Associations of Ultrasound LI-RADS Visualization Score With Examination, Sonographer, and Radiologist Factors: Retrospective Assessment in Over 10,000 Examinations

David T. Fetzer, MD[1], **Travis Browning, MD**[1,2], **Yin Xi, PhD**[1], **Takeshi Yokoo, MD, PhD**[1], **Amit G. Singal, MD**[3]

[1]Department of Radiology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-8896.

[2]Clinical Informatics, Parkland Health and Hospital System, Dallas, TX.

[3]Department of Internal Medicine, Division of Digestive and Liver Diseases, University of Texas Southwestern Medical Center, Dallas, TX.

Abstract

BACKGROUND.—When performing ultrasound (US) for hepatocellular carcinoma (HCC) screening, numerous factors may impair hepatic visualization, potentially lowering sensitivity. US LI-RADS includes a visualization score as a technical adequacy measure.

OBJECTIVE.—The purpose of this article is to identify associations between examination, sonographer, and radiologist factors and the visualization score in liver US HCC screening.

METHODS.—This retrospective study included 6598 patients (3979 men, 2619 women; mean age, 58 years) at risk for HCC who underwent a total of 10,589 liver US examinations performed by 91 sonographers and interpreted by 50 radiologists. Visualization scores (A, no or minimal limitations; B, moderate limitations; C, severe limitations) were extracted from clinical reports. Patient location (emergency department [ED], inpatient, outpatient), sonographer and radiologist liver US volumes during the study period (< 50, 50–500, > 500 examinations), and radiologist practice pattern (US, abdominal, community, interventional) were recorded. Associations with visualization scores were explored.

RESULTS.—Frequencies of visualization scores were 71.5%, 24.2%, and 4.2% for A, B, and C, respectively. Scores varied significantly (*p* < .001) between examinations performed in ED patients (49.8%, 40.1%, and 10.2%), inpatients (58.8%, 33.9%, and 7.3%), and outpatients (76.7%, 20.3%, and 2.9%). Scores also varied significantly (*p* < .001) by sonographer volume (< 50 examinations: 58.4%, 33.7%, and 7.9%; > 500 examinations: 72.9%, 22.5%, and 4.6%); reader volume (< 50 examinations: 62.9%, 29.9%, and 7.1%; > 500 examinations: 67.3%, 28.0%, and 4.7%); and reader practice pattern (US: 74.5%, 21.3%, and 4.3%; abdominal: 67.0%, 28.1%, and 4.8%; community: 75.2%, 21.9%, and 2.9%; interventional: 68.5%, 24.1%, and 7.4%). In multivariable analysis, independent predictors of score C were patient location (ED/inpatient: odds ratio [OR], 2.62; *p* < .001) and sonographer volume (< 50: OR, 1.55; *p* = .01). Among

Address correspondence to D. T. Fetzer (David.Fetzer@UTSouthwestern.edu).

sonographers performing 50 or more examinations, the percentage of outpatient examinations with score C ranged from 0.8% to 5.4%; 9/33 were above the upper 95% CI of 3.2%.

CONCLUSION.—The US LI-RADS visualization score may identify factors affecting quality of HCC screening examinations and identify outlier sonographers in terms of poor examination quality. The approach also highlights potential systematic biases among radiologists in their quality assessment process.

CLINICAL IMPACT.—These findings may be applied to guide targeted quality improvement efforts and establish best practices and performance standards for screening programs.

Keywords

HCC; LI-RADS; quality; screening; ultrasound

Hepatocellular carcinoma (HCC) is the fourth most common cause of cancer-related death worldwide, with a high proportion of cases detected at late stages when curative treatment options are no longer available [1]. Whereas median survival for early-stage HCC exceeds 5 years, advanced stages have a median survival of only 1–2 years. Accordingly, professional societies including the American Association for the Study of Liver Diseases (AASLD) have recommended semiannual surveillance in at-risk patients, including those with cirrhosis [2, 3]. Several cohort studies have shown that HCC surveillance is associated with increased early tumor detection and improved survival, even after adjusting for lead-time and length-time biases [4, 5].

HCC surveillance is typically performed using abdominal ultrasound (US), with or without serum α-fetoprotein, and has a sensitivity of 63% for early HCC detection [6]. However, US is operator dependent, with a wide variation in published sensitivity and specificity [7]. Further, visualization of the liver on US may be impaired by both intrinsic qualities of the liver and by factors extrinsic to the liver, such as obscuration by rib, lung, or bowel. Specific to HCC surveillance examinations, quality has been shown to be lower in patients with obesity and nonalcoholic steatohepatitis (NASH) as well as in patients with Child-Pugh B cirrhosis [8, 9]. Impaired visualization of the liver may increase the risk of a missed early-stage HCC or of a false-positive result, thereby increasing screening-related harms [10, 11]. Widespread adoption of alternative surveillance modalities, such as MRI, continue to be limited by cost and access concerns, and novel markers still require further validation [12, 13]. In the interim, a need exists to evaluate and optimize the effectiveness of US-based surveillance.

LI-RADS includes a module for screening and surveillance US, termed "US LI-RADS," that aims to standardize the imaging technique and reporting of liver US examinations performed in patients at risk for HCC [14–16]. US LI-RADS requires two assessments for each examination, both reported for the whole liver rather than for individual observations. The first is the US category, representing the examination's main result and reported as negative (US-1), subthreshold (US-2), or positive (US-3); each such category is associated with a particular management recommendation. The second is the US visualization score, a qualitative measure of technical adequacy (i.e., examination quality) for HCC detection that

is reported as no or minimal limitations (A), moderate limitations (B), or severe limitations (C).

Examination- and operator-level factors that influence the quality of US examinations performed for HCC screening are not well understood. Identifying potential ranges of expected performance could inform initiatives for targeted training and reeducation, provide metrics for comparing facilities, and help establish quality standards for US HCC screening programs. Therefore, the purpose of this study was to identify associations between examination, sonographer, and radiologist factors and the quality of US examinations performed for HCC screening and surveillance, as assessed by the US LI-RADS visualization score.

Methods

Patient Selection

This retrospective HIPAA-compliant cross-sectional study involved patients from two independent healthcare systems (site 1, Parkland Health and Hospital System, a community county-funded safety net system, and site 2, University of Texas Southwestern Medical Center, a tertiary state-funded university-associated hospital with multiple regional outpatient imaging centers), both served by the same radiology department. The study received institutional review board approval at both sites, with a waiver of the requirement for written informed consent.

The radiology department maintains a single repository of radiology reports and associated workflow data obtained from each site's electronic health record (EHR) and radiology information system (Epic Radiant, Epic Systems). The repository was queried for examinations with the electronic order name "US liver" (the institutions' preferred EHR order for US examinations performed for HCC screening and surveillance, as described later; hereafter referred to as liver US examinations) performed between January 29, 2017, and June 30, 2019. Other electronic order names including "US abdomen complete," "US abdomen right upper quadrant (RUQ)," and "US gallbladder" were not retrieved because these are not typically used in conjunction with HCC screening and surveillance and US LI-RADS reporting. Associated data elements for the identified examinations were retrieved, including patient age, sex, and medical history; order indication and additional comments; examination details including performing site and patient location (i.e., emergency department [ED], inpatient, or outpatient); examination start time (i.e., room entry time) and end time (i.e., room exit time); identifiers for the performing sonographer and reporting radiologist; and examination report content. Among the identified examinations, examinations were excluded for the following reasons: duplicate examination, patient age less than 18 years, patient not at high risk for HCC, and report did not contain the US LI-RADS visualization score. Patients were considered to be at high risk for HCC because of chronic hepatitis B; a clinical diagnosis of cirrhosis as determined by the order indication, associated diagnoses, or the patient's problem list in the EHR; or definitive findings of cirrhosis on prior imaging or on the ordered liver US examination.

Ultrasound LI-RADS Clinical Implementation and Reporting Workflow

The institution adopted US LI-RADS in January 2017. Immediately before adoption, dedicated training in US LI-RADS was provided to sonographers and radiologists through staff meetings and journal clubs at both sites and supplemented by online resources. Subsequently hired sonographers and radiologists received training during onboarding. Ongoing education occurred through a departmental clinical quality assurance feedback system.

At the time of implementation of US LI-RADS, a dedicated order for liver US examination was created in the EHR, intended primarily for use in patients with suspected acute or chronic liver disease, including those at risk for HCC who require screening and surveillance. Sonographers were instructed to convert other abdominal or right upper quadrant orders to this new dedicated liver US order when the order indication or comments indicated chronic liver disease, cirrhosis, or HCC. Sonographers at both sites were instructed to follow a dedicated imaging protocol for all liver US examinations that follows the US LI-RADS technical recommendations [17]. This protocol includes representative gray-scale images in longitudinal and transverse orientation through the left and right hepatic lobes, optimized to ensure complete parenchymal visualization; continuous cine acquisitions through both lobes; and high-resolution images of the hepatic capsule and underlying parenchyma. The protocol also includes assessment for indicators of portal hypertension (i.e., splenic volume, presence of ascites, and portal vein diameter, flow direction, and velocity).

The two sites employed separate groups of sonographers. All sonographers at both sites were licensed and had Registered Diagnostic Medical Sonographer (RDMS) certification. The liver US examinations were performed as part of patients' clinical care using one of the following US systems: iU22 (Philips Healthcare) with C5-1 transducer for deep imaging and L12-5 or L9-3 transducer for superficial imaging; EPIQ 5 or EPIQ 7 (Philips Healthcare) with C5-1 or C9-2, and L12-5, L12-3 or eL18-4 transducers, respectively; or ACUSON Sequoia (Siemens Healthineers) with 5C1 curvilinear and 10L4 or 14L5 transducers, respectively. Images were submitted to the PACS of the site where the examination was performed. The liver US examinations were typically performed during routine daytime hours. For such examinations, sonographers directly discussed the findings with the interpreting radiologist, commenting on any factors compromising examination quality. For all examinations, sonographers entered into the EHR brief free-form notes, which likewise commented on any factors compromising examination quality.

Examinations were interpreted by board-certified radiologists (hereafter, "readers"), who commonly worked at both sites. The radiologists were classified as having one of four clinical practice patterns: US, abdominal, community, or interventional. All radiologists with US and abdominal practice patterns had completed abdominal imaging fellowships; those with a US practice pattern spent at least 50% of their clinical time interpreting US examinations. Radiologists with a community practice pattern may have completed an abdominal imaging fellowship but interpreted a wide range of examination types across multiple subspecialties. Interventional radiologists routinely interpreted diagnostic vascular imaging, including Doppler US examinations of the liver. When interpreting Doppler US

examinations of the liver, interventional radiologists may also interpret a concurrently performed liver US examination.

All liver US examinations were reported using a US LI-RADS structured template [17]. When using the template, readers are prompted to select both a US category (US-1, US-2, or US-3) and a visualization score (A, B, or C) using picklists without a default selection. The US category is considered optional and intended only to be completed in patients at high risk. The visualization score is considered a mandatory reporting element for all liver US examinations but still may be bypassed at the reporting radiologist's discretion. Although the interpreting radiologist receives comments from the performing technologist regarding factors compromising quality, the ultimate reported visualization score is at the discretion of the interpreting radiologist.

Data Analysis

All examinations were classified in terms of the site performed (site 1 or site 2) and patient location (ED, inpatient, or outpatient). Sonographers were classified in terms of years since training (0–3, 4–10, or > 10 years) and the number of liver US examinations performed during the study period in patients considered at high risk (< 50, 50–500, or > 500). Readers were likewise classified in terms of years since training (0–3, 4–10, or > 10 years), number of examinations interpreted during the study period (< 50, 50–500, or > 500), and by practice pattern (US, abdominal, community, or interventional). (The thresholds of 50 and 500 examinations during the study period used for these categorizations correspond with annualized volumes of 21 and 207 examinations, respectively, considering the investigation's 29-month study period). Examination reports were parsed for the US category (US-1, US-2, or US-3) and visualization score (A, B, or C). Examination room times were computed as the time in minutes between room entry and room exit.

Descriptive statistics were calculated as medians and inter-quartile ranges (IQRs) for continuous variables, and as counts and frequencies for categoric variables. When performing analyses involving examination times, examinations with the highest and lowest 5% of values were excluded (presumed to include data entry errors, including such examples as examination times of 0 minutes or 3 days). Examinations at the two sites were compared in terms of patient location, sonographer volume, sonographer experience, reader practice pattern, reader volume, and reader experience. Associations were also assessed between the distribution of visualization scores and the various examination characteristics, and between the distribution of visualization scores and US categories. Comparisons were performed using a chi-square test. A multivariable generalized estimating equation model was used to identify independent predictors of a visualization score of C among those characteristics that had shown a significant association with visualization score in univariable comparisons. The model was adjusted for multiple examinations in individual patients using a working correlation matrix with autoregressive structure. The model combined ED and inpatient as a single patient location. The model did not include examination room time. The Tukey-Kramer method was used to adjust p values from the multivariable model for multiple comparisons.

To visually display performance distribution among sonographers, a plot was created showing the distribution of percentages of examinations with a visualization score of C across sonographers, identifying those with a percentage greater than the upper 95% CI for the sample. An additional plot was created showing the distribution of percentages of visualization score C as a function of each sonographer's volume, with the upper one-sided 95% CI conditional on the sonographer's volume according to a beta distribution (i.e., a continuous probability distribution from 0 to 1). Both plots included only sonographers who performed at least 50 examinations and only outpatient examinations given observed significant variation in visualization scores according to patient location (as described in the Results) and variation among sonographers in terms of their primary work assignments (ED/inpatient vs outpatient).

Data from the two sites were pooled for all analyses other than those analyses comparing the two sites. The *p* values were reported with the significance level set at .05. Descriptive statistics were calculated using an in-house application using R and Shiny (R Foundation for Statistical Computing Development Core Team). Other analyses were performed in SAS 9.4 (SAS Institute).

Results

Study Sample

Selection of the study sample is shown in Figure 1. Of 19,966 liver US examinations performed during the study period, examinations were excluded for the following reasons: duplicate examination ($n = 146$), patient age under 18 years ($n = 60$), and patient not at high risk for HCC ($n = 8735$). These exclusions resulted in 11,025 potentially eligible examinations. Of these, an additional 436/11,025 (4.0%) examinations were excluded because the report did not contain the visualization score, resulting in a final sample of 10,589 examinations. The 10,589 examinations were performed in 6598 unique at-risk patients (3979 men, 2619 women; mean age, 58 years [IQR, 50–64 years]), including 675 (10.2%) with chronic hepatitis B, 4879 (73.9%) with a clinical diagnosis of cirrhosis, and 4552 (69.0%) with definitive imaging findings of cirrhosis. A total of 2415/6598 (36.6%) patients underwent multiple examinations within the sample (maximum of eight examinations in any individual patient). The patient location was available for 10,511/10,589 (99.3%) examinations (ED: 699 [6.7%]; inpatient: 2029 [19.3%]; outpatient: 7783 [74.0%]). The median examination room time was 31.0 minutes (95% CI, 16.0–79.0 minutes).

The examinations were performed by one of 91 sonographers (examination volume of < 50 in 38, 50–500 in 50, > 500 in three; experience of 0–3 years in 12, 4–10 years in 36, > 10 years in 40, unknown in three). The 91 sonographers had a median volume of 71 examinations (range, 1–725 examinations; IQR, 95 examinations) during the study period (annualized volume of 25.7 examinations) and a median experience of 8.0 years (range, < 1 to 35 years; IQR, 12.0 years). The examinations were interpreted by one of 50 radiologists (practice pattern of US in eight, abdominal in 16, community in 13, and interventional in 13; examination volume of < 50 in 20, 50–500 in 24, > 500 in six; experience of 0–3 years in two, 4–10 years in 15, > 10 years in 32). The 50 readers had a median volume of 160 examinations (range, 1–1051 examinations; IQR, 330 examinations) during the study period

(annualized volume of 55.0 examinations) and a median experience of 12.2 years (range, 2–40 years; IQR, 11 years).

A total of 7813/10,589 (73.8%) examinations were performed at site 1 and 2776/10,589 (26.2%) examinations at site 2. Examinations from the two sites were significantly different (all $p < .05$) in terms of patient location (e.g., outpatient in 76.5% vs 67.0%), sonographer volume (e.g., > 500 examinations in 24.1% vs 0.0%), sonographer experience (e.g., > 10 years in 49.0% vs 65.1%), reader practice pattern (e.g., abdominal in 51.2% vs 12.3% and community in 8.9% vs 49.8%), reader volume (e.g., > 500 in 48.3% vs 17.4%), reader experience (e.g., > 10 years, 80.6% vs 84.6%), and examination room time (median, 31 minutes vs 34 minutes) (Table 1).

Ultrasound Visualization Scores

The visualization score was A (minimal or no visual limitations) in 71.5% (7574/10,589), B (moderate limitations) in 24.2% (2566/10,589), and C (severe limitations) in 4.2% (449/10,589) of examinations (Table 2). The visualization score was significantly different ($p = .03$) between site 1 (scores of A, B, and C in 70.8%, 24.9%, and 4.3%) and site 2 (73.5%, 22.5%, 4.0%). The score also varied significantly ($p < .001$) between examinations performed in ED patients (49.8%, 40.1%, and 10.2%), inpatients (58.8%, 33.9%, and 7.3%), and outpatients (76.7%, 20.3%, and 2.9%). The score varied significantly ($p < .001$) with respect to sonographer volume (< 50 examinations: 58.4%, 33.7%, and 7.9%; > 500 examinations: 72.9%, 22.5%, and 4.6%), reader practice pattern (US: 74.5%, 21.3%, and 4.3%; abdominal: 67.0%, 28.1%, and 4.8%; community: 75.2%, 21.9%, and 2.9%; interventional: 68.5%, 24.1%, and 7.4%), reader volume (< 50 examinations: 62.9%, 29.9%, and 7.1%; > 500 examinations: 67.3%, 28.0%, and 4.7%), and reader experience (0–3 years: 73.8%, 19.1%, and 7.1%; > 10 years: 72.0%, 23.9%, and 4.2%). The visualization score was not significantly associated with sonographer experience ($p = .05$). The median examination room time varied significantly ($p < .001$) between examinations with visualization score of A (31.0 minutes), B (33.0 minutes), and C (36.0 minutes). The US category was reported 5748/10,589 (54.3%) examinations; in these examinations, the US category and visualization score showed no significant association ($p = .10$) (Table 3).

The multivariable analysis was performed in a total of 10,278 examinations with complete data. Significant independent predictors of a visualization score of C were patient location (ED/inpatient: OR, 2.62 relative to outpatients; $p < .001$) and sonographer volume (< 50 relative to 50–500: OR, 1.55; $p = .01$; 50–500 relative to > 500: OR, 0.66; $p = .007$). Site, reader practice pattern, reader volume, and reader experience were not independent predictors of visualization score C (all $p > .05$) (Table 4).

Figure 2 shows an example of a patient who had a poorer visualization score when imaged in the ED setting than when imaged as an outpatient. Figure 3 shows an example of a patient who had a poorer visualization score when imaged by a sonographer who had performed fewer than 50 examinations than when imaged by a sonographer who had performed more than 500 examinations in the study period. Figure 4 shows an example of a patient who had a poorer visualization score for an examination interpreted by a radiologist with a US

practice pattern than for an examination interpreted by a radiologist with an abdominal practice pattern.

Figure 5A shows a plot of the distribution of the percentage of outpatient examinations with a visualization score of C by the 33 individual sonographers who performed 50 or more outpatient examinations during the study period. By sonographer, this percentage ranged from 0.8% to 5.4%. A total of 9/33 (27.3%) sonographers had a percentile greater than the upper 95% CI of 3.2%, indicating sonographers with potential outlier performance in terms of frequency of visualization score of C. Figure 5B depicts the same group of sonographers' frequencies of a visualization score of C, though plotted as a function of the individual sonographer's volume, with the upper 95% CI being conditional on the sonographer's volume rather than uniform across the group. By this method, a total of 2/33 (6.1%) sonographers had a percentile greater than the upper 95% CI.

Discussion

In this study of more than 10,000 US examinations performed for HCC screening or surveillance in at-risk patients, we explored associations between a qualitative measure of examination quality (the US LI-RADS visualization score) and examination, sonographer, and radiologist factors. The visualization score was worse for examinations performed in the ED or inpatient setting and for examinations performed by less experienced sonographers. Individual sonographers who were outliers in terms of poor visualization rates were identified. The data could inform improvement measures including feedback programs and targeted retraining to reduce examination variability. Visualization scores also varied significantly across radiologists according to practice pattern or experience, suggesting observer bias as well in the quality assessment process.

US as a primary screening tool for HCC has been scrutinized given the modality's wide range of reported sensitivities [4, 6]. Many factors, including operator dependency, may contribute to this variability across screening US examinations. The observed frequencies of the visualization scores in this study mirror those recently reported [18, 19], as most examinations were categorized as having no or minimal limitations that would be expected to affect sensitivity (visualization score A). However, 24.2% of examinations had moderate limitations, and 4.2% had severe limitations. Visualization scores were particularly worse in ED and inpatient settings compared with outpatient settings, supporting the recommendation by groups such as the US LI-RADS Working Group to limit HCC screening to medically stable outpatients [17]. Nonetheless, additional factors also significantly impacted the visualization score, highlighting the opportunity for continued quality improvement.

We observed a critical role of the sonographer in impacting the quality of liver US examinations. Specifically, high rates of poor visualization were observed among sonographers who performed fewer than 50 high-risk liver US examinations during the study period (i.e., fewer than 21 examinations annually). Thus, when establishing an HCC surveillance program, adoption of a minimum annual volume for sonographers may be warranted, similar to requirements for technologists to attain mammography accreditation [20]. Rates of poor visualization were similar between sonographers with examination

volumes of 50–500 and greater than 500. This may reflect a threshold effect, whereby once sonographers have accrued a certain case volume, further case exposure does not lead to continued improvement. It is also possible that sonographers who accumulated very high liver US case volumes were selected or volunteered to perform the most challenging examinations.

We also observed associations of visualization scores and characteristics of the interpreting radiologists. In the absence of an external reference standard for the score or follow-up data regarding potentially missed HCCs, it is not possible to assess the accuracy of the assigned scores. Nonetheless, the presence of the variation indicates possible systematic biases by the interpreting radiologist. For example, radiologists with varying practice patterns, volumes at the given center, or experience levels, may have varying levels of trust with the performing sonographers or with the modality overall. The visualization score is inherently a subjective assessment of examination quality. In its current form, the score represents an amalgamation of factors that may affect parenchymal visualization, without subcategories or qualifiers for specific limiting factors. In the absence of an objective metric, it remains important to recognize the potential reader bias in the assessment and to standardize the assessments within a practice so as to normalize an organization's reporting. Feedback to individual radiologists may allow readers to further calibrate their visualization scoring to limit identified reader biases.

Examinations with poor visualization had longer mean examination lengths. This potentially reflects the additional time needed in technically challenging examinations to address patient-specific limitations and attempt image optimization. Given this observation, shorter examination times do not appear to be a marker of a rushed or abbreviated examination that may portend poorer quality. Potentially, in patients with a poor visualization score reported on a prior examination, subsequent examinations could be scheduled for a longer time slot or be assigned to a more experienced sonographer.

Our study highlights the utility of standardized methods such as LI-RADS and structured reporting, from which distinct data elements may be automatically extracted, allowing large numbers of clinical reports to be analyzed without manual review. High compliance with the requirement to assign a visualization score (missing in only 4.0% of examinations in this study) further facilitates quality improvement efforts. Programmatically, the examination quality data may be used to guide individual- and site-level quality assurance. For example, frequencies of visualization scores could be used to compare sonographers, identifying those with higher rates of poorly scored examinations for possible intervention. We identified sonographers with frequencies of visualization score C above the upper 95% CI (whether or not adjusting for sonographer volume); these individuals could be considered for targeted improvement efforts. Likewise, sonographers performing well could be surveyed for useful scanning techniques and workflow approaches. Further, because hepatic visualization is assumed to be a key determinant of sensitivity for HCC by US, visualization scores could also be used as overall site measures, compared for the given facility to national benchmarks, similar to the quality assurance process for mammography facilities [21].

We acknowledge study limitations. First, the study was retrospective in nature, relying on the original clinical interpretation and reporting. Therefore, examination-level inter- and intraob-server variability in visualization scores were not assessed. Nonetheless, good interreader agreement of the US LI-RADS visualization score has been recently described [22]. Second, various patient-level factors, such as Child-Pugh score and obesity, have been previously associated with poorer visualization [8, 9]. However, the patient-level factors were not readily available through our data extraction process and thus not assessed in our investigation. Similarly, visualization may be affected by the US scanner model, age, and software version; such information also is not captured in the departmental data warehouse. Third, although we explored associations with visualization scores, we do not know the impact of such associations on actual performance for HCC detection given the lack of follow-up data. Fourth, the examinations were performed at two different sites, which were different for essentially all study variables. Nonetheless, the site was not an independent predictor of visualization score in the multivariable model. Finally, the discussion between the sonographer and radiologist that happens for examinations occurring during daytime hours may represent a workflow that is not followed at other radiology practices. In such situations, the sonographer's assessment of liver visualization could be included in the notes or worksheet completed by the sonographer for the radiologist's later reference at the time of interpretation.

In conclusion, we leveraged the US LI-RADS visualization score to perform a retrospective assessment of the technical quality of a large volume of liver US examinations performed in patients at high risk for HCC. Examinations of lower quality included those performed in ED or inpatient settings and those performed by less experienced sonographers. The approach allowed the identification of individual sonographers who were outliers in terms of poor examination quality and who could be targeted for improvement efforts. Potential systematic biases in quality assessment were also identified among interpreting radiologists. These findings may be applied to help establish best practices and performance standards for HCC screening and surveillance programs.

Acknowledgments

References

1. Moon AM, Singal AG, Tapper EB. Contemporary epidemiology of chronic liver disease and cirrhosis. Clin Gastroenterol Hepatol 2020; 18:2650–2666 [PubMed: 31401364]
2. Marrero JA, Kulik LM, Sirlin CB, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. Hepatology 2018; 68:723–750 [PubMed: 29624699]

3. European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2018; 69:182–236 [Erratum in J Hepatol 2019; 70:817] [PubMed: 29628281]

4. Singal AG, Pillai A, Tiro J. Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: a meta-analysis. PLoS Med 2014; 11:e1001624 [PubMed: 24691105]

5. Choi DT, Kum HC, Park S, et al. Hepatocellular carcinoma screening is associated with increased survival of patients with cirrhosis. Clin Gastroenterol Hepatol 2019; 17:976–987.e4 [PubMed: 30616961]

6. Tzartzeva K, Obi J, Rich NE, et al. Surveillance imaging and alpha fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis: a meta-analysis. Gastroenterology 2018; 154:1706–1718.e1701 [PubMed: 29425931]

7. Singal AG, Nehra M, Adams-Huet B, et al. Detection of hepatocellular carcinoma at advanced stages among patients in the HALT-C trial: where did surveillance fail? Am J Gastroenterol 2013; 108:425–432 [PubMed: 23337478]

8. Simmons O, Fetzer DT, Yokoo T, et al. Predictors of adequate ultrasound quality for hepatocellular carcinoma surveillance in patients with cirrhosis. Aliment Pharmacol Ther 2017; 45:169–177 [PubMed: 27862091]

9. Del Poggio P, Olmi S, Ciccarese F, et al. Factors that affect efficacy of ultrasound surveillance for early stage hepatocellular carcinoma in patients with cirrhosis. Clin Gastroenterol Hepatol 2014; 12:1927–1933.e2 [PubMed: 24582947]

10. Atiq O, Tiro J, Yopp AC, et al. An assessment of benefits and harms of hepatocellular carcinoma surveillance in patients with cirrhosis. Hepatology 2017; 65:1196–1205 [PubMed: 27775821]

11. Singal AG, Patibandla S, Obi J, et al. Benefits and harms of hepatocellular carcinoma surveillance in a prospective cohort of patients with cirrhosis. Clin Gastroenterol Hepatol 2021; 19:1925–1932.e1 [PubMed: 32920214]

12. Singal AG, Hoshida Y, Pinato DJ, et al. International Liver Cancer Association (ILCA) white paper on biomarker development for hepatocellular carcinoma. Gastroenterology 2021; 160:2572–2584 [PubMed: 33705745]

13. Singal AG, Lok AS, Feng Z, Kanwal F, Parikh ND. Conceptual model for the hepatocellular carcinoma screening continuum: current status and research agenda. Clin Gastroenterol Hepatol 2022; 20:9–18 [PubMed: 32961340]

14. Morgan TA, Maturen KE, Dahiya N, Sun MRM, Kamaya A; American College of Radiology Ultrasound Liver Imaging and Reporting Data System (US LI-RADS) Working Group. US LI-RADS: ultrasound liver imaging reporting and data system for screening and surveillance of hepatocellular carcinoma. Abdom Radiol (NY) 2018; 43:41–55 [PubMed: 28936543]

15. Fetzer DT, Rodgers SK, Harris AC, et al. Screening and surveillance of hepatocellular carcinoma: an introduction to Ultrasound Liver Imaging Reporting and Data System. Radiol Clin North Am 2017; 55:1197–1209 [PubMed: 28991560]

16. Rodgers SK, Fetzer DT, Gabriel H, et al. Role of US LI-RADS in the LI-RADS algorithm. RadioGraphics 2019; 39:690–708 [PubMed: 31059393]

17. American College of Radiology website. Ultrasound LI-RADS v2017. www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/LI-RADS/UltrasoundLI-RADS-v2017. Accessed November 13, 2020

18. Millet JD, Kamaya A, Choi HH, et al. ACR Ultrasound Liver Reporting and Data System: multicenter assessment of clinical performance at one year. J Am Coll Radiol 2019; 16:1656–1662 [PubMed: 31173745]

19. Son JH, Choi SH, Kim SY, et al. Validation of US Liver Imaging Reporting and Data System version 2017 in patients at high risk for hepatocellular carcinoma. Radiology 2019; 292:390–397 [PubMed: 31210614]

20. American College of Radiology Accreditation Support website. Radiologic technologist: mammography (revised 12-12-19). accreditationsupport.acr.org/support/solutions/articles/11000049779-radiologic-technologist-mammography. Revised December 12, 2019. Accessed June 2, 2021

21. D'Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. American College of Radiology, 2013. www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads#FollowUpandMonitoring. Accessed June 2, 2021

22. Tiyarattanachai T, Bird KN, Lo EC, et al. Ultrasound Liver Imaging Reporting and Data System (US LI-RADS) visualization score: a reliability analysis on inter-reader agreement. Abdom Radiol (NY) 2021; 46:5134–5141 [PubMed: 34228197]

HIGHLIGHTS

Key Finding

- Independent predictors of visualization score C on HCC screening liver US examinations included patient location (ED/inpatient: OR, 2.62) and sonographer volume (< 50 liver US examinations: OR, 1.55). Visualization scores also showed significant univariable associations with radiologist practice patterns (score C in 2.9% of community radiologists vs 4.3–4.8% of US and abdominal radiologists).

Importance

- Liver US quality assessment is impacted by operator factors and reader biases. US LI-RADS visualization scores may facilitate performance standards and targeted quality improvement efforts.
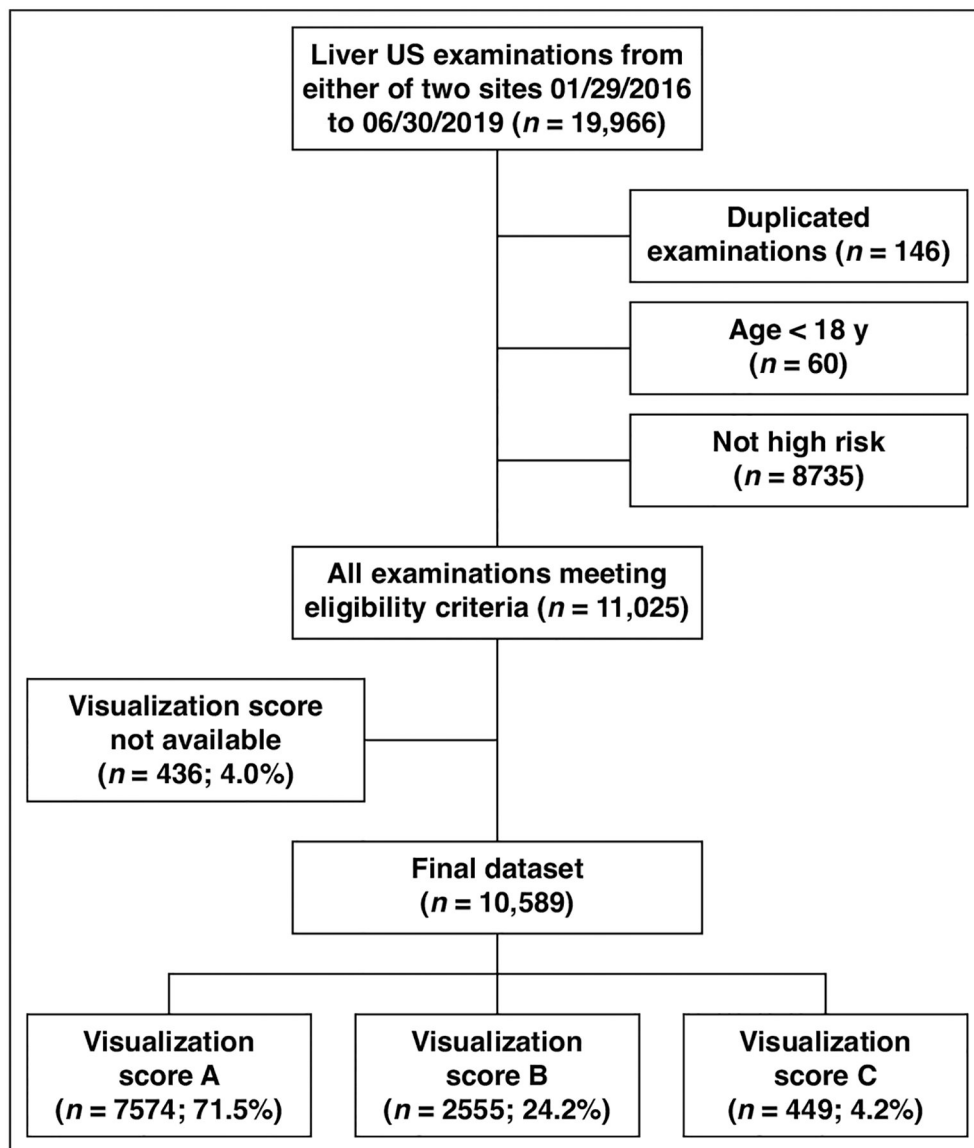
Fig. 1—.

Ultrasound (US) examination characteristics flowchart. All counts represent numbers of examinations and include emergency department, inpatient, and outpatient examinations.
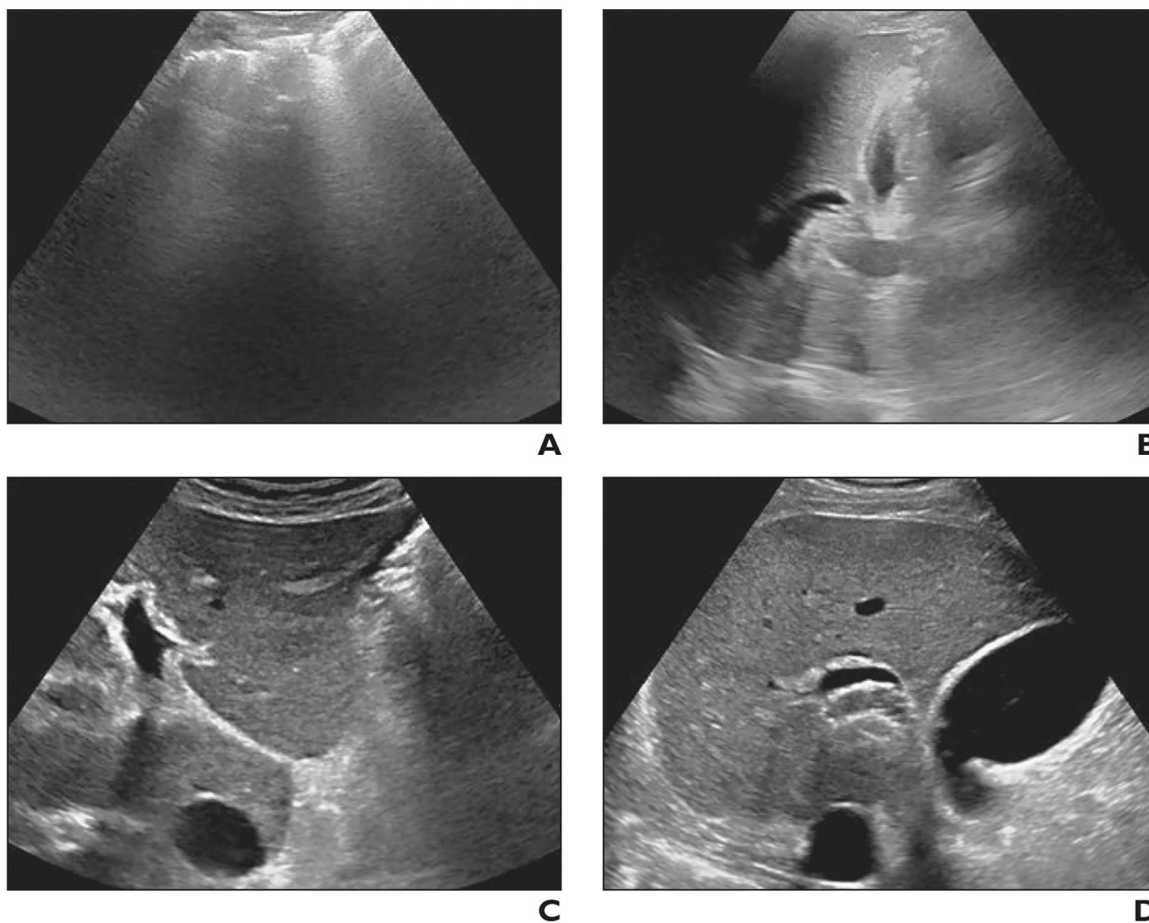
Fig. 2—.

Difference in visualization score between scan locations in 52-year-old man with nonalcoholic steatohepatitis–related cirrhosis. Examinations were performed by high-volume sonographers using same scanner model and interpreted by community radiologists.

A and **B,** Transverse gray-scale ultrasound image of left lobe (**A**) at initial liver ultrasound in emergency department shows complete obscuration of liver by bowel gas. Longitudinal image of right lobe (**B**) shows parenchyma largely obscured by lung and rib. Visualization score C was assigned.

C and **D,** Gray-scale images of left (**C**) and right (**D**) lobes 2 months later when patient underwent liver ultrasound as outpatient show marked improvement in image quality. Visualization score A was assigned.
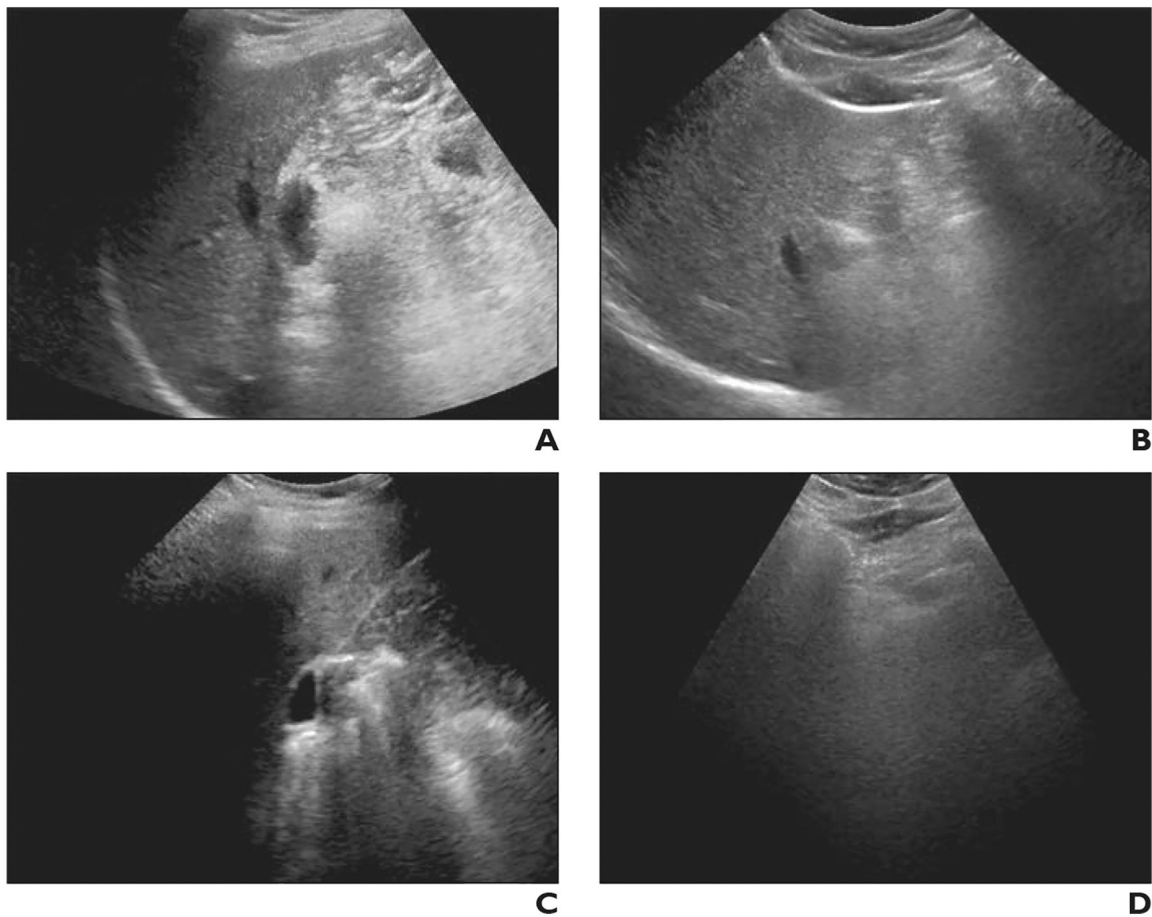
Fig. 3—.

Difference in visualization score between sonographers with different examination volumes in 47-year-old man with chronic hepatitis B. Both examinations were performed in outpatient setting using same scanner model and were interpreted by abdominal radiologists.

A and **B,** Gray-scale ultrasound images of right (**A**) and left (**B**) lobes of liver, obtained from high-volume sonographer (> 500 examinations performed during study period), show near-complete visualization of hepatic parenchyma. Visualization score A was assigned.

C and **D,** Gray-scale images of right (**C**) and left (**D**) lobes 6 months later obtained by low-volume sonographer (< 50 examinations) show that nearly entire right lobe is obscured by rib (**C**) and left lobe obscured by bowel (**D**). Visualization score C was assigned.
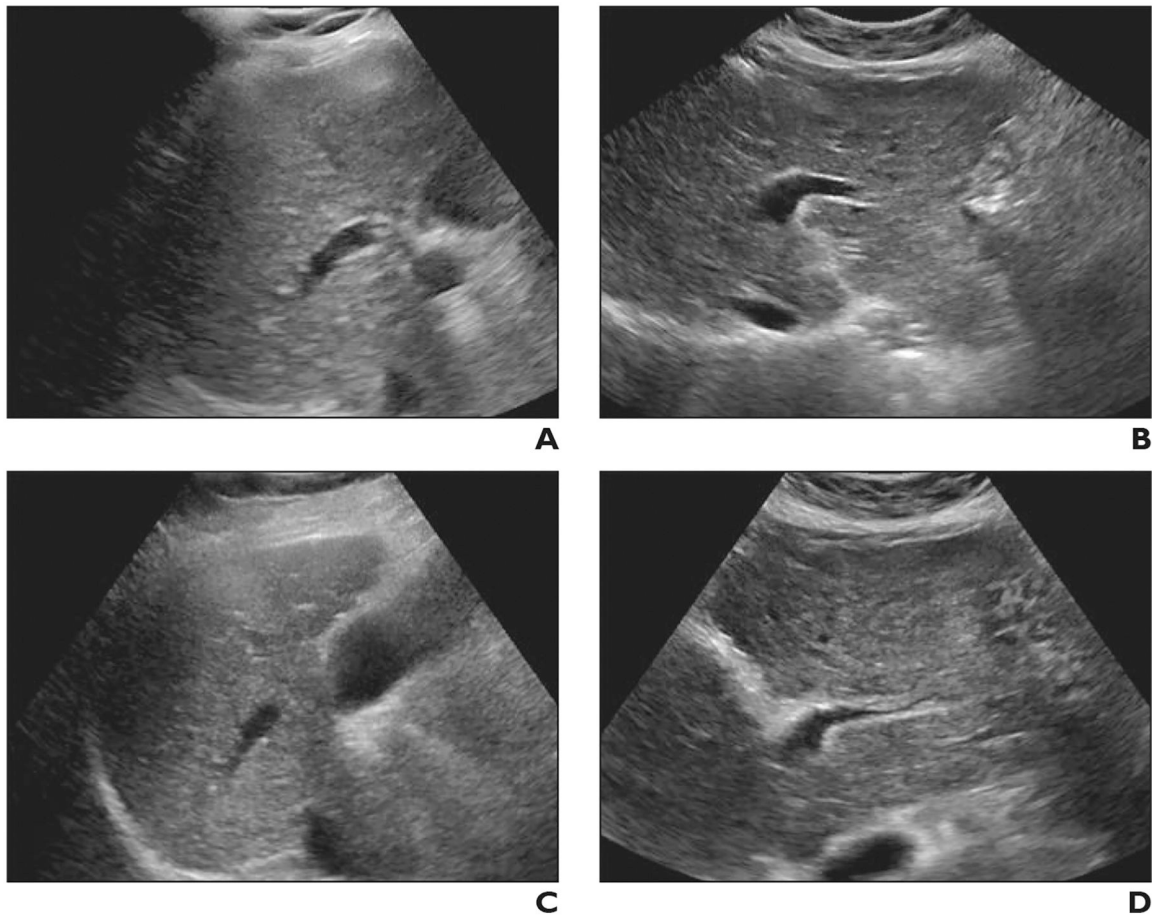
Fig. 4—.

Difference in visualization score between radiologists with different practice patterns in 60-year-old woman with cirrhosis. Both examinations were performed in outpatient setting by same sonographer using same scanner model.

A and **B,** Gray-scale ultrasound images of right (**A**) and left (**B**) lobes of liver obtained from high-volume sonographer (> 500 examinations during study period) show near-complete visualization of hepatic parenchyma with mild rib shadowing and diffusely coarse, heterogeneous parenchyma. Visualization score A was assigned by reader with ultrasound practice pattern.

C and **D,** Two months later, patient underwent liver ultrasound that was interpreted by radiologist with abdominal practice pattern. Gray-scale images of right (**C**) and left (**D**) lobe show similar findings, however, visualization score C was assigned.
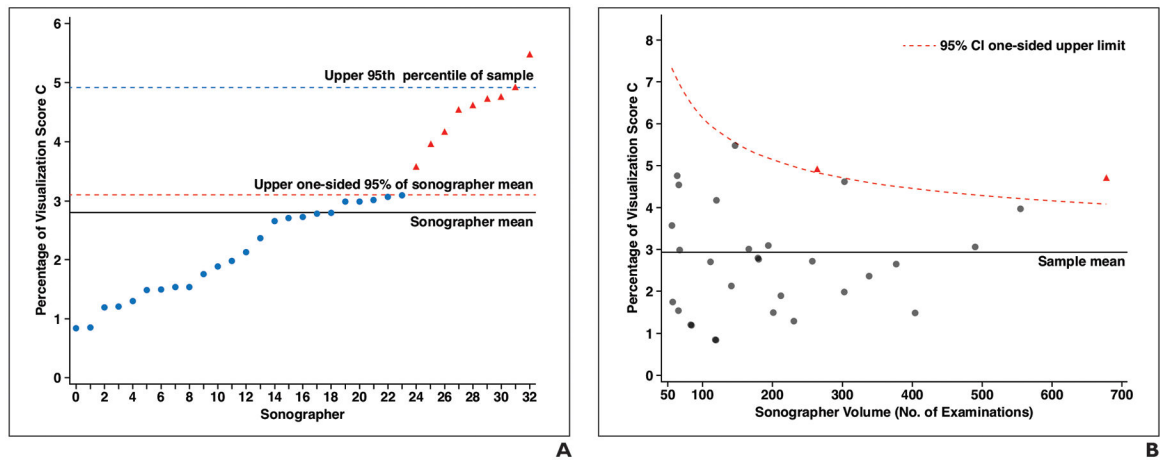
Fig 5—.

Graphs of distribution of outpatient examinations with visualization score of C among sonographers who performed 50 or more examinations. Each point corresponds with individual sonographer; those shown as red triangles can be considered outliers among sonographers in terms of higher percentage of visualization score of C according to analytic method used for each plot.

A, Graph shows distribution of percentage of outpatient examinations with visualization score of C by sonographer ordered from lowest to highest percentage. Each *x*-axis point indicates one of 33 sonographers who performed 50 or more examinations. Solid line indicates mean percentage (2.9%), dashed red line indicates upper one-sided 95% CI of mean, and dashed blue line indicates 95th percentile of sample (3.2%).

B, Graph shows distribution of percentage of outpatient examinations with visualization score of C by sonographer, among sonographers who performed 50 or more examinations, shown as function of sonographer volume. Solid line indicates sonographer mean percentage (2.9%). Dashed red curve indicates upper one-sided 95% CI conditional on sonographer volume according to beta distribution.

TABLE 1:

Comparison of Study Variables Between Sites

Variable, Level	Site 1 (*n* = 7813)	Site 2 (*n* = 2776)	All (*n* = 10,589)	*p*
Patient location				<.001
ED	685 (8.8)	14 (0.5)	699 (6.7)	
Inpatient	1154 (14.8)	875 (32.4)	2029 (19.3)	
Outpatient	5974 (76.5)	1809 (67.0)	7783 (74.0)	
Total no.	7813	2698	10.511	
Sonographer				
Volume (examinations)				<.001
< 50	493 (6.4)	530 (19.2)	1023 (9.8)	
50–500	5327 (69.4)	2237 (80.8)	7564 (72.5)	
> 500	1853 (24.1)	0 (0.0)	1853 (17.7)	
Total no.	7673	2767	10,440	
Experience (y)				<.001
0–3	1129 (14.9)	55 (2.0)	1184 (11.4)	
4–10	2744 (36.1)	912 (33.0)	3656 (35.3)	
> 10	3718 (49.0)	1800 (65.1)	5518 (53.3)	
Total no.	7591	2767	10,358	
Reader				
Practice pattern				<.001
Ultrasound	3113 (39.9)	997 (35.9)	4110 (38.8)	
Abdominal	4003 (51.2)	342 (12.3)	4345 (41.0)	
Community	695 (8.9)	1382 (49.8)	2077 (19.6)	
Interventional	0 (0.0)	54 (1.9)	54 (0.5)	
Volume (examinations)				<.001
< 50	66 (0.8)	131 (4.7)	197 (1.9)	
50–500	3976 (50.9)	2161 (77.9)	6137 (58.0)	
> 500	3769 (48.3)	483 (17.4)	4252 (40.2)	
Experience (y)				<.001
0–3	125 (1.6)	16 (0.6)	141 (1.3)	

Variable, Level	Site 1 (n = 7813)	Site 2 (n = 2776)	All (n = 10,589)	p
4–10	1394 (17.8)	410 (14.8)	1804 (17.0)	
> 10	6292 (80.6)	2349 (84.6)	8641 (81.6)	
Total no.	7811	2775	10,586	
Median (95% CI) time in examination room (min) (n = 8588)	31.0 (16.0–77.0)	34.0 (6.0–94.0)	31.0 (16.0–79.0)	< .001 [a]

Note—Unless otherwise indicated, values are expressed as number of examinations, with percentages in parentheses. Counts do not sum to 10,589 for individual variables because of missing data. Percentages represent distributions across subsets for each site (i.e., down columns). p values were calculated using a chi-square test unless otherwise indicated. ED = emergency department.

[a] Computed using Wilcoxon rank-sum test.

TABLE 2:

Comparison of Ultrasound LI-RADS Visualization Scores for Study Variables

Variable, Level	No. of Examinations	Score A	Score B	Score C	p
Total	10,589	7574 (71.5)	2566 (24.2)	449 (4.2)	
Site					.03
1	7813	5533 (70.8)	1942 (24.9)	338 (4.3)	
2	2776	2041 (73.5)	624 (22.5)	111 (4.0)	
Patient location	10,511				< .001
ED	699	348 (49.8)	280 (40.1)	71 (10.2)	
Inpatient	2029	1193 (58.8)	688 (33.9)	148 (7.3)	
Outpatient	7783	5971 (76.7)	1583 (20.3)	229 (2.9)	
Sonographer					
Volume (examinations)	10,440				< .001
< 50	1023	597 (58.4)	345 (33.7)	81 (7.9)	
50–500	7564	5548 (73.3)	1745 (23.1)	271 (3.6)	
> 500	1853	1350 (72.9)	417 (22.5)	86 (4.6)	
Experience (y)	10,358				.05
0–3	1184	852 (72.0)	293 (24.7)	39 (3.3)	
4–10	3656	2572 (70.4)	910 (24.9)	174 (4.8)	
> 10	5518	4006 (72.6)	1290 (23.4)	222 (4.0)	
Reader					
Practice pattern	10,586				< .001
Ultrasound	4110	3060 (74.5)	875 (21.3)	175 (4.3)	
Abdominal	4345	2913 (67.0)	1223 (28.1)	209 (4.8)	
Community	2077	1562 (75.2)	454 (21.9)	61 (2.9)	
Interventional	54	37 (68.5)	13 (24.1)	4 (7.4)	
Volume (examinations)	10,586				< .001
< 50	197	124 (62.9)	59 (29.9)	14 (7.1)	
50–500	6137	4587 (74.7)	1314 (21.4)	236 (3.8)	
> 500	4252	2861 (67.3)	1192 (28.0)	199 (4.7)	
Experience (y)	10,586				.04
0–3	141	104 (73.8)	27 (19.1)	10 (7.1)	

Variable, Level	No. of Examinations	Score A	Score B	Score C	*p*
4–10	1804	1249 (69.2)	477 (26.4)	78 (4.3)	
> 10	8641	6219 (72.0)	2061 (23.9)	361 (4.2)	
Median (95% CI) examination room time (min)	8588	31.0 (15.0–77.0)	33.0 (16.0–83.0)	36.0 (16.0–79.0)	< .001 [a]

Note—Unless otherwise indicated, values are expressed as number of examinations, with percentages in parentheses. Percentages represent distributions of visualization scores for each level of each variable (i.e., across rows). Counts do not sum to 10,589 for individual variables because of missing data. *p* values were calculated using a chi-square test unless otherwise indicated. ED = emergency department.

[a]Computed using Wilcoxon rank-sum test.

TABLE 3:

Comparison of US LI-RADS Visualization Scores for US Categories

US Category	Total No.	Score A	Score B	Score C
1	4993	3563 (71.4)	1288 (25.8)	142 (2.8)
2	400	287 (71.8)	105 (26.2)	8 (2.0)
3	355	231 (65.1)	110 (31.0)	14 (3.9)

Note—Values are expressed as number of examinations, with percentages in parentheses. Percentages represent distributions of visualization scores for category (i.e., across rows). Distributions of visualization scores not statistically significant between categories ($p = .10$). US = ultrasound.

TABLE 4:

Multivariable Analysis of Predictors of an Ultrasound LI-RADS Visualization Score of C

Comparison	OR	95% CI	p [a]
Site 2 vs site 1	1.01	0.75–1.35	.97
Outpatient vs ED/inpatient	2.62	2.08–3.29	< .001
Sonographer volume (examinations)			
50–500 vs < 50	1.55	1.18–2.05	.01
> 500 vs < 50	1.03	0.71–1.49	.99
> 500 vs 50–500	0.66	0.51–0.86	.007
Reader practice pattern			
Community vs ultrasound	1.38	0.99–1.92	.22
Community vs abdominal	1.39	1.01–1.91	.17
Community vs interventional	1.32	0.42–4.21	.96
Reader volume (examinations)			
> 500 vs < 50	1.13	0.56–2.29	.94
> 500 vs 50–500	0.85	0.68–1.06	.30
Reader experience (y)			
> 10 vs < 3	1.07	0.46–2.49	.99
> 10 vs 3–10	0.94	0.72–1.22	.88

Note—Model includes 10,278 examinations with complete data. First variable listed is reference variable for analysis. ED = emergency department, OR = odds ratio.

[a] Adjusted by Tukey-Kramer method.