THE UNIVERSITY OF CHICAGO

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Longitudinal Imaging Performance Assessment of Grayscale Breast Ultrasound Using the Random Hypoechoic Sphere Phantom

Baihui Yu, PhD[1]; Dufan Wu, PhD[2]; Cristel Baiu, MS[3]; Zheng Feng Lu, PhD[1].
[1]The University of Chicago, [2]The Ohio State University, [3]University of Wisconsin - Madison.

ABSTRACT

Breast ultrasound is widely used for detecting breast abnormalities and for guiding biopsies, yet conventional QC methods offer limited insight to monitor longitudinal drifts in image quality or detect unintentional imaging protocol changes that affect lesion detectability.

Our automated analysis algorithm with freehand scanning of a random hypoechoic sphere phantom (RHSP) provides a quantitative assessment via lesion signal-to-noise ratio (LSNR) and makes objective assessment feasible for frequent routine QC. This study aims to investigate the efficacy and efficiency of this approach for monitoring clinical image quality and longitudinal stability in breast ultrasound systems.

Results reflected the image quality difference between the two breast ultrasound scanners, and indicated no systematic LSNR drift and low variability (< 5%) over a twenty-week period. These findings support the RHSP as a practical tool for monitoring breast ultrasound image quality through routine QC.

CONTACT

Baihui Yu, PhD, DABR
The University of Chicago
5841 S. Maryland Avenue, MC 2026 | Rm. M-512 | Chicago, IL 60637
Email: Baihui.Yu@bsd.uchicago.edu
Office: 773-834-8719

INTRODUCTION

A random hypoechoic sphere phantom (RHSP) with freehand image acquisition and automated software to measure lesion signal-to-noise ratio (LSNR) was developed for quantitative quality control (QC) for grayscale ultrasound [1]. This study aims to investigate the efficacy and efficiency of this approach for monitoring clinical image quality and longitudinal stability in breast ultrasound systems.

METHODS AND MATERIALS

Two breast ultrasound systems of the same vendor and model (Philips EPIQ Elite with eL18-4 transducers), one for diagnosis and the other for biopsy guidance, were monitored using RHSP over 20 weeks at weekly/bi-weekly intervals. Both scanners used the identical imaging protocols to scan the same RHSP containing 2mm spherical voids. In each session, a single operator performed a uniform 3-second freehand cine sweep, repeated six times. The analysis algorithm generated LSNR-versus-depth curves averaged across repeats. Radiologists subjectively assess clinical image quality from patient scans on both ultrasound systems, which were correlated with the LSNR-versus-depth curves. Longitudinal trends for peak LSNR at 1cm depth and mean LSNR over all depths (overall LSNR) were evaluated using linear regression, with variabilities across all time points calculated.

RESULTS

Data were acquired at 20 time points over 20 weeks on two scanners using the Adv. Breast and Dense Breast protocols, respectively. Each configuration was repeated six times, and LSNR values were averaged. Figure 1 shows the phantom picture, the representative RHSP images acquired on scanner 1 using Adv. Breast protocol and Dense Breast protocol.

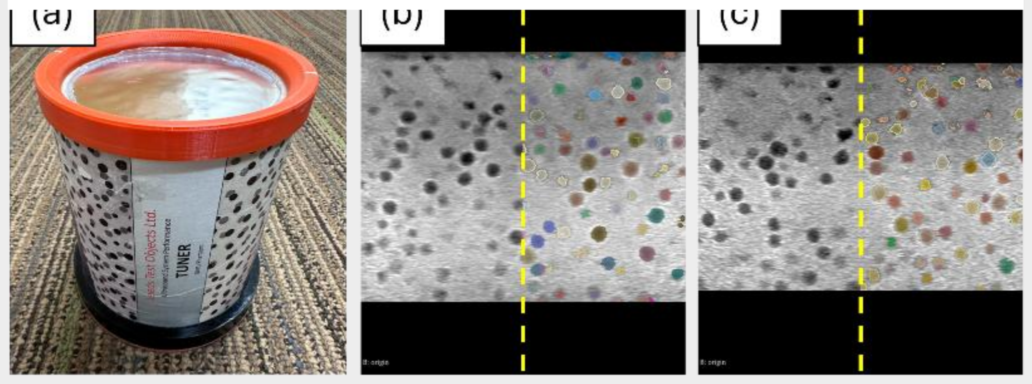

Figure 1. (a) The 2mm sphere phantom. (b-c) frames and segmentations using Adv. Breast protocol and Dense Breast protocol in Scanner 1.

RESULTS

Radiologists consistently rated the clinical image quality from scanner 1 as inferior to that of scanner 2. Similarly, using the same clinical scan protocol on the phantom, LSNR-versus-depth curves from scanner 1 were worse than those from scanner 2 across most depths.

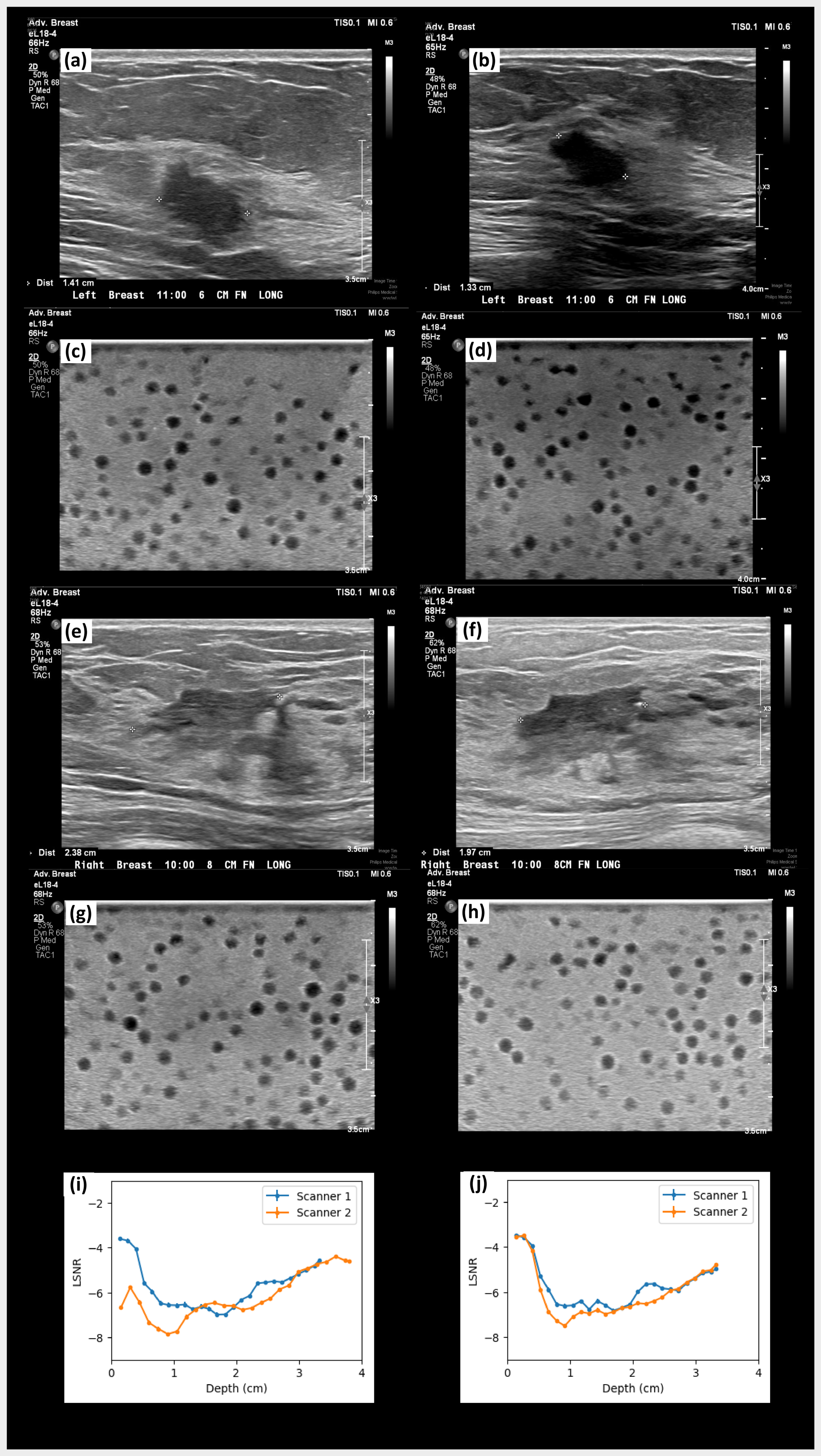

Figure 2. (a) – (b) Breast ultrasound images from patient 1 acquired on scanners 1 and 2. (c) – (d) The corresponding RHSP images using the same scan configurations as patient 1 on both scanners. (e) – (f) Breast ultrasound images from patient 2. (g) – (h) The corresponding RHSP images using the same scan configurations as patient 2. (i) – (j) The LSNR-versus-depth curves for the configurations of patients 1 and 2.

Table 1. Longitudinal variability of LSNR.

Scanner	Protocol	LSNR depth (cm)	LSNR variability	#Sphere variability	LSNR slope/wk	LSNR r^2	LSNR mean
1	Adv.Breast	1	3.2%	27.8%	0.003	0.005	-6.74
1	Dense Breast	1	1.8%	26.7%	0.008	0.197	-5.72
2	Adv.Breast	1	3.6%	21.1%	-0.012	0.064	-7.12
2	Dense Breast	1	2.1%	25.2%	-0.004	0.034	-5.78
1	Adv.Breast	Overall	3.1%	26.3%	-0.006	0.170	-5.55
1	Dense Breast	Overall	2.8%	28.5%	-0.003	0.114	-4.97
2	Adv.Breast	Overall	3.2%	22.3%	-0.004	0.043	-5.68
2	Dense Breast	Overall	2.9%	26.9%	-0.001	0.014	-5.07

RESULTS

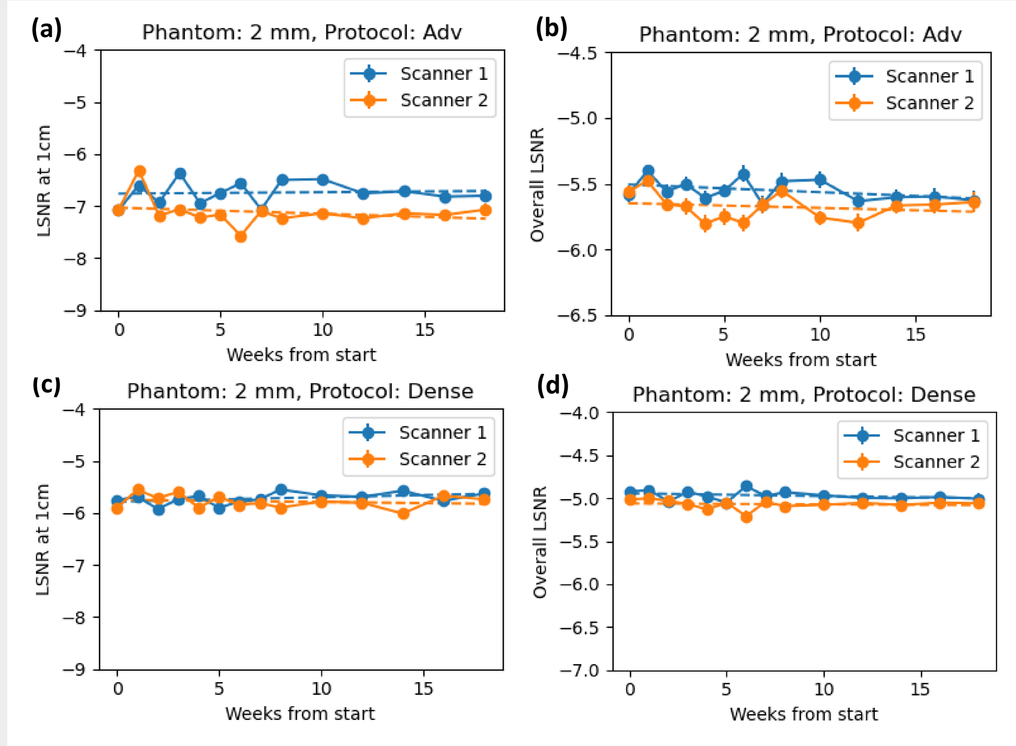

Figure 3. Longitudinal change of peak LSNR at 1cm depth and overall LSNR for scanners 1 and 2 using (a) – (b) Adv. Breast protocol; (c) – (d) Dense Breast protocol. The dashed lines are from linear regression.

Figure 3 and Table 1 summarize the longitudinal variability of the LSNR at 1cm depth and overall depth region. The 1cm depth was selected because it yielded the best LSNR across configurations. Scanner 1 showed worse peak LSNR (-6.74 versus -7.12, 20-week average) and overall LSNR (-5.55 versus -5.68, 20-week average) than scanner 2 for Adv. Breast protocol. Both scanners showed stable LSNR with minimal drift over the 20 weeks: for scanners 1 and 2 using Adv. Breast protocol respectively, their peak LSNRs had regression slopes of 0.003 (r^2=0.005) and -0.012 (r^2=0.064) per week, with variabilities of 3.2% and 3.6%; their overall LSNRs had regression slopes of -0.006 (r^2=0.170) and -0.004 (r^2=0.043) per week, with variabilities of 3.1% and 3.2%.
Both scanners demonstrated significantly lower LSNR variability than sphere-count variability. Small regression slopes and r^2 values means no temporal drift of LSNR, indicating stable scanner performance and minimal phantom aging. No significant protocol-dependent differences were observed. Scanner 1 showed slightly worse LSNR than scanner 2 in both protocols, consistent with clinical impressions that the biopsy-guidance unit has lower image quality than the diagnostic one.

CONCLUSIONS

LSNR-versus-depth curves measured from freehand-acquired RHSP images reflected differences in clinical image quality between the two scanners and correlated consistently with radiologists' assessments. The method revealed minimal drift and low variability for both scanners during a 20-week longitudinal study. Current ultrasound QC relies heavily on subjective assessment. This study demonstrates an automated quantitative QC method using RHSP that correlates well with subjective clinical image quality and shows low variability, supporting its potential use for routine ultrasound QC.

REFERENCES

1. Yu B, Wu D, Baiu C, LuZ. Assessing imaging performance of ultrasoundsystems using a random hypoechoic spherephantom with freehand scanning. Med Phys.2026;53:e70278