose of quality assurance (QA) testing is to adequately characterize specific performance parameters for imaging equipment. The results are compared to vendor specifications, published values, or previous measurements to determine if the image quality is adequate. In most x-ray modalities, performance tests are well defined and typical values and tolerances are suggested. If the measured values differ significantly from the reference values, corrective action must be taken. While performance testing in x-ray modalities is almost universally practiced, in ultrasound, performance testing remains somewhat controversial in the medical physics community and proliferation of

quality assurance programs in ultrasound has not yet been realized. There are several reasons commonly given by those who oppose initiating quality assurance testing for ultrasound imaging equipment, including the following:

1. Ultrasound is considered to be an established and safe modality; therefore, no quality assurance is necessary.
2. Ultrasound scanners are very stable; therefore, no quality assurance is necessary.
3. No one really knows what to test, how often to test, what measured values to expect, or what equipment to use in an ultrasound quality assurance program. Consequently, performance testing of ultrasound scanners is more trouble than its worth.
4. There are no regulations requiring a well-defined quality assurance program; therefore, allocating the resources and expenses to a quality assurance program is not justified.

The first two of these reasons are most often used by radiologists, the third one by physicists, and the last one by administrators. Each of these reasons is a misconception that could be detrimental to a healthcare institution.

Ultrasound: An Established and Safe Modality?

Consider the common argument, "Ultrasound imaging has been around for about 50 years, no biological effects have ever been shown at diagnostic levels, and nothing ever goes wrong with the machines. No physics support has been used in the past, and no one has ever complained." Now consider this: ultrasound has changed dramatically in just the past few years, and the cost-effectiveness of ultrasound is driving a frenzy of competition

among the ultrasound scanner vendors. Although clinicians and patients ultimately benefit from the competitive efforts of the vendors, new technologies continue to emerge at nearly a blinding rate—challenging physicists and institutions to keep pace. Ultrasound scanners are no longer the simple devices that they are often thought to be. Comparing a modern scanner to the early clinical scanners is like comparing a calculator to a desktop computer. Ultrasound practices are being re-defined as technology provides more echo information and processing options. In this respect, the need for physics support has become twofold: quality assurance to verify that the scanners are operating properly, and technical consultation to assure that the scanners are used appropriately, both of which have implications on patient care and safety.

For most medical professionals, the question of potential biological effects caused by diagnostic ultrasound is almost universally dismissed as being a non-issue. However, in 1993, the FDA relaxed the restrictions on the maximum allowable power outputs of diagnostic medical ultrasound scanners. This was done with the understanding that the vendors would display the output indices on the scanners, implying that the sonographers would assume responsibility for moderating the output levels. Consequently, output power levels have increased. The maximum limits are now in terms of displayed output indices, which depend on the scan protocol, and the user is assumed to understand implications of the indices. Unfortunately, this assumption is not justified, as biological effects are not often considered to be of concern.

So, is diagnostic ultrasound still "safe?" The answer has not yet been proven to be "no," but there is some evidence[1] that suggests that biological effects may occur at diagnostic power levels. Add to this the fact the some of the newer technologies, such as

harmonic imaging, use increased power levels or require multiple acquisitions of the same scan line (or from multiple overlapping or intersecting scan lines). Also consider that scans that need to be repeated or require extra scan time due to poorly operating equipment result in excess patient exposure (not to mention additional resource expenditure). A quality assurance program can minimize scan times, reduce repeat rates, and assure that scan protocols employ proper scanner settings, including using appropriate technologies and lower power levels when possible.

Ultrasound: A Stable Modality?

Unlike the failure of an x-ray tube, there is no expected "life-time" of an ultrasound scanner after which it will shut down and need repair. Some changes in performance may change quickly and obviously, and are usually noted by the operator shortly after they occur. However, other problems may occur over an extended period of time and may be very subtle. One example involves the gradual breakdown of a transducer cable. In this particular case, the institution where this occurred had a single scanner, and as the transducer cable—which probably was defective at installation—was subjected to routine use, the noise in the image became increasingly worse. However, since the change was gradual, and since another scanner was not available for visual comparison, the problem persisted until the image quality was substantially compromised. A QA program could have identified the problem before the degradation in image quality had become significant, possibly even at installation of the scanner.

Some scanner problems may go unnoticed because they are masked or camouflaged by patient anatomy. For example, non-functioning elements on an array transducer may be

quite challenging to notice on a patient image, yet when a uniform field is imaged (a QA test) the artifact is readily apparent. Once the artifact is detected, it is typically surprisingly obvious on the patient images.

Review of a service report for scanners at one institution showed an average of 1.8 hardware and software problems per scanner (a total of 29 scanners) over a one year period. The problems ranged from "bad trackball" and "broken key on keyboard" to "lines in image," and "system hangs on start-up." Although some scanners had no reported errors, others had multiple substantial problems.

For clarification, it should be stated that not every scanner is expected to exhibit regular failures, and major malfunctions are rare. However, ultrasound equipment is not immune to hardware and software problems, and, if such problems go unnoticed or ignored, the impact can range from staff inconvenience to compromised patient diagnosis. A QA program can detect a range of problems, thereby increasing the confidence of diagnosis, reducing patient exposure, and increasing patient throughput. Another advantage of a QA program is that all scanner problems are channeled to a single source and documented. Documentation of all scanner-related problems can be useful in purchasing decisions and for contract (purchasing and service) negotiations.

Testing: What, when, with what, and what to expect?

Regarding ultrasound performance tests, a substantial effort has been made to determine the most pertinent tests, the recommended frequency of testing, the most useful phantom design(s), and acceptable measured values[2-24]. However, the underlying problem in attempting to define these issues is the fact that there are many variables that must be

considered. In fact, the number of potential variables is so large that the conclusions can become generalized to the point of being of questionable utility. This may be the most valid argument against initiating a quality assurance program. For example, consider that there are at least three common transducer types—linear, sector, and curva-linear. Each of these has specific operating and performance characteristics. Now consider that each transducer-type typically has 2 to 4 different available frequencies—again, each with associated performance characteristics. At each frequency, several operator-controlled parameters can be adjusted, including focal zone(s) placement, post-processing, gain, TGCs, power, compression, and maximum imaging depth—each also having associated performance characteristics. Furthermore, the TGC implementation provides only a graphical display of the setting, which makes replicating the setting non-trivial; and the power setting is a relative measure, rendering comparison of the absolute power outputs from scanners of different vendors nearly impossible. All of the mentioned items are used in some combination on clinical images; however, some options (or combinations of scan parameters) are not available on all scanners, and implementation of the scan parameters can vary among vendors (for example, look-up tables are typically labeled as a number, with no standardization among vendors). Now add vendor-specific features that can change the appearance of the image, such as harmonic imaging, temporal and spatial optimization(s), lateral gain, and elevational focusing, to name a few. With all of this in mind, it is indeed daunting to devise a test and define an acceptable range for the result that makes a clinically meaningful statement regarding the performance of the scanner/transducer imaging system.

While US QA at first it may seem an insurmountable challenge, there is a straightforward approach that can be taken to implement a QA program. The program can be thought of as consisting of two separate components—"basic" tests (i.e., those that do not require a phantom), and phantom tests. The basic tests are easy to implement and have been shown to be useful. These tests include hard-copy/processor QA, display settings, visual checks, room lighting, and programmed pre-set consistency.[20, 23] These basic tests address a large percentage of image quality problems that are encountered in ultrasound imaging. However, the basic tests can not detect the most dangerous problems—those that are intimately integrated into the image data.

Phantom tests can provide insight to specific performance characteristics of an ultrasound scanner. For example, resolution tests quantify the resolving capabilities of the scanner. A uniformity test can shed light on numerous scanner malfunctions, and depth of penetration measurements can be related to the scanner's output power. However, not all tests are suitable to performed routinely and not all are necessarily reasonable indicators of *clinical* scanner performance.

For the phantom tests, the scan parameters (including transducer frequency), the performance metric to be measured, and the phantom to be used must be determined. The frequency of testing must also be addressed. The scan parameters should be those that are commonly used in the clinical practice. That is, the physicians and sonographers need to be consulted with to determine the scanner settings (including transducer frequency) that are most frequently used.

The phantom design and composition also need to be considered. The phantom should be composed of a tissue-mimicking material. Other types of materials, such as

polyurethane and zerdine, have been used as phantom materials, but the speed of sound in these phantoms is considerably lower than that of soft-tissue (1430-1450m/s compared to 1540 m/s). Therefore, object sizes and placement within the phantom must compensate for the speed of sound mismatch. However, compensation for the speed of sound mismatch is not adequate for all multi-element transducers and the focusing characteristics of such transducers may not be correctly represented.[24] These concerns do not apply to the tissue-mimicking phantoms, which can be made to match the speed of sound, attenuation, and backscatter properties of soft-tissue.

The size, shape, and object contrast of "targets" within the phantoms determine the purpose of the phantom. Filaments, drawn across the short axis of the phantom, are often used as high-contrast point sources, and cylindrical-shape targets (of various object-contrast materials) are typically used to represent circular objects on the ultrasound display. Ramps, spheres, and cones have also been used in phantoms. Therefore, the selection of objects within the phantom is largely dependent on the performance metric that is to be measured.

Ideally, the performance tests should be sensitive to small changes in the measured parameter, easy to complete, and be related to the *clinical* performance of the scanner. Typically, not all of these features are noted in a single test. Therefore, some compromise is necessary. This manuscript contains a survey of performance tests that were carried out over a one-year period on 23 scanners and 14 transducers. The survey contains representative results, including variability measurements, for many common ultrasound QA tests. The discussion that follows the survey provides insight to tests that are useful for routine QA and for acceptance testing.

Regulations

In x-ray modalities, state and federal regulations require regular performance verification of all equipment[25]. Consequently, administrators, radiologists, technologists, and physicists are aware that resources must be allocated to assure that the equipment is properly maintained. This is not the case with ultrasound equipment. There is no perceived need for ultrasound performance testing and resources are focused elsewhere. In effect, the benefits of an ultrasound quality assurance program are overshadowed by the cost of resources needed to establish and maintain such a program. This is primarily because the benefits are often not as tangible as in other modalities, and assigning a monetary value to them is difficult.

One last issue must be mentioned regarding performance testing and regulation of ultrasound. Ultrasound imaging has been very prolific in the healthcare environment. Scanners are available that represent a range of technologies, and the price for a ultrasound imager can range from under $20,000 to about a quarter of a million dollars. Is it reasonable to think that the image quality produced by all scanners is comparable? Obviously, the answer to this question is “No.” Whether the image quality scanner is “low end” or “state-of-the-art” is irrelevant to how it can be legally used and how much the patient can be charged for the exam.

II. ULTRASOUND PERFORMANCE MEASUREMENTS

A listing and a detailed description of the tests and procedures that have been suggested for quantifying ultrasound scanner performance is summarized by an AIUM publication[11] and by Goodsitt, et. al.[11] and will not provided in detail in this document. A study of over 2,000 measurement values is presented and the utility of several common ultrasound QA tests are discussed. The data provide representative ultrasound quality assurance values from a large number of ultrasound scanner configurations from a single vendor. Statistical calculations regarding the ranges of measured values are presented for each unique scanner configuration and comparisons with suggested values and tolerances are presented where applicable. The results can be used as reference data for determining appropriate tests, test procedures, and tolerance values when establishing and maintaining and ultrasound quality assurance program.

A Survey of Ultrasound Quality Assurance Measurements

Ultrasound QA tests were performed on 23 different scanners (model 128XP, Acuson Corp., Mountain View, CA) using 14 unique transducer models. The number of transducers of each model ranged from 3 to 23. Of the 14 transducer models, nine could be used at more than one frequency (multi-Hertz), resulting in 24 unique transducer/frequency combinations and 285 unique scanner/transducer/frequency combinations (scan configuration). Each combination was evaluated with a single set of six QA tests (one measurement for each test), yielding a total of 2,231 measurements. Additionally, a set of reference repeatability measurements were recorded for three different scanner/transducer/frequency combinations. A total of 12 measurements of each

test, over a four-week period, were performed for each of the three repeatability scan configurations. All measurements were completed by a single individual with over seven years of dedicated ultrasound quality assurance experience.

The QA tests incorporated a general purpose tissue-mimicking phantom and a slice thickness phantom (models 403 and 421, respectively, Gammex/RMI, Middleton, WI). The scan parameters were set as follows: Power: 0%, gain: 0%, pre-processing: 1, persistence: 3, and post-processing: 0.

The time-gain compensation controls (TGCs) were set using computer-assisted feedback to assure that fibers depictions were displayed at gray scale levels within 10 pixel values of each other throughout the field of view (axially) and for each scan configuration—thereby assuring TGC repeatability and image uniformity.

The QA tests that were performed are described as follows.

Depth of Penetration (DOP). The depth of the field-of-view (FOV) was set to maximum. The depth (to the nearest centimeter) at which the presence of electronic noise was comparable to echo signal was recorded as the DOP.

Vertical and Horizontal Distance Accuracy. The electronic scanner calipers were placed in the centers of the images of two vertical fibers. The distance displayed on the scanner was recorded as the vertical distance accuracy (VDA). The horizontal distance accuracy (HDA) was recorded in a similar manner using horizontal fibers. The fibers for both measurements were selected such that they were separated by the maximum distance allowed by the FOV. VDA and HDA measurements for the reference data were performed using electronic calipers on a spatially calibrated frame-grabbed image.

Axial Resolution. The set of phantom fibers dedicated for measuring axial resolution (AR) was used to record the AR. The AR was recorded as the separation of the two most closely spaced fibers that did not present an overlap in their depiction on the displayed image. AR measurements were recorded for all possible depths (within the FOV) allowed by the phantom (3, 8, and 14 cm). The FOV was magnified ("zoomed") to encompass the axial resolution fibers for each depth of interest. Magnified FOV was not used for the repeatability data. The focal zone was placed at the fiber depth in all cases. Note that the phantom design allows only for discrete values of 2, 1, 0.5, and 0.25 mm for axial resolution.

Lateral Resolution. Lateral resolution was measured using a frame-grabber (model LG-3, Scion, Corp., Frederick, MD) and personal computer (Powerbook DuoDock II, Apple Computer, Cupertino, CA). The lateral resolution was measured as the full-width-at-half-maximum (FWHM) of the fiber depictions. The lateral resolution was recorded for three depths (if allowable by the FOV). The focal zone was placed slightly below the depth of the fiber of interest. (Note: For the repeatability data, the focal zone was placed at the maximum depth regardless of the depth of the fiber.)

Ring-Down (Dead Zone). Ring-down, or dead zone, depth measurements were recorded using the dedicated ring-down fibers on the phantom. The criteria that the fiber be visually resolved was used to determine the ring-down depth. Note that the phantom design allows only for discrete values of 1, 4, 7, and 10 mm for the ring-down measurement. The focal zone was located at the maximum depth for all measurements of ring-down depth.

Slice Thickness. Using the slice thickness phantom, the minimum depth range of the horizontal band (the image of the slice thickness "ramp") was recorded as the slice thickness. That is, the transducer was traversed across the phantom and the narrowest ramp depiction was determined. The endpoints for the depth range were determined visually.

Throughout the one-year period of data acquisition, all tested scanners produced clinically acceptable images, were subjected to additional QA measurements (such as "non-phantom" tests and subjective visual assessments), and were checked periodically by the vendor as part of a preventive maintenance program.

Results and Discussion

To best demonstrate the variability in US QA measurements, all of the data were considered valid (i.e., no "flyers" were removed). Note that due to differences in data acquisition, the reference data are shown for comparison with regard to variability and were not included in the analyses of the remaining data. The key for the plots is given in Figure 1. All tables and figures are included at the end of this manuscript.

Depth of Penetration (Table I). Lower frequency transducers typically showed a greater depth of penetration, as expected. However, this was not always true with transducers that are capable of transmitting more than one nominal frequency.

The depth of penetration (DOP) measurements showed no variation for any of the scanner/transducer/frequency combinations evaluated. The lack of variability in the DOP

measurement is most likely due to the method of measurement. Namely, that the measured DOP is rounded to the nearest centimeter. The measured values do not contradict the published suggested tolerances of ±1 cm for the DOP.

Distance Accuracy (Figures 2 and 3). The error associated with the vertical distance accuracy (VDA) is typically between 1-2% and, in all cases, the measured distance was less than the known distance. The average standard deviation of the VDA is approximately 0.5%, which is consistent with the average standard deviation of the reference data. One possible explanation for a consistently negative error in the VDA measurement is that the speed of sound in the phantom is slightly greater than 1540 m/s. The manufacturer specification for the speed of sound in the phantom is 1540 ±10 m/s, which could introduce a maximum error of ±1% in the VDA. 278 of the 285 measurements showed errors less than 2%, suggesting that a 2% tolerance limit is reasonable.

The average error of the horizontal distance accuracy (HDA) was less than 1%. The average standard deviation of the HDA among all transducers is approximately 0.7%, which is comparable to the average standard deviation of the reference data. Of the 285 HDA measurements, only 14 showed an error greater than 2.5%, 9 of which were of the curva-linear scan format. These results may also be biased by the potential speed of sound error noted with the vertical distance accuracy. A tolerance value of 2.5% seems reasonable for the HDA measurement.

Axial Resolution (Table II). General trends in the axial resolution (AR) measure are consistent with expectations. Namely, AR is superior with higher frequency transducers and degrades slightly with depth.

The data show that few transducer types yielded identical AR measurements for all transducers. However, due to the discreteness and subjectiveness of the measurement, it is difficult to draw conclusions regarding the source of variability. The reference data (most notably with the L538 transducer) imply that the subjectiveness plays a major role in the variability. Therefore, a change of ±1 fiber would have to be acceptable. However, this range includes almost all of the fibers, and a change of more than 1 fiber would essentially never occur. The conclusion can be made that the method of measuring axial resolution is inadequate to obtain meaningful data for routine QA purposes. Other methods have been suggested[20] and may be of more utility.

Lateral Resolution (Figure 4). General trends show superior lateral resolution (LR) with higher frequencies, as expected. No correlation of the LR measurement or variability of the measurement with scan format is evident, even with consideration for depth. However, the LR is generally worse at the deeper depths, as expected.

The standard deviation is typically between 0.1 and 0.5 mm, indicating that a reasonable tolerance for LR measurements is ±0.5 mm for most scan configurations when using computer-assisted analysis. One would expect that extrapolating the results to account for variability caused by visual analyses (using the electronic calipers of the scanner) would most likely increase the tolerances. However, computer-assisted analysis is susceptible to local "disruptions" of pixel values near the fibers caused by speckle.

These variations are somewhat compensated for by the human visual system. A previous study[26] shows that the standard deviation of lateral resolution as measured via scanner calipers is comparable to that shown in this study. Therefore, for most scan configurations, a tolerance of ±0.5 mm may be of utility for LR measurements that are made using the electronic calipers on the scanner. However, as with all QA tests, additional measurements should be acquired if the results are not within tolerance. Note that the results show that under some circumstances (e.g., the V4 transducer at the deepest tested depths), a tolerance of ±0.5 mm would not be reasonable.

With the exception of a few transducer models (V4 at 2.5 MHz and at 3.5 MHz, S228, L382), the standard deviation varies by no more than 0.1 mm among the LR measurements recorded at different depths. Therefore, the depth at which the measurement is performed does not influence the stability of the measurement, and a single depth may be used for routine LR measurements as suggested in references 20 and 23.

Ring-Down, or Dead Zone (Table III) The measurements of the ring-down depth show consistent results for most transducers—the variability being more prominent with the low frequency transducers. The lack of variability is most likely caused by the discrete nature of the measurement. Namely, only two (1 and 4 mm) of the four available depths were needed to measure the ring-down depths for all scan configurations. Nearly all of the transducers were able to consistently resolve the fiber at shallowest depth, suggesting that the test does not adequately stress the imaging capabilities of the scanners. Finer increments (and starting at a shallower depth) of the fiber placement would produce more

detailed information regarding the ring-down characteristics of a particular scan configuration but may not be of clinical utility. The data does not contradict the tolerance suggested by the manufacturer of the phantom[17].

Slice Thickness (Figure 5). The slice thickness measurements showed an expected trend of decreasing values with increasing transmitted frequencies. Additionally, the slice thickness for the linear format transducers was generally smaller than that from other scan format transducers.

With the exception of two transducer types (C544 at 3.5 MHz and S228), the standard deviations were less than 0.5 mm. This suggests that a reasonable tolerance for slice thickness measurements is ±0.5 mm for most transducers.

Comparison to Published Values

Table IV shows a comparison of the suggested tolerances from several references to those concluded from the measured values. In many cases, such as depth of penetration, axial resolution, axial resolution, and ring-down, the method for performing the measurement limits the sensitivity of the measurement. For example, all identical transducers (at equal frequencies) showed the same depth of penetration. This is due to the fact that the depth of penetration is measured to the nearest centimeter. Therefore, a tolerance level can be stated as ± 1 cm, but it must be acknowledged that this tolerance is imposed based on the sensitivity of the measurement, and a smaller tolerance may be more appropriate if the measurement technique were improved.

III. CONCLUSIONS ON THE CLINICAL UTILITY OF PHANTOM TESTS

A large number of quality assurance measurements and suggested tolerances have been presented. These data provide insight to the expected values and variability for a specific set of scanner from a single vendor. However, the question remains as to whether the tests are clinically relevant. Personal experience indicates the following:

Resolution tests are reasonable for acceptance testing but are of little utility for routine performance checks. It is always important to verify vendors' claim when a unit is installed. However, a significant change in a resolution value would indicate that there is serious malfunction with the scanner and mostly likely other major (and obvious) image performance problems would be noticed. This is stated as speculation because in six years of involvement in ultrasound quality assurance, I have never noted a significant degradation in axial, lateral, or elevational resolution.

Ring-down testing is irrelevant with most modern day scanners. I suspect that some older scanners could potentially have objectionable ring-down, but in most cases, the ring-down is clinically irrelevant. The proximal 4 mm just isn't of interest in most clinical exams (and note that ring-down was only measured as 4 mm for a few transducers, all of which were operated at a frequency of less the 3.5 MHz—these transducers are *not* used for imaging superficial structures).

Depth of penetration is a useful test for measuring a decrease in the power output of the scanner or any increase in noise (such as from a defective transducer cable, etc.). Both of these can produce very subtle image degradation that can be very difficult to detect by non-invasive methods (with the exception, of course, of the depth of penetration test).

Both require baseline data, which requires that a depth of penetration measurement be performed and archived (via a hard-copy or electronically) when the scanner is installed.

Distance accuracy measurements are also important tests. There is little uncertainty in the measurement and the expected value is known. Important clinical decisions are made based on distance measurements, and verification of the electronic calipers is essential. Note, however, that there is dependence on the speed of sound in the phantom material for the distance accuracy measurement and, as seen in the survey data, phantom manufacturing tolerances should be considered. Ideally, phantom-specific speed of sound values should be obtained, if available.

A uniformity test, which was not evaluated in the survey, is perhaps the most useful phantom test. Many scanner malfunctions are manifested as some anomaly in the image and are immediately apparent in a uniformity image. The test is quick, simple, and of unquestionable clinical relevance. A hard-copy of a uniformity phantom image also should obtained at the time the scanner is installed.

IV. POTENTIAL ULTRASOUND QA TESTS

There are many potential developments that could benefit ultrasound quality assurance efforts. A novel image quality test is currently under development that involves size-dependent signal-to-noise ratio measurements of simulated spherical lesions in a perfect array within a tissue-mimicking phantom.[22, 27, 28] The importance of the technique is that uses spherical test objects, which equally consider axial, lateral, and elevational resolution. Spheres of different diameters and object contrasts can be used, covering a

range of imaging situations. Additionally, the measurements are correlated with the impressions of trained human observers, thereby directly addressing clinical relevance.

Another major "breakthrough" in ultrasound QA would be the integration of some tests into the scanner software. This would make the tests very practical, and quick checks of the scanner could be performed regularly (such as daily or weekly). I am quite surprised that no vendors currently offer such features. "Features" of this nature could be promoted as an indication of the manufactures' commitment to quality scanning.

"First-scan" testing is also a noteworthy goal. The test would be a quick and simple visual check, such as a uniformity test, that would be completed every day prior to scanning patients. Scanner problems could be detected before any clinical images are obtained, thereby assuring the institution that the images are free of artifacts caused by equipment malfunctions. A test of this nature would most likely prove its worthiness to the department after only a few scanner problems have been detected.

V. CONCLUSIONS

Ultrasound scanners are extremely complex imaging devices and developing clinically relevant performance tests is a challenge. Consequently, ultrasound quality assurance has been in a state of evolution for many years. Today, a convergence of philosophies, combined with gained experience, is beginning to be noted. It remains our responsibility as scientists to promote and improve the practice of ultrasound quality assurance testing. As interest and enthusiasm increases in the physics community, refinement and proliferation of quality assurance tests will continue. With proliferation of testing, the benefits of performance testing will become more widespread and will eventually lead to a standard

practice of quality assurance in ultrasound departments. As professionals in the healthcare environment, our ultimate goal is to improve patient care, and assuring that adequate and appropriate diagnostic ultrasound image quality is achieved and maintained directly serves that purpose.

ACKNOWLEDGEMENTS

I would like to thank Debra Groth, for her assistance in acquiring the data for this manuscript.

REFERENCES

1. Holland, C., et al., *Direct evidence of cavitation in vivo from diagnostic ultrasound.* Ultrasound Med. Biol., 1996. **7**(22): p. 917-925.
2. Brendel, K., et al., *Methods of measuring the performance of ultrasonic pulse-echo diagnostic equipment.* Ultrasound in Med & Bio, 1976. **2**: p. 343-350.
3. Carson, P.L., *Rapid Evaluation of Many Pulse Echo System Characteristics by Use of a Triggered Pulse Burst Generator with Exponential Decay.* J. Clin. Ultrasound, 1976. **4**(4).
4. Carson, P. and G. Dubuque, *Ultrasound Equipment Quality Control Procedures, CRP Report No. A3*. 1979.
5. Carson, P.L. and J.A. Zagzebski, *Pulse Echo Ultrasound Imaging Systems: Performance Tests and Criteria*. 1980, AAPM: College Park, MD.
6. Goldstein, A., *Quality Assurance in Diagnostic Ultrasound—a Manual for the Clinical User*. 1980, HHS Publications, FDA.
7. Smith, S.W. and H. Lopez, *A contrast-detail analysis of diagnostic ultrasound imaging.* Med. Phys., 1982. **9**(1): p. 4-12.

8. Banjavic, R.A., *Design and Maintenance of a Quality Assurance Program for Diagnostic Ultrasound Equipment.* Seminars in Ultrasound, 1983. **4**(1).
9. Donofrio, N.M., et al., *Investigating the Efficacy of Current Quality Assurance Performance Tests in Diagnostic Ultrasound.* J. Clin. Ultrasound, 1984. **12**: p. 251-260.
10. Goldstein, A., *Slice thickness measurements.* J. Ultrasound Med, 1988. **7**: p. 487-498.
11. AIUM, *Methods for measuring performance of ultrasound scanners: American Institute of Ultrasound in Medicine Standard*. 1990, American Institute of Ultrasound in Medicine: Bethesda, MD.
12. Lopez, H., et al., *A clinical evaluation of contrast-detail analysis for ultrasound images.* Med. Phys, 1990. **17**(1).
13. NCRP, *Quality Assurance for Diagnostic Imaging, NCRP Report No. 99*. 1990, National Council on Radiation Protection: Bethseda.
14. Lopez, H., M.H. Loew, and D.J. Goodenough, *Objective analysis of ultrasound images by use of a computational observer.* IEEE Transactions on Medical Imaging, 1992. **2**(4): p. 496-506.
15. Metcalfe, S.C. and J.A. Evans, *A Study of the Relationship Between Routine Ultrasound Quality Assurance Parameters and Subjective Operator Image Assessment.* Brit. J. Radiology, 1992. **65**: p. 570-575.
16. Metcalfe, S.C. and J.A. Evans, *Optimization of Ultrasound Scanner Characteristics: A Preliminary Study.* Brit. J. Radiology, 1993. **66**: p. 609-613.

17. Gammex/RMI, *The QC Cookbook for Ultrasound*. 1994, Middleton, WI: Gammex/RMI.
18. Hoskins, P.R. and W.N. McDicken, *Techniques for the Assessment of the Imaging Characteristics of Intravascular Ultrasound Scanners.* Brit. J. Rad, 1994. **67**: p. 695-700.
19. Madsen, E.L., et al., *Performance Testing of Transrectal US Scanners.* Radiology, 1994. **190**: p. 77-80.
20. AIUM Technical Standards Committee (EL Madsen, Chair, *Quality Control Manual for Gray-Scale Ultrasound Scanners–Stage 2*. 1995, American Institute of Ultrasound in Medicine: Laurel, MD.
21. Carson, P.L. and M.M. Goodsitt. *Pulse Echo Acceptance and Quality Control Testing, proceedings of the*. in *AAPM 1995 Summer School on CT and Ultrasound: Current Technology and Applications of Computed Tomography and Ultrasound*. Advanced Medical Publishing, Madison, WI 1995.
22. Rownd, J.J., et al., *Phantoms and automated system for testing resolution of ultrasound scanners.* Ultrasound Med. Bio, 1996(Submitted Oct.).
23. Goodsitt, M., et al., *Real-Time Quality Control Test Procedures.* Med. Phys., 1998(25): p. 1385-1406.
24. Goldstein, A. *The effect of urethane rubber acoustic velocity on Quality Assurance Measurements*. *AIUM 2000 Annual convention*. 2000. San Francisco, CA: American Institute of Ultrasound in Medicine.
25. 21 CFR part.1020, *Regulations for the Administration and Enforcement of the Radiation Control for Health and Safety Act of 1968, part 1020:Performance*

Standards for Ionizing Radiation Emitting Products, in *Code of Federal Regulations*. 1993.

26. Kanal, K., J. Kofler, and D. Groth, *Comparison of selected ultrasound performance tests with varying overall receiver gain and dynamic range using conventional and magnified field of view.* Med. Phys., 1998. **25**: p. 642-647.
27. Kofler, J., *Quantification of Diagnostic Ultrasound Scanner Performance as Perceived by Human Observers (tentative)*, Doctorate Thesis, University of Wisconsin– Madison, 2000 (tentative).
28. Wharton, D.B. and E.L. Madsen, *Refinement of an automated lesion detection algorithm to better mimic human observers.* (in progress).

TABLES AND FIGURES

Scan Format	Linear									
Model	L382	L5		L538		L558	L7		L7384	
MHz	3.5	3.5	5.0	5.0	5.0	5.0	5.0	7.0	7.0	7.0*
Sample Size	23	16	16	11	Ref	18	22	22	5	Ref
Measurement (cm)	18	8†	8†	8†	8†	8	8†	6	6	6

Scan Format	Curva-Linear					
Model	C366		C544		C7	
MHz	2.5	3.5	3.5	5.0	5.0	7.0
Sample Size	7	7	18	18	4	4
Measurement (cm)	20	16	16	12	10	8

Scan Format	Sector				Vector						
Model	S228	S5192			V714		V328		V4		
MHz	2.5	3.5	4.0	5.0	5.0	7.0	3.5	3.5	2.5	3.5	4.0
Sample Size	4	7	4	7	3	3	9	Ref	19	19	19
Measurement (cm)	20	14	12	10	7	6	16	16	20	20	16

Ref: Indicates repeatability data. 12 measurements (over a four-week period) using a single transducer at the indicated frequency.

† Indicates maximum allowable depth.

Table I. Depth of penetration measurements. No variability in the depth of penetration was noted for any of the transducer types.

Scan Format	Linear									
Model	1.382	1.5		1.538		1.558	1.7		1.7.384	
MHz	3.5	3.5	5.0	5.0	5.0*	5.0	5.0	7.0	7.0	7.0*
0.5 mm	0	0	0	0	0	0	0	2	0	0
1 mm	19	6	16	1	4	7	22	20	5	10
2 + mm	4	10	0	10	8	11	0	0	0	2

a)

Scan Format	Curva-Linear					
Model	C366		C544		C7	
MHz	2.5	3.5	3.5	5.0	5.0	7.0
0.5 mm	0	0	0	0	0	0
1 mm	2	5	0	9	3	4
2 + mm	5	2	18	9	1	0

b)

Scan Format	Sector				Vector						
Model	S228	S5192			V714		V328		V4		
MHz	2.5	3.5	4.0	5.0	5.0	7.0	3.5	3.5*	2.5	3.5	4.0
0.5 mm	0	0	0	0	0	1	0	0	0	0	0
1 mm	0	0	0	5	3	2	1	0	2	4	3
2 + mm	4	7	4	2	0	0	8	12	17	15	16

c)

* Indicates repeatability data. 12 measurements (over a four-week period) using a single transducer at the indicated frequency.

Table II. Axial resolution measurements. Data was acquired using three different depths (when possible), however, only the data acquired using depth of 3 cm is shown. The axial resolution

values are listed on the left and the number of occurrences of the listed resolution is stated for each transducer type. a) Linear-array scan heads; b) Curva-linear transducers; c) Sector and vector format scan heads.

Scan Format	Linear									
Model	L.382	L.5		L.538		L.558	L.7		L.7384	
MHz	3.5	3.5	5.0	5.0	5.0*	5.0	5.0	7.0	7.0	7.0*
1 mm	12	16	16	11	12	18	22	22	5	12
4 mm	11	0	0	0	0	0	0	0	0	0

a)

Scan Format	Curva-Linear					
Model	C366		C544		C7	
MHz	2.5	3.5	3.5	5.0	5.0	7.0
1 mm	1	1	16	18	4	4
4 mm	6	6	2	0	0	0

b)

Scan Format	Sector				Vector						
Model	S228	S5192			V714		V328		V4		
MHz	2.5	3.5	4.0	5.0	5.0	7.0	3.5	3.5*	2.5	3.5	4.0
1 mm	0	7	4	7	3	3	9	12	5	16	17
4 mm	4	0	0	0	0	0	0	0	14	3	2

c)

* Indicates repeatability data. 12 measurements (over a four-week period) using a single transducer at the indicated frequency.

Table III. Ring-down, or dead-zone, measurements. The measurement values are listed on the left and the number of occurrences of the listed ring-down measurement is stated for each transducer type. a) Linear-array scan heads; b) Curva-linear transducers; c) Sector and vector format scan heads.

QUALITY ASSURANCE TEST	SUGGESTED TOLERANCES*				
	Ref. 20	Ref. 17	Ref. 23	Ref. 13	Indicated in Results Section
Depth of Penetration	± 1 cm	± 1 cm	± 1 cm	± 1 cm	±1 cm
Distance Accuracy Vertical	±2%	±1 %	±1.5% or 1.5 mm	±1 %	±2%
Distance Accuracy Horizontal	±3%	±2% or 2 mm	±2% or 2 mm	±3%	±2.5%
Axial Resolution	0	CDB**	1 mm (>4 MHz), 2 mm (<4 MHz)	-	N/A†
Lateral Resolution	0 mm	±1 mm	>1 mm or > Equation 1††	-	±0.5 mm
Ring-Down (Dead Zone)	-	CDB**	≥7 mm (≤3 MHz) ≥5 mm ((3-7 MHz) ≥3 mm (≥7 MHz)	-	CDB**
Slice Thickness	-	-	-	-	±0.5 mm

* Relative to baseline values.

**Consistent measurable Difference from Baseline value.

† Measurement method (using axial resolution fibers) deemed inadequate for routine QA purposes.

††Equation 1: >2.5 x focal length/[MHz x aperature diameter (mm)]

Table IV. A comparison of published suggested tolerances with tolerances derived from the results of this study.

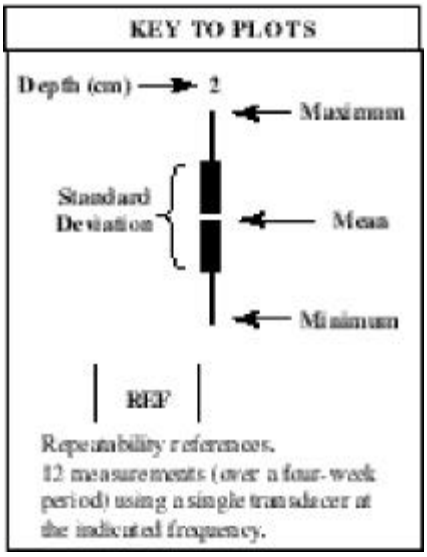

Figure 1. Key for plots.

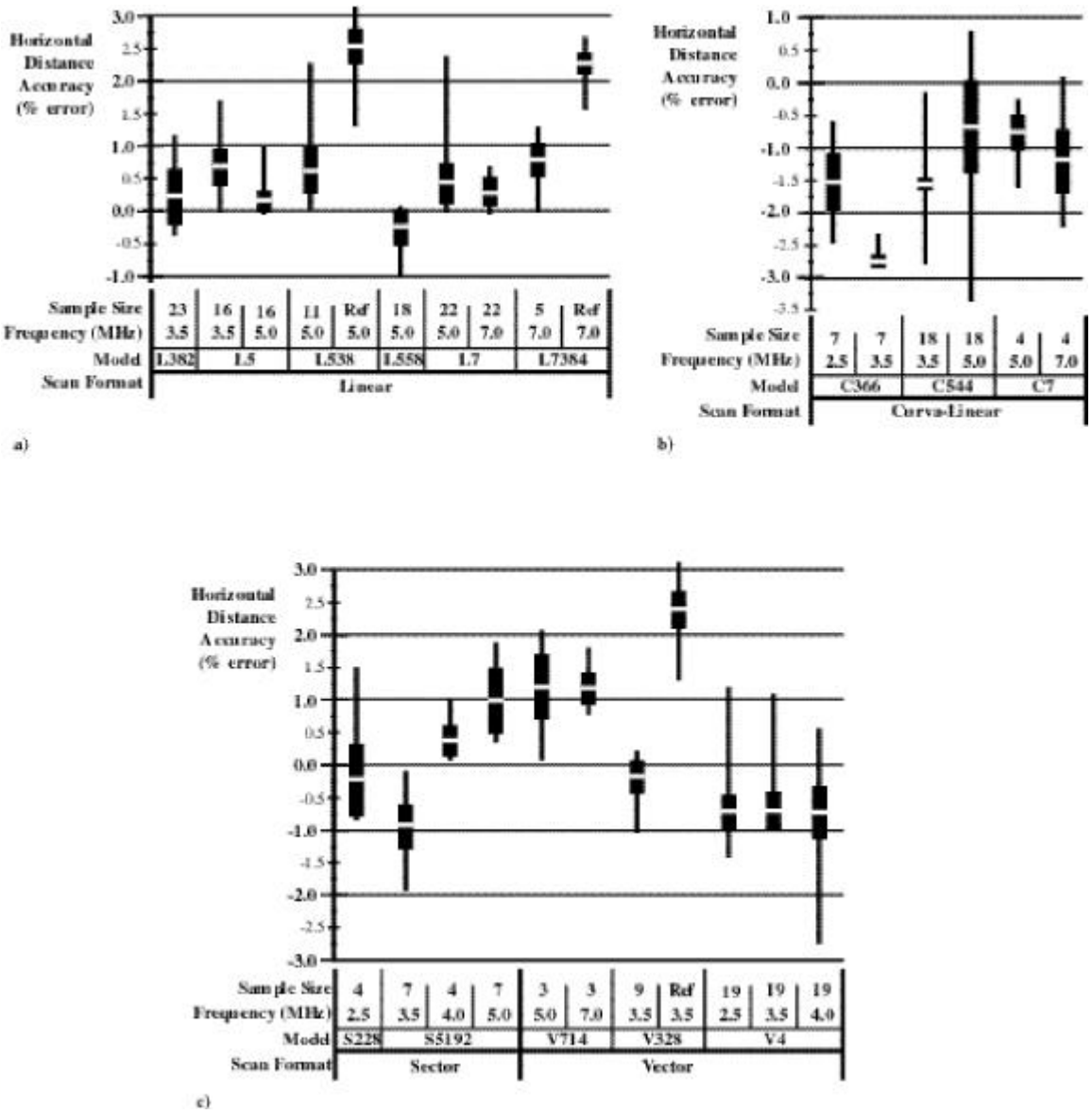

Figure 2. Horizontal distance accuracy errors. Note the slightly larger negative errors associated with the curva-linear transducers as compared to the remaining transducer types. a) Linear-array scan heads; b) Curva-linear transducers; c) Sector and vector format scan heads.

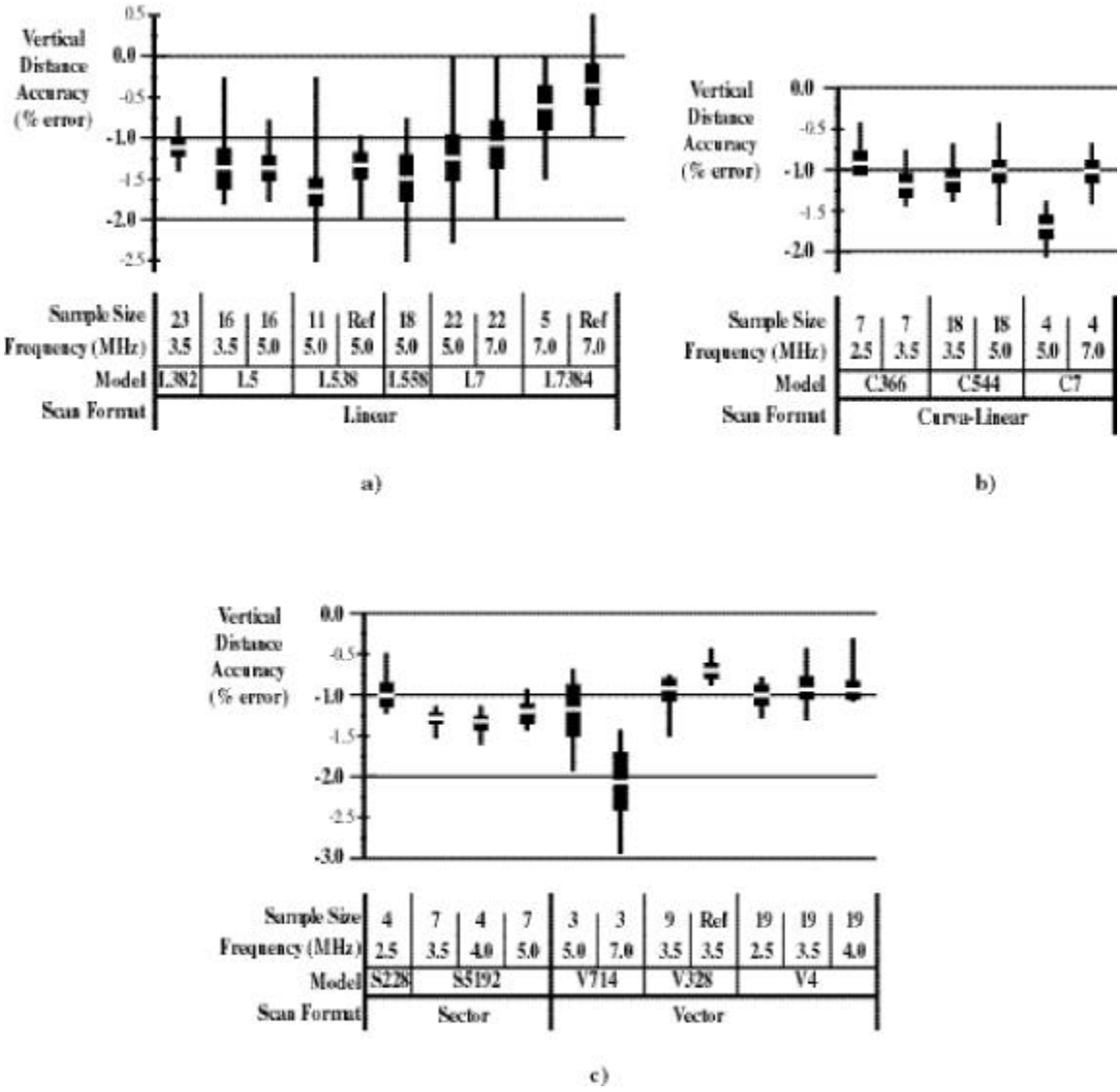

Figure 3. Vertical distance accuracy errors. The consistently negative errors of the vertical distance accuracy indicates that the speed of sound in the phantom may be slightly greater than 1540 m/s. a) Linear-array scan heads; b) Curva-linear transducers; c) Sector and vector format scan heads.

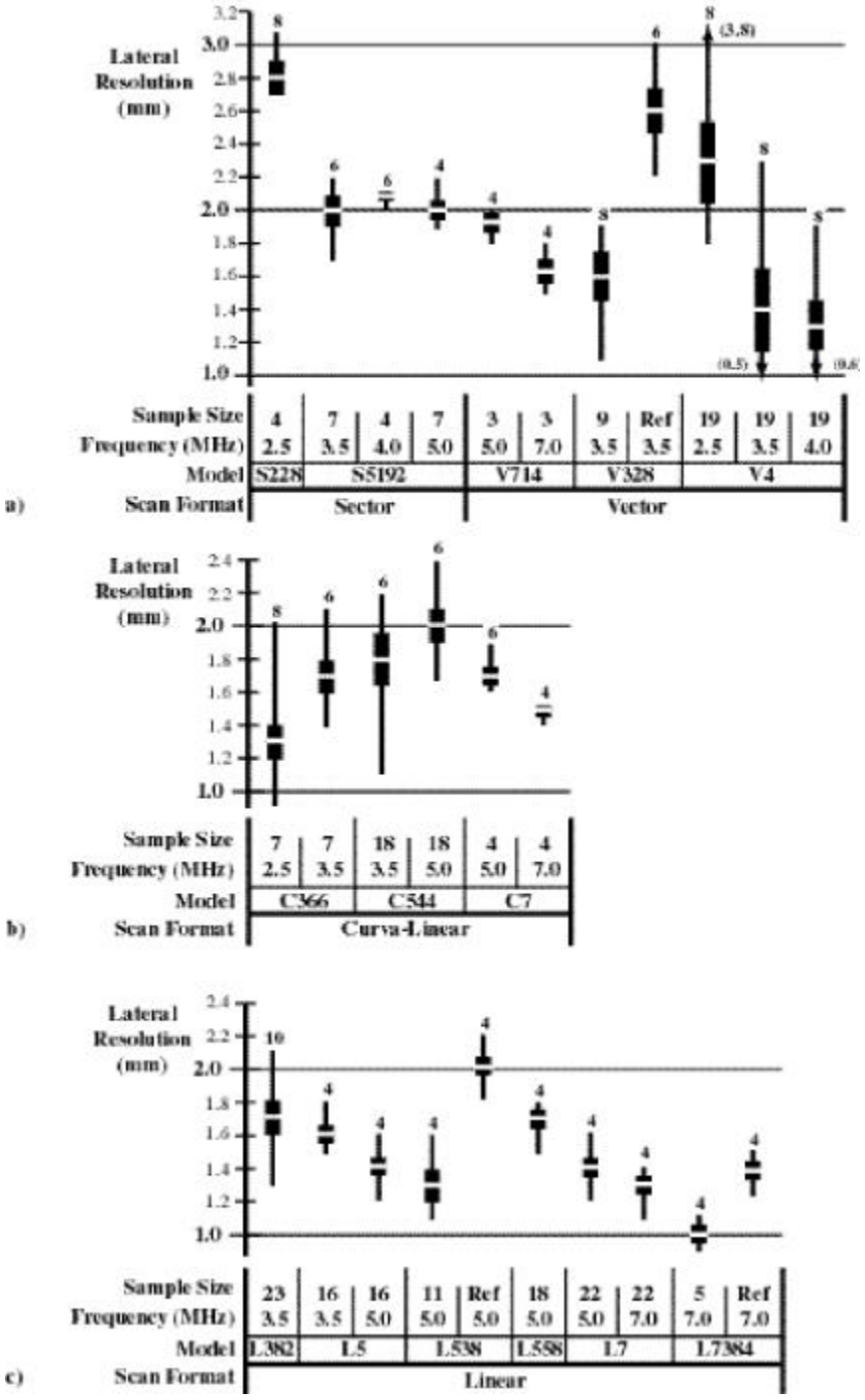

Figure 4. Lateral resolution measurements. Measurements were made at three different depths (when possible), however, only data form the mid-level depth is plotted. The number above the data point indicates the depth of the measurement in centimeters. a) Linear-array scan heads; b) Curva-linear transducers; c) Sector and vector format scan heads.

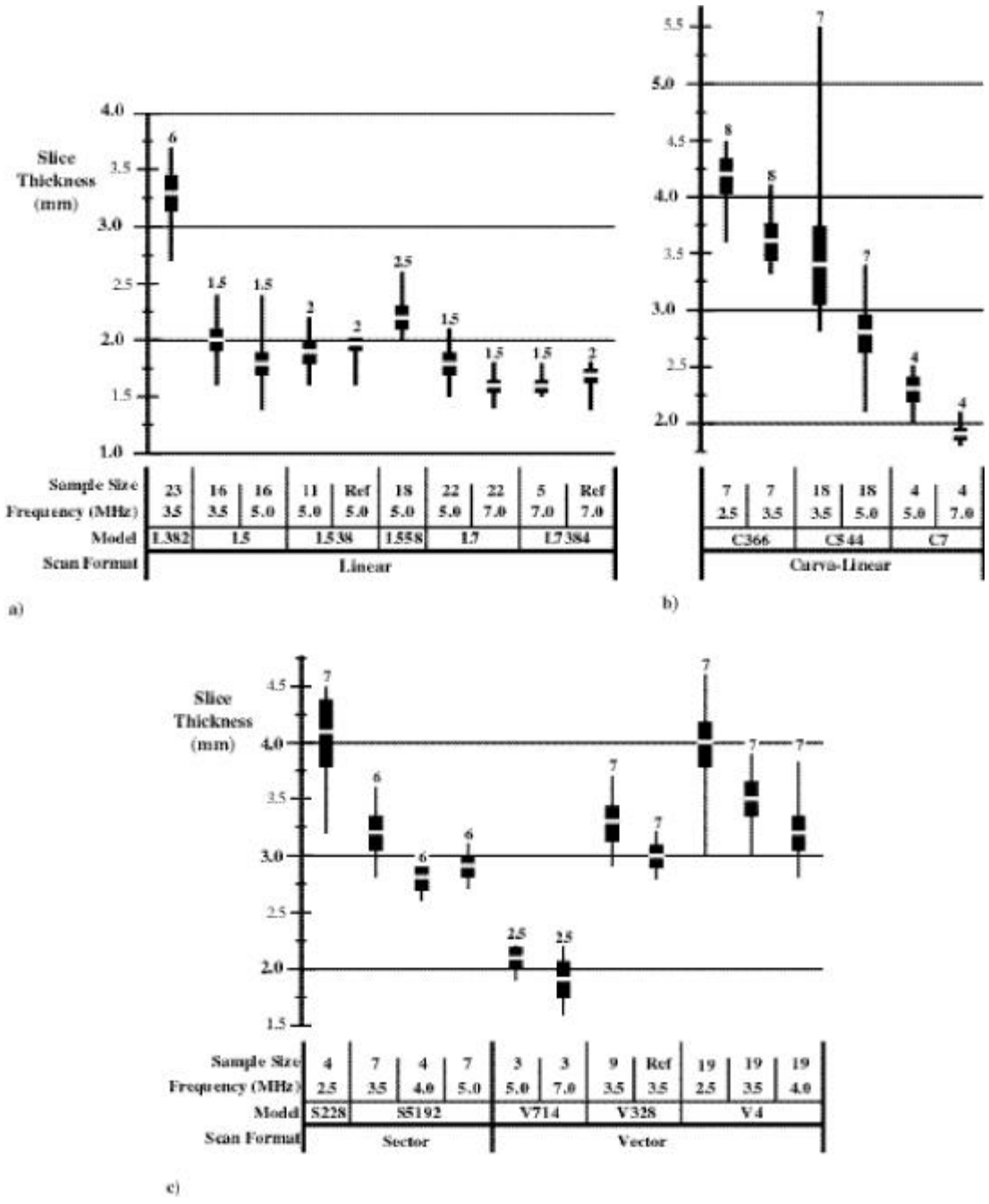

Figure 5. Slice thickness measurements. The transducer was translated across the top of the phantom to locate the depth corresponding to the minimum slice thickness—the depth of the

measurement is indicated above each data point. a) Linear-array scan heads; b) Curva-linear transducers; c) Sector and vector format scan heads.