

INTRODUCTION

In Korea, the quality assurance of CT, MRI and mammography have been regulated since 2004 by the Korean Institute for Accreditation of Medical Image (KIAMI) under the Ministry of Health, Welfare and Family Affairs. The goal of this program is to evaluate the image quality in medical examinations for the improvement of national health and achieved quality improvement of CT, MRI and mammography successfully. Based on these successes, Koran Government planned to expand the quality assurance system of diagnostic imaging to other imaging devices including ultrasound, PET-CT, fluoroscopy and angiography.

The importance of ultrasound (US) image quality assurance (QA) is widely recognized and recommendation for performing QA in US have been made by the major international scientific bodies, including American Institute of Ultrasound in Medicine (AIUM), the American Association of Physicists in Medicine (AAPM) and American College of Radiology (ACR). However, a standardized QA test has not yet been solidly established for US imaging, primarily because US examinations are conducted by highly diverse professional groups for their own purposes, and in most cases, there is no legal regulation system for US such as for ionizing diagnostic imaging modalities. Furthermore, the application of uniform standards is not easy because the technical development of US equipment has been rapid. For example, ACR standard for monitoring the performance of real-time US equipment relegates the determination of the standards and methods of QA and the analysis of the results to the users. Therefore, it is reasonable to establish separate QA standards for each professional group that performs its own specific examinations. In Korea, US examinations of the liver for the group at risk for hepatocellular carcinoma (HCC), i.e., carriers of hepatitis-B or hepatitis-C viruses, and patients with liver cirrhosis, are included in the National Cancer Screening Program. This program is run by the National Cancer Control Institute, which is a part of the National Cancer Center of Korea under the Ministry of Health, Welfare and Family Affairs, and is funded through taxes. The government of Korea decided to evaluate the quality of US examinations through this tax-funded program. A 3-year survey was planned for the period between 2008 and 2010 for all medical institutes participating in the program. The plan included the evaluation of all general hospitals in 2008, small hospitals other than general hospitals in 2009, and private clinics in 2010.

The Korean Society of Radiology, under the provision of National Cancer Center of Korea, performed this national survey with the help of the Korean Institute for Accreditation of Medical Image (KIAMI) and this presentation is the report of the results of this 3-year, national-wide survey for the investigation of the quality of US examination performed for the screening of HCC in high-risk patients and the preparation of the quality assurance regulation of US examination in Korea.

ACKNOWLEDGEMENT

This study was supported by a grant from the National Cancer Control Institute of the Nation Cancer Center of Korea and by the Ministry of Health, Welfare and Family Affairs of Korea. This study was performed by the help of the Korean Society of Radiology and the Korean Institute for Accreditation of Medical Image.

INVESTIGATION PROCESS

The investigation was performed for the all medical institute participating in the National Cancer Screening Program for hepatocellular carcinoma all over the country, for three years, from 2008 to 2010. General hospitals were investigated in 2008, small hospitals other than general hospital in 2009, and private clinics in 2010.

The evaluations for quality assurance of imaging examination can be divided into three categories: personnel evaluation, phantom image evaluation, and clinical image evaluation. For personnel, we investigated who (radiologists, physicians other than radiologists or radiological technicians) were performing screening ultrasound for the detection of hepatocellular carcinoma.

For phantom image evaluation, we used the ATS-539 multipurpose phantom (ATS laboratories, Bridgeport, CT, the USA). Research assistants transported the standard phantom to the medical institutes and obtained phantom images at each ultrasound scanners with a 3.0 to 5.0 MHz curved-array probe and software settings for abdominal ultrasound. Image settings were optimized by the physician on site and our research assistants. The dead zone, vertical and horizontal measurement, axial & lateral resolution, sensitivity, and gray scale/dynamic ranges were evaluated. We would like to survey the real situations and decided to perform the tests with same settings the patients underwent US scanning at. Therefore, we asked the physician on site to set up the scanner for optimization. The power output, brightness, contrast levels and time-gain control were controlled and optimized by research assistants and physicians on site. We obtained phantom images using the measurement methods described in the manufacturer's manual and AAPM guideline. The scanning of the phantom was performed by research assistants at the presence of physician on site because we considered the scanning conducted by the research assistants to be superior due to a lack of sufficient understanding of the US phantoms by the hospital staff.

For the clinical image evaluation, we acquired clinical images of the patients who already underwent ultrasound examination. For assessment of clinical images, we adopted the standard images established by the Korean Society of Radiology and the Korean Society of Ultrasound in Medicine. These standard images includes transverse and longitudinal scans of left hemiliver, subcostal and intercostal scans of right hemiliver, transverse plane of right and left portal veins, hepatic veins at hepatic dome, longitudinal view of gallbladder, long axisview of extrahepatic duct. Also, technical information including name, gender, age of the patients, name of the medical institute, name and date of examination, identification number of the patient, overall image quality, appropriateness of depth, location of focuses, annotations and presence of any artifacts were also evaluated.

All phantom and clinical ultrasound images were reviewed by the experienced abdominal radiologists with over 5-years of experience in ultrasound examination. Results of the investigation was analyzed for each categories of medical institutes, the year of manufacture (we categorized into three groups; within recent 5 years, between 5 and 10 years, more than 10 years), data storage form (digital or analogues) and personnel who performing ultrasound scanning.

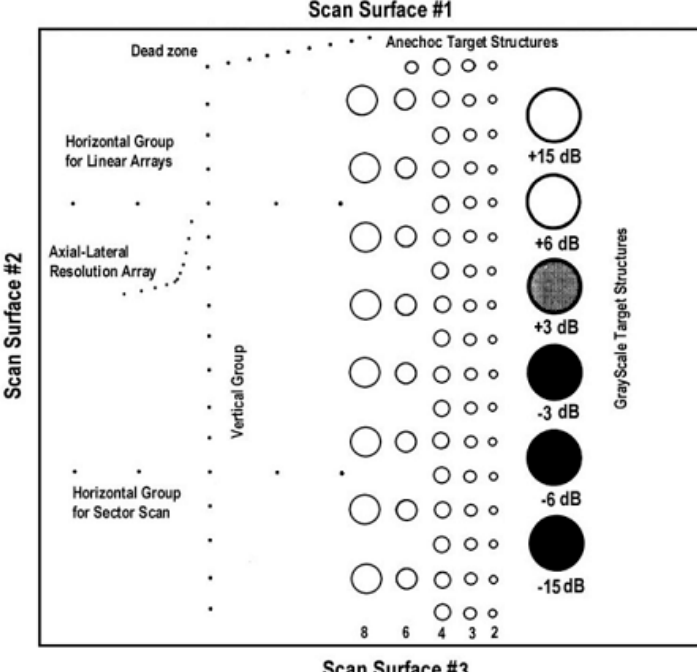
TEST ITEMS FOR QUALITY ASSURANCE AND CUT-OFF VALUES

Phantom Image Evaluation

Phantom image evaluation can be used to evaluates the performance and quality of hardware of US scanners. The items consist of conditions and parameters of contrast, spatial resolution, accuracy of measurements and penetration of US beam. For our examination, we used the ATS-539 multipurpose phantom (ATS laboratories, Bridgeport, CT, the USA), which is specified as the standardized phantom for US phantom images by the Korean Society of Radiology and the Korean Society of Ultrasound in Medicine in 2003. This phantom is constructed of a rubber-based tissue-mimicking material and is used to evaluate the accuracy and performance of US scanners. The phantom mimics the acoustic properties of human tissue and provides test structures within the simulated environment (Figure 1). The tests performed using this phantom focused on the dead zone, vertical and horizontal measurements, sensitivity, axial & lateral resolution, and gray scale/dynamic range. The focal zone and the functional resolution were also measured, but they were excluded from evaluation due to the difficulty in defining objective standards for these parameters.



Fig. 1 Outward appearance and target diagram of the standardized phantom. It has 4 scanning surfaces and many internal structures with which various measurement can be performed.



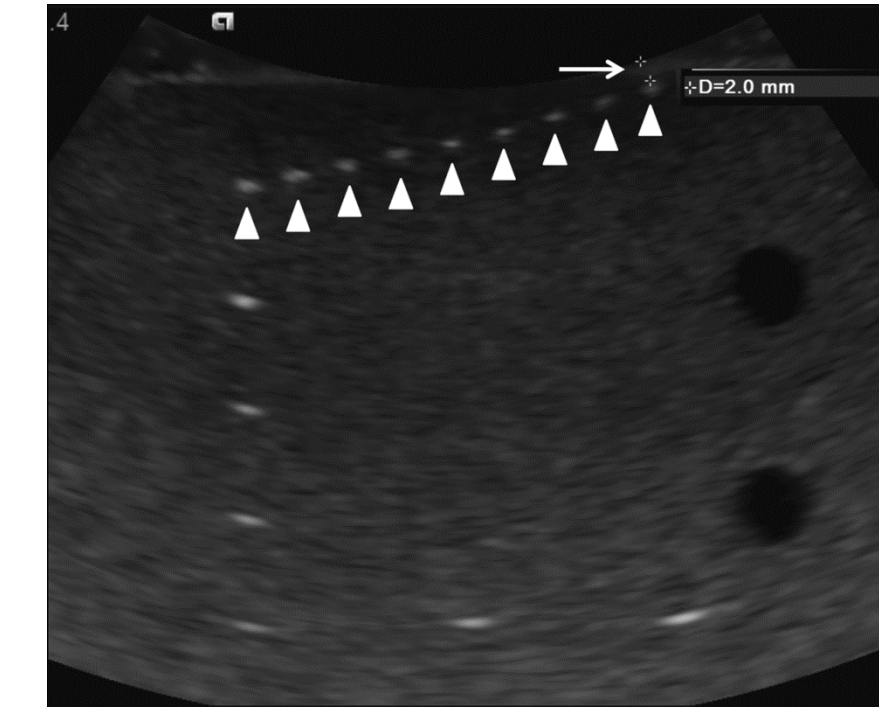
Both, subjective visual methods and objective computer-based approaches may be used to perform the phantom image evaluation. Due to subjectivity, manual measurement and visual assessment of phantom images are known to be less accurate than computerized automated measurements. However, considering the large number of scanners included in our future survey (more than 2000 US scanners in numerous hospitals and clinics) and the different formats of storing data (many of the private clinics used thermal paper or film), subjective visual assessment had to be accepted. To overcome the subjectivity stemming from manual measurement and visual assessment, the results were accepted only if both readers reached a consensus in 2008. Prior to evaluation, both readers received two hours of training on US phantom image interpretation. However in 2009 and 2010, only one radiologist reviewed the phantom image because more than 95% of agreement rate, which means the rate of agreement of "pass or fail" between two reviewers with the decided cut-off values, were observed in all test items of the phantom image evaluation in 2008 investigation.

Dead Zone

The dead zone is the distance from the front face of the transducer to the first identifiable echo at the phantom or patient interface. No clinical data can be collected in the dead zone. The target group was composed of 9 line targets with the first line target positioned 2 mm below the scan surface. Subsequent targets were spaced 1 mm apart to a depth of 10 mm. We measured the distance from the scan surface to the first identifiable line target. If the first line target of the nine targets was identifiable, the dead zone was less than 2 mm (Figure 2).

Figure 2. Dead zone.

Nine line targets are positioned between 2 to 10 mm below the scan surface. In this image, all nine line targets are clearly visualized (arrowheads). The distance between scan surface and 1st line target is the dead zone. In this case, dead zone is 2 mm. Focus is located as near as possible (arrow).



Vertical and Horizontal Measurement

The vertical and horizontal distance measurements were obtained both parallel and perpendicular to the axis of the sound beam. Accurate measurement of the size, depth, and volume of a structure is one of the critical factors in making a proper diagnosis. We measured 10.0 cm along the axis of the sound beam for vertical measurement (from 1.0 depth line target to 11.0 cm depth line target) and 8.0 cm perpendicular to the sound beam for horizontal measurement, and the resulting measurements were compared to the actual distance between the line targets in the phantom using US scanner's calipers (Figure 3). Focal zones were at the depth of the horizontal targets and be sure to use as little pressure as possible when applying the transducer to the scanning membrane to avoid displacement of the line targets in the phantom. For vertical measurement, the caliper markers were placed at the top of the echo from line target and for horizontal measurement, we placed the caliper markers above the centers of the echoes from line targets

Figure 3. Vertical and horizontal measurement

10 cm distance along the US beam axis (arrows) and 8cm distance perpendicular to US beam axis (arrowheads) are measured. Measurement should be done at the center and top of the each line target. In this US scanner, vertical measurement is 10.04 cm, and horizontal measurement is 8.28 cm. The discrepancies of vertical and horizontal measurement are 0.4% and 3.5%, respectively.

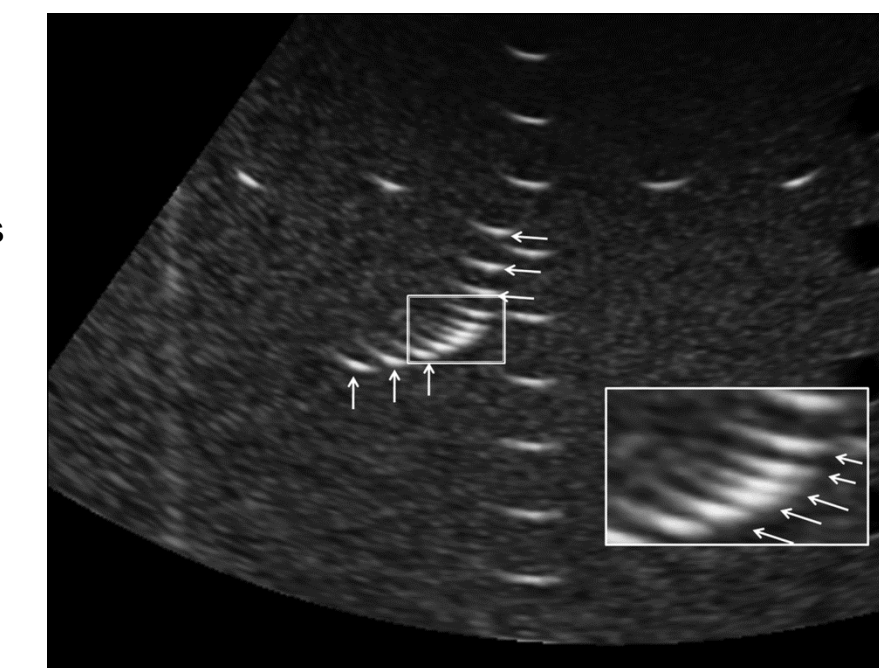


Axial & Lateral Resolution

Resolution is defined as the minimum reflector separation between two closely spaced objects that can be imaged separately. If a system has poor resolution, small structures lying close to each other will appear as one structure. The axial resolution is dependent on the pulsing system of the imaging device and the condition of the transducer, whereas the lateral resolution is affected by the beam width. The line targets in the phantom were spaced at 5.0, 4.0, 3.0, 2.0, and 1.0 mm intervals, both axially and laterally. Eleven line targets were present in the phantom, and we counted the number of line targets that were identifiable separately (Figure 4). The focus was located at target group and image zoom was applied.

Figure 4. Axial & lateral resolution

Eleven line targets with curved array are clearly visible separately. The distance between line targets are from 1 mm in central area to 5 mm in peripheral area. The curved array of line targets is for the test of axial and lateral resolution. In this image, all 11 line targets are visualized separately and clearly (arrows). The central part of the line targets are zoomed in rectangle.



Sensitivity

Sensitivity, which is a test of the penetration depth of the US beam, refers to the ability to image small objects located at specified depths. Anechoic, 8-mm round structures were located in the phantom along the direction of the US beam. The distance between the structures was 2.0 cm. We recorded the deepest target structure that was displayed in the US images. i.e., if the eighth structure was visible and appeared round, the sensitivity was more than 16 cm, and if the sixth structure was the deepest visible structure, then the sensitivity was 12 cm (Figure 5). The focus was located as deep as possible.

Figure 5. Sensitivity

8 mm sized anechoic, round structures are well visualized. 8 structures are clearly visible as round structure and the sensitivity is more than 16cm in this case (arrowheads). This test is for penetration depth of US beam. The focus is located as deep as possible (arrow) and the depth was set to 18 cm.

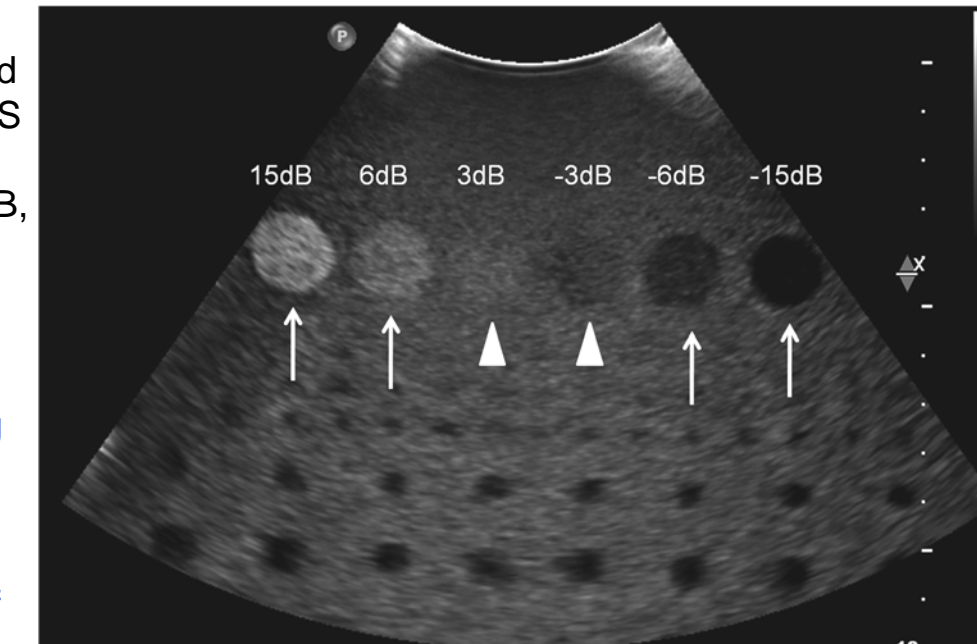


Gray Scale / Dynamic Range

The gray scale/dynamic range, which is a test for the contrast in US images, uses the amplitude of the received echoes to vary the degree of brightness in the displayed image. Six cylindrical targets with varying degrees of brightness were visible in the US images. These targets appeared circular in the US image plane. The contrasts of these targets relative to background material were +15 dB, +6 dB, +3 dB, -3 dB, -6 dB, and -15 dB. We counted the number of cylindrical targets that appeared as discrete round structures through more than 180 degrees (Figure 6).

Figure 6. Gray scale/dynamic range

Four or more cylindrical structures should be clearly visible over 180 degree among six structures for passing the test of gray scale/dynamic range. The contrasts of these targets relative to background material were +15 dB, +6 dB, +3 dB, -3 dB, -6 dB, and -15 dB. In this case, four cylindrical structures (+15 dB, +6 dB, -6 dB, +15 dB) are clearly visible as round ones (arrows). However, two cylindrical structures of contrast of +3 dB and -3 dB are not clearly visible as round ones (arrowheads).



Cut-off values

We have defined the cut-off values and the standards for US scanner phantom images by analyzing the results of the evaluation of phantom images acquired in general hospitals, 2008. We selected cut-off values for each test based on the following criteria: 1) the criteria recommended by the phantom manufacturer's manual and the major international scientific bodies such as AIUM, AAPM, or ACR; 2) the highest cut-off values that allowed at least 90% of the scanners to pass the QA testing. The results of our analyses for determining cut-off values for phantom images were described in the article; Choi JI, Kim PN, Jeong WK, et al. Establishing cutoff values for a quality assurance test using an ultrasound phantom in screening ultrasound examinations for hepatocellular carcinoma. An initial report of a nationwide survey in Korea. Journal of Ultrasound in Medicine 2011;30:1221-1229.

Fail of any of 6 test items was considered as "failed" scanner. Cut-off values of test items are summarized in Table 1.

Table 1. Cut-off values for test times of US phantom

Test Item	Cut-off Values
Dead zone	Less than 2 mm for the dead zone
Vertical measurement	Within 5 % discrepancy (10.0±0.5cm)
Horizontal measurement	Within 7.5 % discrepancy (8.0±0.6cm)
Axial & lateral measurement	All 11 identifiable line targets
Sensitivity ¹⁾	More than 14 cm
Gray scale / Dynamic range	More than 4 cylindrical structures identifiable

Clinical Image Evaluation

Clinical image evaluation assesses the quality of the patient image for diagnostic purposes. The clinical images were evaluated using the standard protocol established by the Korean Society of Radiology (KSR) and the Korean Society of Ultrasound in Medicine (KSUM). We selected 8 images as standards images for the clinical image evaluation among recommended 15 standards images by KSR and KSUM.

Standards images are comprised of six liver images and two biliary images; Transverse scan of left hemiliver, longitudinal scan of left hemiliver, transverse plane of left and right portal vein, hepatic dome including three hepatic veins, subcostal scan of right hemiliver, intercostal scan of right hemiliver, longitudinal scan of gallbladder, long axial scan of extrahepatic duct.

Hospitals should acquire the pertinent part of the image and they then submit this image to a research assistant when he/she visits the site or medical institutes can submit the clinical images by post. Because this investigation was a survey, not regulation though this was a foundation work for the regulation, we asked medical institutes to submit their best clinical images instead of designating clinical images of a specific patients. The clinical image can be submitted via hard copy or soft copy.

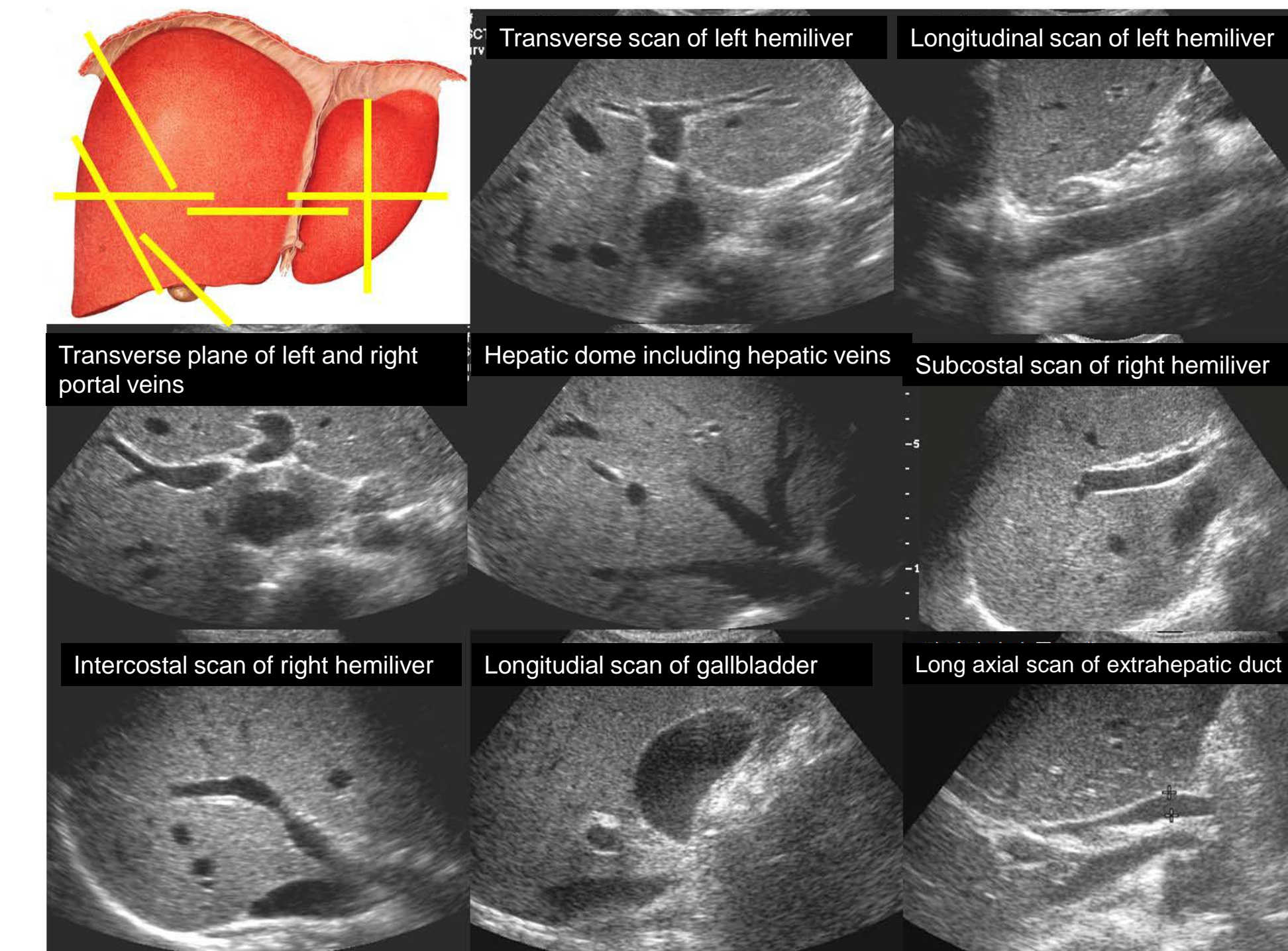


Figure 7. Eight standard images of the clinical image evaluation based on standard US protocol by the Korean Society of Radiology and the Korean Society of Ultrasound in Medicine.

¹Department of Radiology, School of Medicine, The Catholic University of Korea, Seoul, ²Department and Research Institute of Radiology, College of Medicine, University of Ulsan, Seoul, ³Department of Radiology, College of Medicine, Hanyang University, Seoul, ⁴Department of Radiology, College of Medicine, Kyung Hee University, Seoul, ⁵Department of Radiology, College of Medicine, Korea University, Ansan, ⁶Department of Radiology, College of Medicine, Yonsei University, Seoul, ⁷Department of Radiology, Research Institute and Hospital, National Cancer Center, Goyang, ⁸Department of Radiology, Ilsan Baek Hospital, College of Medicine, Inje University, Goyang, Republic of Korea

Clinical image evaluation is for testing appropriateness of the protocol of the US scanning, and anatomical / medical knowledge and skill of US scanning of the physicians who perform US examinations.

Test items of the clinical images include number of good images (16 points), identification (6 points), information from equipment (30 points), standard images (40 points) and artifacts (8 points). Score for number of good images was perfect when there are 8 good images because we recommended 8 standard images for US scanning. The importance of standard images was emphasized because fulfillment of all 8 images can guarantee the whole liver scan without missing area. For the information from equipment, proper position of focal zone and control of depth was included for test items to encourage the fine control of the scanning parameters during US scanning instead of scanning only with preset conditions.

Total score for the clinical image evaluation was 100 points, and the passing grade for this evaluation is a score over 60 points. Test items and valuation basis are summarized in Table 2.

Table 2. Valuation basis for the clinical image evaluation. Appraisal standards are as below.

Articles	Items	Score
1. No. of good images	1. No. of qualified images	2 or 0 point/image (Total 16 points)
2. Identification	1. Patient's name 2. Age/Sex 3. Registration NO. 4. Examiner 5. Date 6. Medical institution	1 or 0 point/items (Total 6 points)
3. Information from equipment	1. Similar and appropriate brightness 2. Proper position of focal zone 3. Control of depth 4. Display of scaler 5. Display of direction or body mark 6. Kind of transducer or frequency	5, 3, or 0 point/item (Total 30 points)
4. Standard images	1. Longitudinal scan of left hemiliver 2. Transverse scan of left hemiliver 3. Transverse plane of left and right portal veins 4. Hepatic dome including hepatic veins 5. Subcostal scan of right hemiliver 6. Intercostal scan of right hemiliver 7. Longitudinal scan of gallbladder 8. Long axis scan of extrahepatic duct	5, 3, or 0 point/item (Total 40 points)
5. Artifacts	1. Motion artifact 2. Mechanical artifact from damage of elements	4, 2, or 0 point/item (Total 8 points)

RESULTS OF THE QUALITY ASSURANCE TESTS

271 general hospitals with 357 ultrasound scanners, 467 hospitals other than general hospitals with 547 scanners, and 1184 private clinics with 1375 scanners were evaluated and overall, 1922 medical institutes and 2279 ultrasound scanners were tested. Number of participating US scanners for phantom and clinical image evaluation were summarized in Table 3.

	No. of institutes	No. of US scanners for phantom image evaluation	No. of US scanners for clinical image evaluation
General hospitals (2008)	271	357	266
Hospitals other than general hospitals (2009)	467	417	547
Private clinics (2010)	1184	363	1333

Table 3. Number of participating institutes and US scanners

Personnel Performing US Scanning

In Korea, there is no licensing system for sonographers and all US examination including scanning should be performed by a physician. Especially, for the patients of National Cancer screening, a physician should scan and interpret US examination and this is elucidated by the law. In all 271 general hospitals, US examinations (scanning and interpretation) were only performed by radiologists. Meanwhile in small hospitals other than general hospitals, radiologists only comprised 33.5% of performers of US examination scanings. In private clinics, only 17.1% of US examination scanings were performed by radiologists. However, there were many physicians whose majority was unknown in survey and therefore, true proportion of the radiologists was not clear. In private clinics, 4.2% of US scanings were performed by radiation technicians even though it is illegal for the patients of National Cancer Screening.

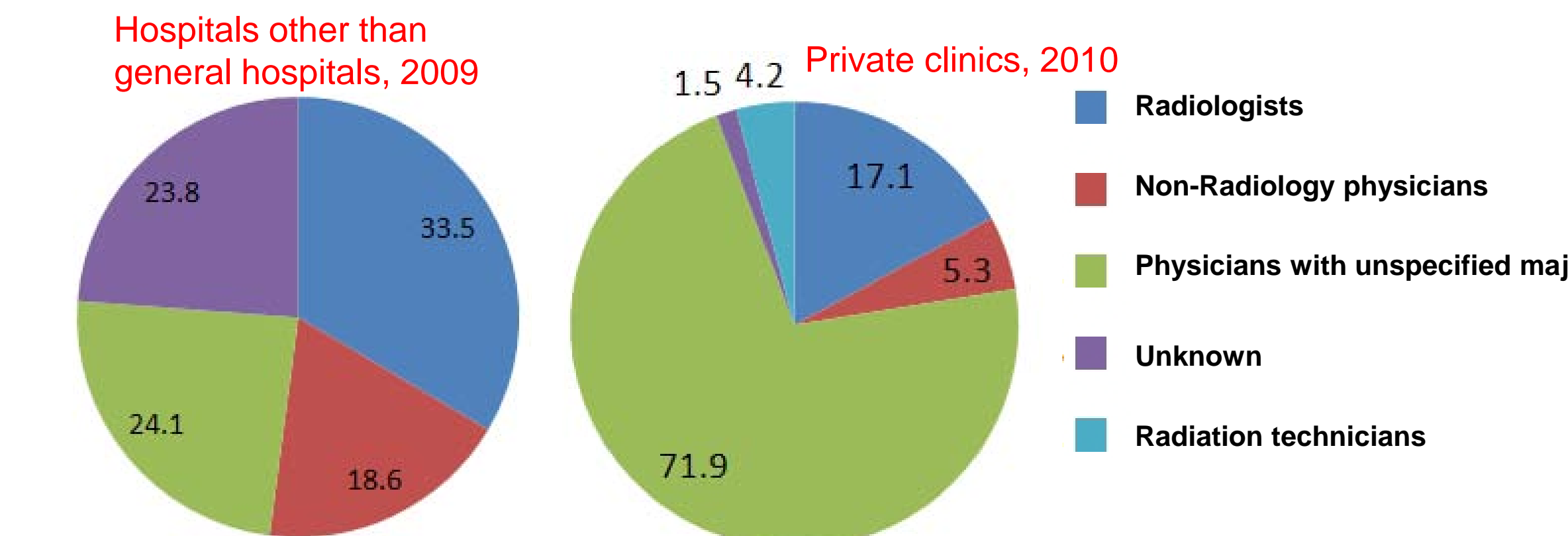


Fig. 8 Proportion of personnel who performed US examination scanning. Numbers are percentages.

Phantom Image Evaluation

Results of the failure rate according to the groups of medical institutes

Table 4 summarizes the results of the phantom image evaluation according to the groups of medical institutes. There was no significant difference of the failure rate among the different groups of hospitals. In other words, more than 20% of the US scanners constantly failed in the phantom image evaluation regardless of the groups of the hospitals and this means the needs of quality improvement of US scanners.

	General hospitals (2008)	Hospitals other than general hospitals (2009)	Private clinics (2010)	Overall
No. of scanners with the results of "Pass"	283	315	277	875
No. of scanners with the results of "Fail"	75	102	88	265
Failure rate, %	20.9	24.5	24.1	23.2
p-values comparing to general hospitals		0.265	0.328	

Table 4. Comparison of the failure rates of the phantom image evaluation according to the groups of the hospitals. P-values are calculated using Fisher's exact test, comparing to the failure rate of the radiologists.

Analysis of the causes of the failure for the phantom image evaluation

The failure rate of each test item was summarized in Table 5 and Figure 9. The most common cause of the failure was gray scale/dynamic range, which was responsible for 42.6% of the failure. If we omit the test items of dead zone and measurements which are less important to image quality of liver US, then about 23.7% of the failed cases will be redeemed.

	Dead zone	Vertical measurement	Horizontal measurement	Axial & lateral resolution	Gray scale/ Dynamic range	Sensitivity
No. of failed cases	47	11	16	39	133	66
Failure rate (%)	5.37	1.26	1.83	4.46	15.20	7.54

Table 5. Number of failed cases and failure rate of each test item of the phantom image evaluation.

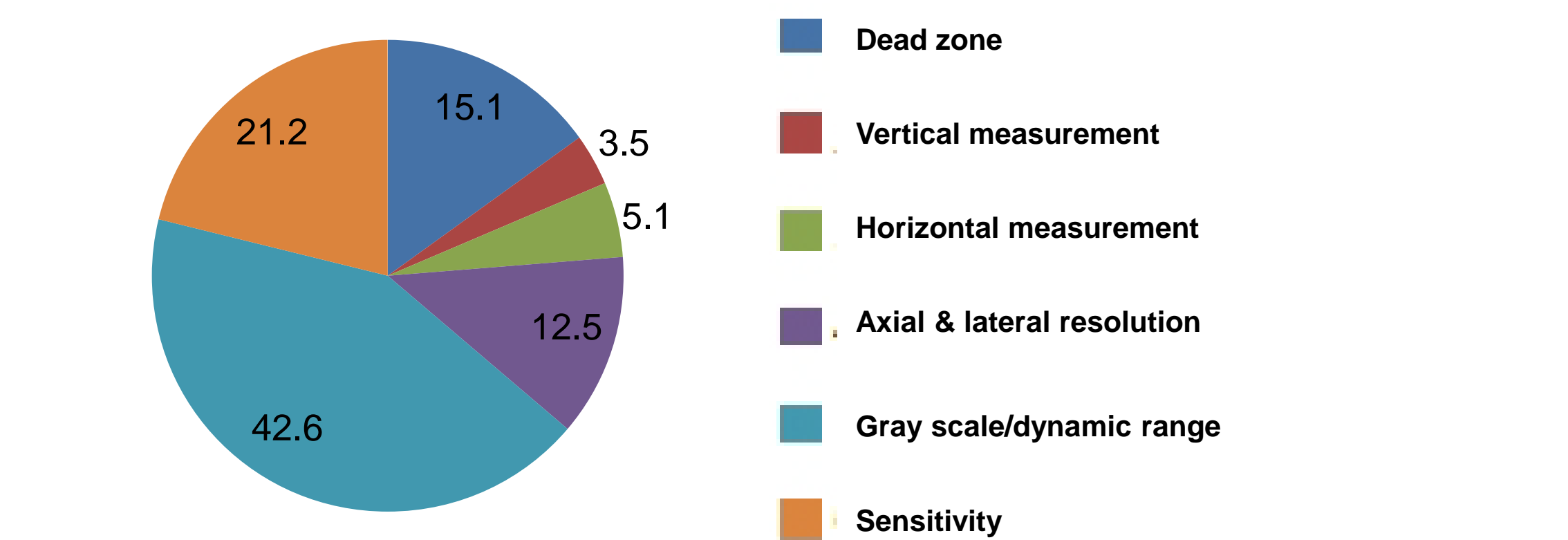


Fig. 9 Proportion of causes of the failure for the phantom image evaluation. Numbers are percentages.

Analysis of the failure rate according to years of manufacture

We evaluated the failure rate of the phantom image evaluation by years of manufacture because the phantom image evaluation is the test for the hardware itself of the US scanners. Hardware performance may fall considerably as times go on. Year of manufacture was available in only 2008 and 2010 data.

The failure rates for the phantom image evaluation with respect to the years of manufacture are summarized in Table 6. On the contrary to our expectation, there was no significant differences in the failure rates among the 3 groups based on the years of manufacture. This might be due to selection bias. Older scanners with poor performance might be already discarded and small number of US scanners more than 10 years old is the partial proof of this hypothesis

Years of manufacture	Recent 5 years	5 – 10 years	More than 10 years
No. of scanners with the results of "Pass"	232	144	69
No. of scanners with the results of "Fail"	54	43	15
Failure rate, %	18.8	23.0	17.9
p-values comparing to "recent 5 years"		0.265	0.328

Table 6. Failure rates by years of manufacture. P-values were calculated using Fisher's exact test.

Clinical Image Evaluation

Results of the failure rate according to the groups of medical institutes

Table 7 summarizes the results of the clinical image evaluation according to the groups of medical institutes. Results of general hospitals was the best with the failure rate of 5.5%. Hospitals other than general hospitals showed the worst results with the significant difference among three groups. This is probably due to the lack of concern for pay-doctors of the small hospitals; US examiners of private clinics might be the owners of the medical institutes but most of the US examiners of small hospitals are not. The best results of general hospitals might be originated from the fact that only radiologists performed US scanning in general hospitals, this is the results of another analysis that follows. (see Table 10)

	General hospitals (2008)	Hospitals other than general hospitals (2009)	Private clinics (2010)	Overall
No. of scanners with the results of "Pass"	241	398	1206	875
No. of scanners with the results of "Fail"	14	69	127	265
Failure rate, %	5.5	14.8	9.5	10.2
p-values comparing to general hospitals		0.0001	0.0405	
p-value comparing to private clinics		0.0024		

Table 7. Comparison of the failure rates of the clinical image evaluation according to the groups of the hospitals. P-values are calculated using Fisher's exact test.

Analysis of "passed" and "failed" cases by articles

In all articles, average scores of "failed" cases were lesser than "passed" cases. Especially, average scores of No. of good images and standard images were markedly differed and might be the main cause of the failure of the clinical image evaluation.

	No. of good images	Identification	Information from equipment	Standard images	Artifacts
Total score	16	6	30	40	8
Average of all cases	14.3	4.6	23.3	26.4	7.8
Average of "passed" cases	15.1	4.7	19.4	28.2	7.8
Average of "failed" cases	6.6	3.2	18.1	8.8	7.5
p-value	0.0001	0.0001	0.0001	0.0001	0.01

Table 8. Average scores of each test item for "passed" and "failed" cases. p-values were calculated using unpaired t-test.

Data storage form

Average score for No. of good images was markedly different between "passed" and "failed" cases. We have found out that No. of good images was significantly smaller in cases in which images for the clinical image evaluation were submitted in analogue format (films or thermal papers) than digital images. Table 9 Summarizes the results.

Format	Average score for No. of good images	Failure rate (Failure/Total)
Analogue	13.8	14.1% (116/821)
Digital	15.3	3.3% (25/753)
p-value	0.0001	0.0001

Table 9. Average scores of No. of good images and failure rate. p-values were calculated using unpaired t-test.

Standard images

We also evaluated the completeness of standard images. We defined complete standard images as images with score of 5. The proportion of complete standard images were summarized in Table 10; In all items, proportion of the complete standard images were significantly higher in "passed" cases than "failed" cases. Proportions of complete images of "failed" cases were less than 10% in longitudinal scan of left hemiliver and longitudinal scan of gallbladder. Proportion of complete image was less than 50% even in "passed" cases in long axis scan of extrahepatic duct, and more careful scan for extrahepatic duct is recommended even in more than half of "passed" cases.

	Longitudinal scan of left hemiliver	Transverse scan of left hemiliver	Transverse plane of left and right portal veins	Hepatic dome including hepatic veins	Subcostal scan of right hemiliver	Intercostal scan of right hemiliver	Longitudinal scan of gallbladder	Long axis scan of extrahepatic duct
"Passed" cases	73.4%	77.3%	64.6%	75.3%	82.8%	75.6%	81.4%	46.3%
"Failed" cases	8.2%	17.6%	25.8%	21.2%	36.5%	12.9%	20.0%	1.2%
p-values	0.0001	0.0001	0.0001	0.0001	0.0004	0.0004	0.0001	0.0001

Table 10 Comparison of the proportion of complete standard images of "passed" and "failed" cases. p-values are calculated using Fisher's exact test.

Comparison of the failure rate of clinical image evaluation according to the personnel

Because the clinical image evaluation is for the test of the protocols and personnel, we compared the failure rate of different groups of personnel who scanning US examination.

Failure rate of the radiologists was 6.4% and significantly better than other groups of the personnel. Non-radiology physicians showed poor failure rate which was similar to that of the radiation technicians. Physicians with unspecified major showed the result between that of the radiologists and non-radiologist physicians and this is probably due to they are made up of the mixture of radiologists and non-radiology physicians. Radiologists are experts of US examinations and this result might be the cause of the results presented in Table 7.

	Radiologists	Non-radiology physicians	Physicians with unspecified major	Radiation technicians
No. of scanners with the results of "Pass"	634	137	984	44
No. of scanners with the results of "Fail"	43	36	106	12
Failure rate, %	6.4	20.8	9.7	21.4
p-values comparing to radiologists		0.0001	0.0136	0.0004

Table 11. Comparison of the failure rates of the clinical image evaluation according to the personnel who performed US scanning. P-values are calculated using Fisher's exact test, comparing to the failure rate of the radiologists.

CONCLUSION

We have performed a nation-wide survey of screening US examination for the patients at risk of hepatocellular carcinoma. Proportions of radiologists in small hospitals and private clinics were lower than expected. Failure rates of the phantom image evaluation were 20.9-24.5% in all groups of hospitals and those of the clinical image evaluation were 5.5 – 14.8%. Failure rate of small hospitals was the worst for the clinical image evaluation. There was no significant difference of failure rate among the groups of US scanners with different year of manufacture. Medical institutes submitting clinical images with the analogue images resulted poorer failure rate and possibly due to the small number of standard images submitted. For the personnel who performs US scanning, radiologists showed best results for the clinical image evaluation, comparing to other groups including non-radiology physicians, physicians with unspecified major, and radiation technicians.

REFERENCES

- AIUM technical standards committee. Quality control manual for gray-scale ultrasound scanners-stage 2. American Institute of Ultrasound in Medicine; 1995.
- ACR technical standard for diagnostic medical physics performance monitoring of real time ultrasound equipment: 2009. http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/med_phys/
- Kim PN, Kim KW, Byun JH. Quality assessment of hepatic ultrasound images examined after a medical check-up. J Korean Soc Ultrasound Med 209;28:31-37.
- Kim PN, Lim JW, Kim HC, et al. Quality assessment of ultrasonographic equipment using an ATS-539 multipurpose phantom. J Korean Radiol Soc 2008;58:533-541.
- ATS Laboratories. Test performed. Bridgeport, CT: ATS Laboratories, Inc., 2000.
- Choi JI, Kim PN, Jeong WK, et al. Establishing cutoff values for a quality assurance test using an ultrasound phantom in screening ultrasound examinations for hepatocellular carcinoma. An initial report of a nationwide survey in Korea. J Ultrasound Med 2011;30:1221-1229.