

Effectiveness of on-site education for quality assurance of screening ultrasonography for hepatocellular carcinoma

Joon-II Choi¹, Seung Eun Jung¹, Woo Kyoung Jeong², Hyun Cheol Kim³, Chandana Lall⁴, Yeol Kim^{5,6}, Kui Son Choi^{5,6}, Mina Suh⁵, Boyoung Park^{5,6}

¹Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea, ²Department of Radiology, Samsung Medical Center, Sungkyunkwan University, Seoul, Republic of Korea, ³Department of Radiology, Kyung Hee University Hospital at Gangdong, Seoul, Republic of Korea, ⁴Department of Radiological Sciences, University of California, Irvine, USA, ⁵National Cancer Control Institute, National Cancer Center, Goyang, Gyeonggi-do, Republic of Korea, ⁶Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, Gyeonggi-do, Republic of Korea

Abstract

Objective: To evaluate the usefulness of on-site education for clinical imaging evaluation using quality assurance (QA) testing of surveillance ultrasonography (US) for hepatocellular carcinoma (HCC). **Material and methods:** Thirty-eight medical institutes underwent on-site education in 2012 for QA testing of clinical imaging evaluation of surveillance US for HCC. Failure rates and mean scores of clinical imaging evaluation for surveillance US of the 2011 survey, the 2012 survey after on-site education and the 2013 survey were compared. **Results:** Failure rates and mean scores of the 2011 survey, the 2012 survey after education and the 2013 survey were 81.6%, 18.4%, 21.1% and 61.7, 82.7 and 74.6, respectively. Pair-wise analyses demonstrated that the failure rate of the 2011 survey was significantly larger compared to that observed in the other surveys. Mean score of the 2013 survey was worse than that of the 2012 survey after on-site education. **Conclusions:** On-site education positively impacts the failure rate and scores of clinical imaging evaluation of screening US for HCC. However, the impact may be reduced over time, and repeated, annual education might be necessary to maintain US quality.

Keywords: ultrasonography, quality assurance, education, surveillance

Introduction

Given the relationship between imaging diagnoses and patient safety in modern medicine, the importance of quality assurance (QA) in medical imaging has recently received increased attention. In Korea, QA of computed tomography (CT), magnetic resonance imaging (MRI), and mammography have been regulated since 2004 by the Korean Institute for Accreditation of Medical Image

(KIAMI) under the control of the Ministry of Health, Welfare and Family Affairs[1,2]. The goals of this accreditation program were to evaluate CT, MRI, and mammography images to ultimately improve the quality of medical imaging for public health. However, the accreditation program for ultrasonography (US) was not enforced due to the complexity of US examinations, in particular the variety of US units, the myriad of examinations used, and the many expert groups involved.

In order to perform QA on US, one of the most widely utilized modalities for cancer surveillance, the Korean government and the Korean Society of Radiology (KSR), increased focus on the US surveillance for hepatocellular carcinoma (HCC) because surveillance US for HCC can be standardized and the required hardware can be simplified. Surveillance US for HCC is included in the National Cancer Screening Programs in Korea and funded by tax dollars, but the reported cancer detection

Received 29.02.2016 Accepted 06.04.2016

Med Ultrason

2016, Vol. 18, No 3, 275-280

Corresponding author: Seung Eun Jung

Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul, 06591, Republic of Korea
Phone: 82-2-2258-1431, Fax: 82-2-599-6771
E-mail: sejung@catholic.ac.kr

rate was below expectations. Therefore, KSR and KI-AMI performed QA tests for surveillance US of HCC in this national program [3-5]. The QA program consists of personnel evaluation, phantom image evaluation for the hardware and software of US units, and clinical image evaluation for the protocol and scanning methods. The results of the clinical imaging evaluation can be improved with the appropriate knowledge of scanning protocols and techniques. However, some medical institutes repeatedly failed in clinical image evaluation, even after a related lecture-type, cluster education for QA of US was performed repeatedly. Therefore, KSR and the National Cancer Center of Korea planned visits and on-site education for the QA of clinical imaging examinations of surveillance US for HCC

On-site education is a type of hands-on education in which educators visit medical institutes and counsel the institutional staff. It takes substantial time and effort on the part of the educators; however, it is thought to be more effective than lecture-type, passive, cluster education.

The purpose of this study was to evaluate the usefulness of on-site education for clinical imaging evaluation of surveillance US for HCC.

Material and methods

Investigation process

This study was approved by the Institutional Review Board of National Cancer Center of Korea. Written informed consent was waived because of the retrospective nature of the study.

As mentioned above, QA of the imaging examination can be divided into three categories: personnel evaluation, phantom image evaluation, and clinical image evaluation. We concentrated on clinical image evaluation, as it is the only examination that deals with protocol appropriateness, which can ultimately be improved through education. Only selected medical institutes were included in this study due to the nature of the demonstration project. First, we selected 123 institutions that failed the clinical image evaluation at least twice between 2008 and 2011. Among them, 75 institutes refused on-site education, and 48 medical institutes underwent on-site education in 2012 for QA testing of clinical imaging evaluation of surveillance US for HCC. Among them, 10 institutes were excluded for analyses because they lacked results from the 2012 or 2013 surveys. Finally, 38 medical institutes were included for this retrospective study. Before the annual survey in 2012, on-site education was performed by expert radiologists involved in QA testing of surveillance US for more than three years. On-site education was provided to doctors that perform US scanning and

interpretation. Visiting educators reviewed images and reports of previous examinations performed at the medical institutes and explained the test items and protocol for clinical image evaluation. Scanning education was also performed, and question-and-answer (Q&A) sessions were held for "student" questions. Several months after on-site-education, the 2012 surveys were given to evaluate the effects of on-site education. We also collected the results of the 2011 and 2013 surveys at the same medical institutes for comparison purposes. Failure rates and mean scores of clinical imaging evaluation of the 2011 survey, the 2012 survey after on-site education, and the 2013 survey were compared.

Test items and scoring system for QA testing of clinical image evaluation

Clinical image evaluation is performed to evaluate the appropriateness of the scanning protocol, as well as the relevant anatomical and medical knowledge of physicians that perform US examinations. Because this investigation was a survey and not a regulation, we asked medical institutes to submit their best clinical images instead of the images from specific patients.

Scoring systems for the clinical imaging evaluation were developed by consensus from experts in the KSR and the Korean Society of Ultrasound in Medicine (KSUM) [4]. Test items included *number of good images* (16 points), *presence of proper reports* (4 points), *identification* (4 points), *information from equipment* (18 points), *standard images* (40 points) and *artifacts* (12 points). Score for *number of good images* was perfect when there were eight good images. For *information from equipment*, proper position of focal zone and control of depth were included as test items to encourage the fine control of scanning parameters during US scanning.

Standard images were comprised of six liver images and two biliary images: left hemiliver axial scan, left hemiliver sagittal scan, left and right portal vein transverse planes, hepatic dome including three hepatic veins, right hemiliver subcostal scan, right hemiliver intercostal scan, gallbladder longitudinal scan, and an extrahepatic duct long-axial scan. These eight standard images were selected from the 15 standard images for abdominal US recommended by the KSR and KSUM in 2001 [4]. The importance of standard images was emphasized, as fulfillment of all eight images guarantees whole-liver scanning.

Failure of clinical image evaluation was defined as: 1) less than 60 out of 100 points in clinical image examinations, 2) absence of essential information including patient name, patient sex/age, hospital identification, or date of examination because missing of these data can result the wrong patient identification. Reviewers, who are abdominal radiologists familiar with QA tests for US, scored the

Table I. The scoring system for clinical image evaluation of surveillance ultrasonography for hepatocellular carcinoma.

Items	Sub-items	Score
1. Number of good images	1. Number of qualified images	2 or 0 point/image (Total 16 points)
2. Proper report	1. Presence of proper report	4 or 0 point
3. Identification 1	1. Patient's name 2. Age/sex or registration number 3. Date of examination	Compulsory
4. Identification 2	1. Name of medical institute 2. Name of examining doctor	2, 0 point/sub-item (Total 4 points)
3. Information from equipment	1. Appropriate brightness and contrast 2. Proper position of focal zone 3. Proper depth of images 4. Display of direction or body mark	*6, 3, 0 point/sub-item (Total 24 points)
4. Standard images	1. Sagittal scan of left hemiliver 2. Axial scan of left hemiliver 3. Transverse plane of right and left portal veins 4. Hepatic veins at hepatic dome 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	¶6, 3, or 0 point/item (Total 12 points)

**Information from equipment*: 1) Six points in cases with adjustments of over half of the images; 2) Three points in cases with adjustments of under half of the images, §*Standard images*: 1) Five points in cases with complete visualization of each anatomic structure; 2) Three points in cases with partial visualization of each anatomic structure, ¶*Artifacts*: 1) Motion artifact: a. Six points in cases of no artifacts; b. Three points in cases of noticeable artifacts below half of the images; c. Zero points in cases of noticeable artifacts over half of the images; 2) Damage of elements: a. Six points in cases of no artifacts; b. Three points when an artifact is at the periphery of a transducer; c. Zero points for central artifacts.

clinical images according to the score tables. Test items for clinical image evaluation are summarized in Table I.

Statistical analysis

Failure rates and mean scores of clinical image evaluation for surveillance US of the 2011 survey, the 2012 survey after on-site education, and the 2013 survey were compared. Additionally, scores for each test item were compared among surveys. Failure rates were compared using the Friedman test and the paired McNemar's test. Mean scores were compared using one-way, repeated measure analysis of variance (ANOVA) and the Bonferroni test as a posthoc analysis. p-values less than 0.05 were considered statistically significant.

Results

In the 2011 survey, 19 medical institutes failed clinical image evaluation due to a low score, while 17 failed due to the absence of essential information. Five institutes failed due to both a low score and lack of essential information. In the 2012 survey, only two medical institutes failed due to a low score, while six failed because of the absence of essential information. One institute failed due to both low score and lack of essential information.

In the 2013 survey, seven medical institutes failed due to a low score and two due to the absence of essential information. One institute failed due to both a low score and lack of essential information.

Failure rates in the 2011 survey, the 2012 survey after education, and the 2013 survey were 81.6% (31/38), 18.4% (7/38), and 21.1% (8/38), respectively. The Friedman test revealed a significant difference in the failure rate among the three surveys ($p < 0.001$). Figure 1 illustrates a representative case of failure by low score (< 60). Pair-wise analyses using the paired McNemar's test indicated that the failure rate of the 2011 survey was significantly inferior to the results of other surveys. Failure rates by score only (< 60 points, excluding failure by absence of essential information) were 50.0% (19/38), 5.3% (2/38), and 18.4% (7/38) for the 2011 survey, the 2012 survey after education, and the 2013 survey, respectively. The Friedman test also revealed a significant difference in the failure rate by scores only among the three surveys ($p < 0.001$). Pair-wise analyses using the paired McNemar's test also showed that the failure rate of the 2011 survey was significantly inferior to the results of other surveys.

The mean scores from the 2011 survey, the 2012 survey after education and the 2013 survey were 61.7,

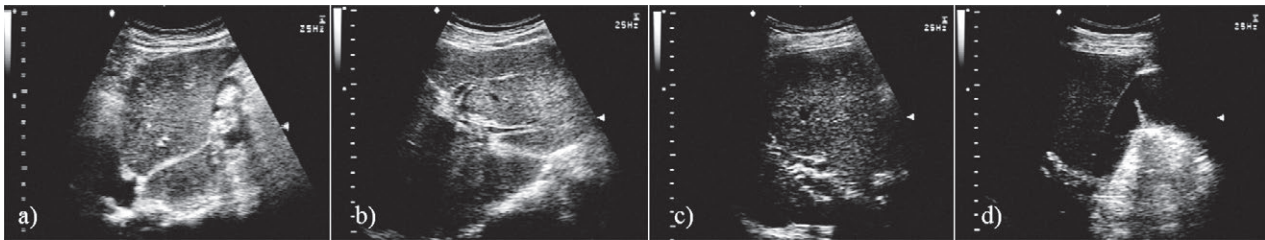


Fig 1. A representative failed case in a clinical image evaluation. Only four images were acquired in this examination for liver and biliary system. Four images are a) sagittal scan of left, b) axial scan of left hemiliver, c) subcostal scan of right hemiliver, and d) longitudinal scan of gallbladder. Also, brightness and contrast are poor and some motion artifacts are noted. Total score for this examination was 53.

Table II. Mean scores for each survey item.

	Number of good images	Proper report	Identification	Information from equipment	Standard images	Artifacts	Total scores
2011 survey	12.0	3.8	1.6	12.0	21.1	11.2	61.7
2012 survey	15.3	3.5	2.0	14.4	36.1	11.4	82.7
2013 survey	15.1	3.8	2.0	13.7	28.4	11.5	74.5
p-values	<0.001	0.277	0.564	0.085	<0.001	0.761	<0.001

P-values were calculated using one-way repeated measure analysis of variance (ANOVA).

Table III. Results of posthoc analyses.

	2011 vs. 2012	2012 vs. 2013	2011 vs. 2013
Total failure rate*	<0.001	1.000	<0.001
Failure rate by score only*	<0.001	0.063	0.004
Total scores§	<0.001	0.015	0.001
Number of good images§	<0.001	1.000	0.001
Proper reports	0.495	0.498	1.000
Identifications	0.996	1.000	1.000
Information from equipment§	0.097	1.000	0.344
Standard images§	<0.001	<0.001	<0.001
Artifacts	1.000	1.000	1.000

* Calculated using the paired McNemar test. § Calculated using the Bonferroni test.

82.7, and 74.6, respectively. One-way ANOVA yielded a p-value of less than 0.001. Posthoc analysis using the Bonferroni test revealed that the score from the 2011 survey was the worst. The score from the 2013 survey was worse than the score from the 2012 survey after education (p=0.015). Also, scores for the *number of good images* were worst in the 2011 survey. For the scores of standard images, the 2011 survey was the worst, while the 2012 survey after education was the best. Results and posthoc analyses are summarized in Tables II and III.

Discussions

Early detection of malignancies via screening/surveillance testing is one of the most effective ways to prevent death due to cancer. Imaging examinations such

as US for HCC, mammography for breast cancer, and fluoroscopic examination for stomach or colon cancers play a crucial role in early cancer detection. However, for optimal screening examination results, high-quality imaging studies are also crucial. US surveillance of HCC every six months in populations at an elevated risk has been proven to reduce HCC related deaths and is recommended as a standard protocol in many countries, including the United States, Europe, Japan, and Korea [6-10]. Recent guidelines for the management of HCC from the American Association for the Study of Liver Disease (AASLD) and the European Association for the Study of Liver (EASL) recommended maintaining high-quality US examinations for optimal surveillance [6,7].

Acquiring standard images is very important for the US surveillance of HCC, as it guarantees that the entire liver is imaged (except the hepatic dome, which cannot be scanned with US). According to a meta-analysis by Singal et al, pooled sensitivity of US for HCC screening is about 60% for small HCCs [11]. In our speculation, a primary cause of surveillance failure may be a lack of scanning over areas where HCC exists. Therefore, we designed the clinical image evaluation to emphasize acquiring standard images, which can guarantee the scanning of the whole liver. Compliance of standard protocol is one of the building blocks for quality assurance in most of the businesses. However, even with the repeated cluster educations, some medical institutes repeatedly failed in clinical image evaluation. Therefore, we planned on-site education

On-site education is a type of hands-on education in which educators visit medical institutes and counsel the institutional staff. It takes substantial time and effort on the part of the educators; however, it is thought to be more effective than lecture-type, passive, cluster education. On-site education and field education have been reported to be effective for medical personnel in various fields [12-14]. In our study, on-site education encompasses hands-on training of individuals that perform US examinations. Educators reminded “students” of recommended US surveillance protocols for HCC, and Q&A sessions were held. The results of our study imply that on-site education was effective for improving the results of clinical image evaluation. Failure rates were reduced after on-site education, and total scores were improved. Particularly, standard image scores were markedly improved after on-site education. This implies that lack of knowledge of test items and standard images were the primary causes of failure in clinical image evaluation. Contrary to phantom image evaluation (which is closely related to the performance of the hardware and software of US units), clinical image evaluation is a test of protocols and scanning techniques, factors that can be improved with appropriate education.

Results of our studies are in accordance with the previous study done in Korea. Failure rates of clinical image evaluation from 2008 to 2010 were 5.5 to 14.8%, and the primary causes of failure were low scores for the *number of good images* and *standard images* [4]. According to the results of our study, the *number of good images* and *standard images* can be significantly improved with on-site education.

However, on-site education has two critical drawbacks: cost and time. Providing well-trained educators to institutions is costly, both monetarily and temporally. To overcome these issues, on-site education should be limited to medical institutes that repeatedly failed QA testing. Additionally, online, interactive education could be considered as an alternative to on-site education.

Another concerning trend with our results is that the positive effect of education on the score became less pronounced with time. The results of the 2013 survey were inferior to those of the 2012 survey after on-site education, and the mean total score of the 2013 survey was inferior to that of the 2012 survey after on-site education. Failure rate by score only for the 2013 survey was 18.4%, while that of the 2012 survey was 5.3%. Although the p-value was 0.063, we may have uncovered a significant difference if the sample size was slightly larger. Therefore, frequent re-education may be necessary in order to maintain the effects of on-site education for clinical image evaluation. However, as mentioned above, it is costly

to repeatedly perform on-site education, and we must seek other ways to maintain the effect of on-site education. Annual supplementary education, whether online or clustered, may be another option for repeated education.

Our study has a significant drawback. We have insufficient evidence to claim that the results of the clinical image evaluation are associated with the performance of surveillance US for HCC. Thus we cannot make the claim that failed examinations for clinical image evaluation result in poor HCC detection. One can detect and diagnose HCCs even if he/she is not complying with the standard protocol. However, that kind of evidence cannot be obtained if we are unable to include either standardized patients or very large numbers of patients. We believe that adherence to the standard protocol and obtaining standard images may be helpful for whole-liver scanning to reduce the possibility of missing HCCs, especially in the setting that many non-experts do the US scanning and interpretation.

Our study has some limitations. Firstly, the number of medical institutes was too small, primarily due to budgeting concerns. However, the results of our study can serve as a basis for rebuilding the nationwide educational program for QA of screening examinations. Secondly, too many candidate institutes refused to join the on-site education program, and many medical institutes did not undergo follow-up survey in 2012 (after on-site education) or 2013. Many medical institutes consider QA testing unnecessary and bureaucratic. Also, many medical institutes were reluctant to the visiting of educators because they thought that the educators were auditors from the government. An effort should be made to overcome this stereotype. Thirdly, the scoring system for clinical image evaluation was consensus-based and not derived from scientific evidence. Even though previous studies have been performed, guidelines were arbitrarily set by experts. However, this is a demonstration program and not a legal regulation, and the results of our study will be useful for establishing the education system for the QA of US. Fourthly, we evaluated the “best images” from medical institutes, again because this survey was a demonstration program that may not reflect reality (and failure rates might therefore be underestimated).

In **conclusion**, on-site education positively impacts failure rates and mean scores of clinical image evaluation of surveillance US for HCC. However, the impact may be reduced after some time, and repeated, annual education may be necessary to maintain the quality of surveillance US

Acknowledgement: We would like to thank our reviewers and educators for on-site education and scoring.

Additionally, we greatly appreciate the assistance of the researchers from the Korean Institute for Accreditation of Medical Image (KIAMI). This study was supported in part by National Cancer Center Grant [1560460-1]. This study was also partially supported by a grant from the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (150160).

Conflict of interest: none

References

1. Park HJ, Jung SE, Lee YJ, et al. Review of failed CT phantom image evaluations in 2005 and 2006 by the CT accreditation program of the Korean Institute for Accreditation of Medical Image. *Korean J Radiol* 2008; 9: 354-363.
2. Park HJ, Jung SE, Lee YJ, et al. The relationship between subjective and objective parameters in CT phantom image evaluation. *Korean J Radiol* 2009; 10: 490-495.
3. Lee S, Choi JI, Park MY, et al. Intra- and interobserver reliability of gray scale/dynamic range evaluation of ultrasonography using a standardized phantom. *Ultrasonography* 2014; 33: 91-97.
4. Choi JI, Jung SE, Kim PN, et al. Quality assurance in ultrasound screening for hepatocellular carcinoma using a standardized phantom and standard clinical images: a 3-year national investigation in Korea. *J Ultrasound Med* 2014; 33: 985-995.
5. 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.
6. Bruix J, Sherman M; American Association for the Study of Liver Diseases. Management of hepatocellular carcinoma: an update. *Hepatology* 2011; 53: 1020-1022.
7. European Association for The Study of The Liver, European Organisation for Research and Treatment of Cancer. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* 2012; 56: 908-943.
8. Korean Liver Cancer Study Group (KLCSG), National Cancer Center Korea (NCC). 2014 Korean Liver Cancer Study Group-National Cancer Center Korea practice guideline for the management of hepatocellular carcinoma. *Korean J Radiol* 2015; 16: 465-522.
9. Kudo M, Matsui O, Izumi N, et al. Surveillance and diagnostic algorithm for hepatocellular carcinoma proposed by the Liver Cancer Study Group of Japan: 2014 update. *Oncology* 2014; 87 Suppl 1: 7-21.
10. Lee JM, Park JW, Choi BI. 2014 KLCSG-NCC Korea Practice Guidelines for the management of hepatocellular carcinoma: HCC diagnostic algorithm. *Dig Dis* 2014; 32: 764-777.
11. Singal A, Volk ML, Waljee A, et al. Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Aliment Pharmacol Ther* 2009; 30: 37-47.
12. Renzulli M, Golfieri R; Bologna Liver Oncology Group. Proposal of a new diagnostic algorithm for hepatocellular carcinoma based on the Japanese guidelines but adapted to the Western world for patients under surveillance for chronic liver disease. *J Gastroenterol Hepatol* 2016; 31: 69-80.
13. Kim HJ, Hong JI, Mok HJ, Lee KM. Effect of workplace-visiting nutrition education on anthropometric and clinical measures in male workers. *Clin Nutr Res* 2012; 1: 49-57.
14. Treloar D, Hawayek J, Montgomery JR, Russell W; Medical Readiness Trainer Team. On-site and distance education of emergency medicine personnel with a human patient simulator. *Mil Med* 2001; 166: 1003-1006.