A Phantom-Based Assessment of Repeatability and Reproducibility of Transvaginal Quantitative Ultrasound

Shi Chen, Barbara L. McFarlin, Barbara T. Meagher, Tara A. Peters, Douglas G. Simpson, *Member, IEEE*, William D. O'Brien, Jr., *Life Fellow, IEEE*, and Aiguo Han, *Member, IEEE*

***Abstract*— This article evaluated the repeatability and reproducibility (R&R) of quantitative ultrasound (QUS) biomarkers attenuation coefficient (AC) and backscatter coefficient (BSC) in transvaginal QUS reference phantoms for obstetric applications. Five phantoms were scanned by three sonographers according to the scanning protocol. Each sonographer scanned each phantom with four transvaginal transducers of the same model (MC9-4) and three probe cover types (latex cover, nonlatex cover, and no cover). The AC and BSC were estimated by using a reference phantom method. The R&R analysis was performed for the frequency-averaged AC and logBSC ($= 10\log_{10}$BSC) (5.4–5.8 MHz) by using three-factor random effects Analysis of Variance with interaction. The total R&R variabilities for AC and logBSC are small (AC: 0.042–0.065 dB/cm-MHz; logBSC: 0.50–0.68 dB), indicating high measurement precision. These values are small compared to the ranges of AC (0.28–0.99 dB/cm-MHz) and logBSC (−33.16 to −20.35 dB) of the five phantoms. The AC and logBSC biomarkers measured on transvaginal QUS phantoms using the reference phantom method are repeatable, and reproducible between sonographers, transducers, and probe covers.**

***Index Terms*— Attenuation coefficient (AC), backscatter coefficient (BSC), quantitative ultrasound (QUS), repeatability and reproducibility (R&R).**

I. Introduction

THE Quantitative Imaging Biomarkers Alliance (QIBA) defines a quantitative imaging biomarker as an "objective characteristic derived from an *in vivo* image measured on a ratio or interval scale as an indicator of normal biological processes, pathogenic processes, or a response to a therapeutic intervention" [1]. The attenuation coefficient (AC, dB/cm-MHz) and the backscatter coefficient (BSC, 1/cm-sr) are two common biomarkers used in quantitative ultrasound (QUS). AC is an objective measure of the spatial rate of ultrasonic energy loss in tissue, and BSC is an objective measure of the fraction of ultrasonic energy returned from tissue. AC and BSC provide quantitative information of tissue microstructures [2]. In recent years, AC and BSC have shown promise in hepatic fat quantification [3]–[5] and early prediction of preterm birth [6], [7].

For AC and BSC biomarkers to be useful clinically, their precision needs to be evaluated rigorously. Precision deals with measurement variability, which is present whether the measurement conditions remain unchanged or vary between replicate measurements. It is necessary to demonstrate that the biomarkers can not only be used to repeat a measurement reliably, but also be used with a more general set of conditions. Therefore, two types of precision are considered: repeatability and reproducibility (R&R). Repeatability is "the measurement precision with conditions that *remain unchanged* between replicate measurements (repeatability conditions)" [8]. Reproducibility is "the measurement precision with conditions that *vary* between replicate measurements (reproducibility conditions)" [8]. Rigorous R&R studies allow us to evaluate separately the components contributing to variability.

In previous studies, the R&R of AC and BSC biomarkers have been determined in both phantoms [9] and human liver in adults with known or suspected nonalcoholic fatty liver disease (NAFLD) [10]–[12]. The phantom study [9] assessed the repeatability, between-transducer reproducibility, and between-sonographers reproducibility in liver-tissue mimicking reference phantoms. The first human liver study [10] evaluated the repeatability and between-transducer reproducibility of AC and BSC. The second human liver study [11] assessed inter-sonographer reproducibility of AC and BSC. The third human liver study [12] assessed the inter-platform reproducibility of AC and BSC. All data acquisitions for these previous four studies [9]–[12] were obtained from a single site (UCSD) and the frequency ranges used to analyze R&R were around the center frequency of 3.0 MHz, typical for ultrasound liver studies.

This study examined the fundamental R&R for an obstetric application of transvaginal QUS using five reference phantoms and four transvaginal transducers for QUS biomarkers (frequency range 5.4–5.8 MHz) that were being developed to assess preterm birth in pregnant human subjects. The R&R phantom data were acquired at the University of Illinois at

Manuscript received February 1, 2019; accepted June 5, 2019. Date of publication June 14, 2019; date of current version August 26, 2019. This work was supported in part by the National Institutes of Health Grant R01 HD089935. *(Corresponding author: Aiguo Han.)*

S. Chen, W. D. O'Brien, Jr., and A. Han are with the Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: chenshi613@gmail.com; wdo@uiuc.edu; han51@illinois.edu).

B. L. McFarlin, B. T. Meagher, and T. A. Peters are with the Department of Women Children and Family Health Science, UIC College of Nursing, University of Illinois at Chicago, Chicago, IL 60612 USA.

D. G. Simpson is with the Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA.

Digital Object Identifier 10.1109/TUFFC.2019.2921925

Fig. 1. One of the reference phantom surfaces showing the slot in which one of the transvaginal transducer probes is positioned in order to obtain a full-field B-mode image (as well as the raw RF data). Scale: a U.S. coin (dime) has a diameter of 1.8 cm.

Chicago (UIC) Medical Center by three sonographers using the same research protocol that is being used to acquire the IRB-approved, HIPPA compliant human subjects QUS data [6], [7]. The study focused on repeatability, between-sonographer reproducibility, between-transducer reproducibility, and between-probe cover reproducibility of AC and BSC biomarkers in phantoms.

II. Methodology

A. Study Design

The purpose of this study design was to assess baseline repeatability as well as sonographer, transducer, and transducer probe cover reproducibility of AC and BSC biomarkers using the well-established reference phantom method [13]. There were three sonographers (denoted S1, S2, and S3), four same-model ultrasonic transducers (denoted blue, green, orange, and red), and three probe covers (denoted latex cover, LC; nonlatex cover, NLC; and no cover, NC). Each sonographer scanned five phantoms (denoted UIC1, UIC2, UIC3, UIC4, and UIC5) according to the scanning protocol with each of the four transducers.

The reference phantoms' membrane surface has a recess at the midline that closely matches the contour of a Siemens MC9-4 transvaginal ultrasonic transducer. Specifically, the MC9-4 active element lens surface has a 1.1-cm radius of curvature and each reference phantom (CIRS, Inc., Norfolk, VA, USA) has a slot that is approximately 1.5-cm inside diameter and 1.1-cm deep (Fig. 1). The reference phantom slot allows for acquisition of the B-mode images (as well as the raw RF data) over much of the 176° array angle.

The three sonographers each used four Siemens MC9-4 transvaginal ultrasound transducers (4–9 MHz nominally) to scan the phantoms with a clinical ultrasonic imaging system (Siemens S2000, Issaquah, WA) with the Axius Direct Ultrasound Research Interface to acquire the RF data [14]. Specifically (see Fig. 2), one sonographer scanned one of the UIC phantoms (e.g., UIC1) three times with the same latex cover (LC, Civco, Coralville, IA), three times with the same nonlatex cover (NLC, Parker Laboratories, Fairfield, NJ), three times with no probe cover (NC), and then one of the other phantoms (e.g., UIC2) with no cover (NC); for this sequence of scans, the last phantom scanned (e.g., UIC2) served as the reference phantom. Three RF data frames were acquired during each scan. The same sonographer repeated the same scanning protocol using each of the other four phantoms separately as reference. Then, the same sonographer repeated the same scanning protocol for each of the other three transducers plus respective probe covers.

Fig. 2 represents 50 scans performed by one sonographer using the same transducer probe. Without considering the reference phantom scan (denoted by the asterisk *) for each of the five sequence sets, 45 [50–5] QUS scans were performed from which QUS biomarker outcomes were estimated. For three sonographers and four transducer probes, a total of 540 [$45 \times 3 \times 4$] QUS scans were performed. Viewed differently, with three sonographers, four transducer probes, and 45 [15 LC + 15 NLC + (20–5) NC] (five NCs deleted because they are the reference phantom scans) transducer probe cover combinations, a total of 540 [$3 \times 4 \times 45$] QUS scans were performed. Each QUS scan yielded data from which two QUS biomarkers (AC and BSC) were estimated.

B. AC and BSC Processing

AC and BSC were calculated within a field of interest (FOI) rather than the whole image region for simplicity and comparability; the FOI is the largest FOI across all images. To segment the FOI, the B-mode image of each scan was generated using the corresponding RF data. The same region shown in Fig. 3 (axial: 30 mm; lateral: 92.5°, or 245 scan lines) for each of the 540 B-mode images was used as the FOI to estimate AC and BSC. Important parameters used in processing AC and BSC included the −6-dB bandwidth (3.7–7.5 MHz), the center frequency (5.6 MHz), and the pulse length (0.4 mm).

AC was calculated by the spectral difference reference phantom technique [13] using the RF data. The QUS processing methodologies, described in detail in [9], were used to calculate the AC of one phantom.

For BSC computation, the attenuation of the phantom and the transmission loss of the probe cover were both compensated. The attenuation loss was compensated by using the estimated AC of that phantom. The probe cover's transmission loss was compensated by using the round-trip pressure transmission coefficient (Fig. 4) of that transducer cover type. Because the latex (LC) and nonlatex covers (NLC) have different acoustic properties, different transmission loss values were used for LCs and NLCs, determined as follows.

The round-trip pressure transmission coefficients (Fig. 4) were estimated using a subset of the acquired data described in Fig. 2. Specifically, the transmission coefficients of LC and NLC shown in Fig. 4 were estimated by averaging the transmission coefficients of 60 LC samples and 60 NLC samples, respectively. The sample size of 60 [$1 \times 5 \times 3 \times 4$] LC samples was a result of one LC sample in each box of

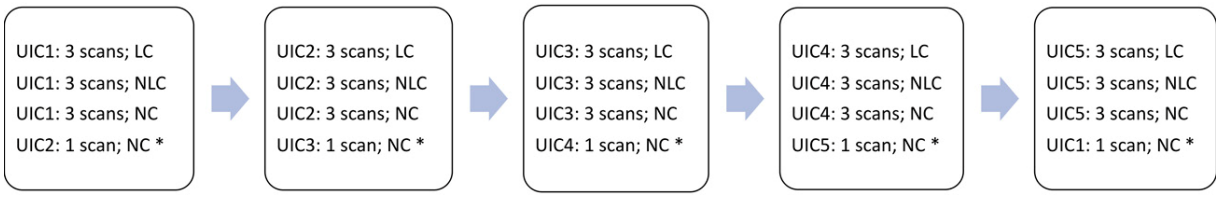

Fig. 2. Sequence of data acquisitions conducted by the same sonographer using the same transducer probe for which all three probe covers are used, and all five reference phantoms are scanned. The asterisk * denotes the reference phantom scan in each of the five sequence sets.

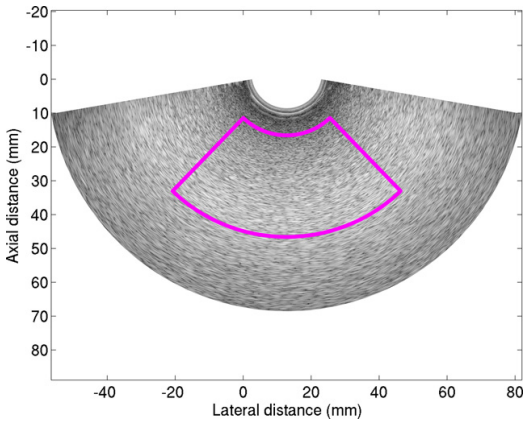

Fig. 3. FOI used for calculating AC and BSC.

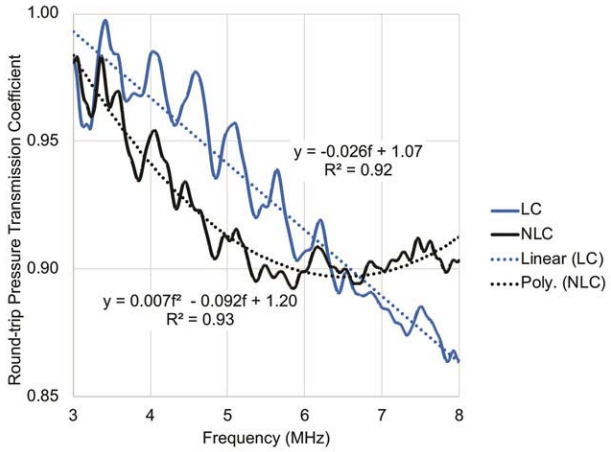

Fig. 4. Round-trip pressure transmission coefficients (%) for both the latex probe covers (LC) and nonlatex probe covers (NLC) relative to the 100% no cover (NC).

Fig. 2 times five boxes times three sonographers times four transducers. The 60 NLC samples were similarly obtained. Each LC sample was used in three scans (first row in each box of Fig. 2), and so was each NLC sample (second row in each box of Fig. 2). The transmission coefficient of each LC sample was estimated by

$$T_{\mathrm{LC}} = \frac{1}{3}\sum_{i=1}^{3}\sqrt{\frac{\mathrm{PS}_{\mathrm{LC},i}}{\mathrm{PS}_{\mathrm{NC},i}}} \tag{1}$$

where $\mathrm{PS}_{\mathrm{LC},i}$ is the power spectrum of the echo data acquired by the ith scan of the three scans shown in the first row of Fig. 2, and $\mathrm{PS}_{\mathrm{NC},i}$ is the power spectrum of the echo data acquired by the ith scan of the three scans shown in the third row of Fig. 2. Similarly, the transmission coefficient of each NLC sample was estimated by

$$T_{\mathrm{NLC}} = \frac{1}{3}\sum_{i=1}^{3}\sqrt{\frac{\mathrm{PS}_{\mathrm{NLC},i}}{\mathrm{PS}_{\mathrm{NC},i}}} \tag{2}$$

where $\mathrm{PS}_{\mathrm{NLC},i}$ is the power spectrum of the echo data acquired by the ith scan of the three scans shown in the second row of Fig. 2, and $\mathrm{PS}_{\mathrm{NC},i}$ is the power spectrum of the echo data acquired by the ith scan of the three scans shown in the third row of Fig. 2. A linear equation was fitted to the average transmission coefficient estimated from 60 LC samples, and a quadratic equation was fitted to the average transmission coefficient estimated from 60 NLC samples (Fig. 2). These equations were found to provide adequate fit to the data over the frequency range of the experiment. The linear and quadratic fitted values were used to correct for the transmission loss of LC and NLC samples, respectively, during the BSC estimation process.

After compensating the attenuation loss and the transmission loss, the BSC was calculated by the reference phantom technique [13] with the QUS processing methodologies described in detail in ref. [9].

C. R&R Methodology

The purpose of this study was to estimate the R&R of AC and BSC biomarkers using the reference phantom methodology. Repeatability means the closeness of measurements obtained by the same method under the same conditions, that is, the closeness of results by the same sonographer, same transducer probe, and the same probe cover. The reproducibility means the closeness of measurements obtained by the same method under different conditions, that is, the closeness of results by different sonographers, the different transducer probes, and different probe covers.

A balanced analysis of variance (ANOVA) approach was used to analyze various components of the R&R. This study includes three effects: the sonographer effect, the transducer effect, and the probe cover effect. The interactions between these three effects were unknown. Therefore, a three-factor model with interactions was used. Random effects were assumed because the sonographers, transducers, and probe covers were considered random samples of a larger pool of sonographers, transducers, and probe covers. The three-factor random model was used to calculate repeatability, reproducibility, and interactions between factors [15].

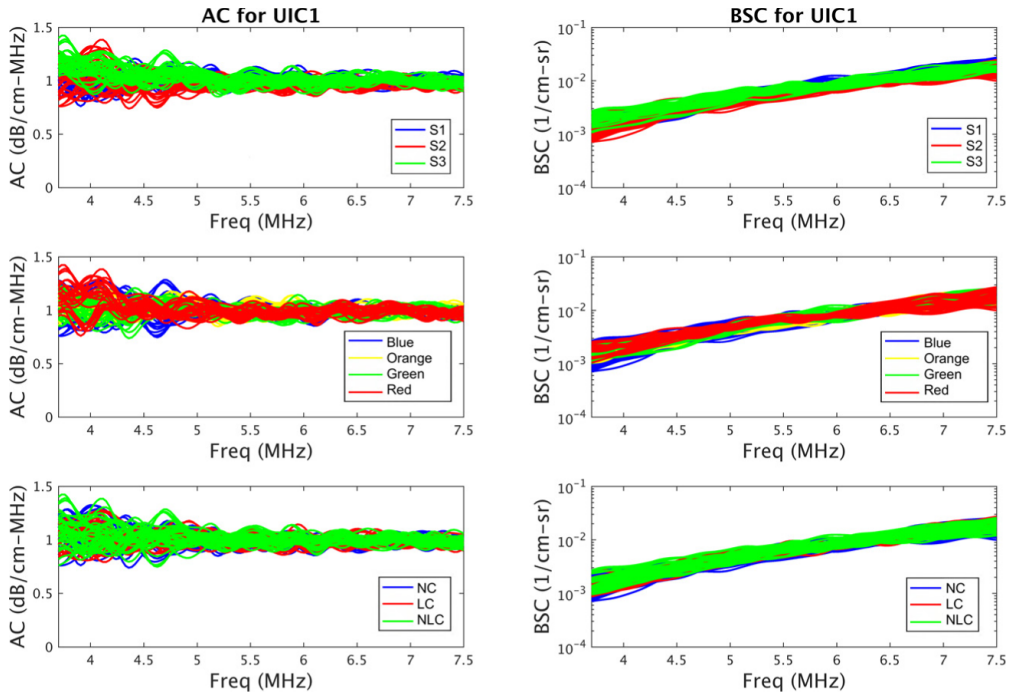

Fig. 5. Frequency-dependent (3.7–7.5 MHz) AC and BSC curves from 108 measurements of phantom UIC1. AC and BSC curves color coded by sonographers (top row). AC and BSC curves color coded by transducers (middle row). AC and BSC curves color coded by probe covers denoted by the probe cover (LC or NLC) or no probe cover (NC) (bottom row).

The R&R analysis was performed separately for each of the five phantoms. Also, AC and logBSC (defined as $10\log_{10}$BSC) were separately assessed; the log transformation was performed to normalize the distribution. The p-values of Shapiro–Wilk test for logBSC of five phantoms are 0.027, 0.075, 0.078, 0.099, and 0.513. After Bonferroni–Holm adjustment for multiple testing, none of these five tests is significant at level $\alpha = 0.10$. Frequency-averaged AC and logBSC values (5.4–5.8 MHz), rather than AC and logBSC versus frequency spectra in the entire bandwidth (3.7–7.5 MHz), were used for the R&R analysis, because R&R analysis over the entire bandwidth would have required functional ANOVA [16] as logBSC is correlated with the frequency. Therefore, the bandwidth of frequencies for analysis was selected around the RF data center frequency in order to utilize the less complex ANOVA approach.

The three-factor random effects with interaction [15] was modeled as

$$Y_{ijkl} = \mu_Y + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + (ABC)_{ijk} + \varepsilon_{ijk}, \quad (i = 1, \ldots, n_A, j = 1, \ldots, n_B, k = 1, \ldots, n_C, l = 1, \ldots, n_{\text{acquisition}}) \quad (3)$$

where Y_{ijkl} is the measured AC or logBSC of a phantom, μ_Y is a constant, and $A_i, B_j, C_k, (AB)_{ij}, (AC)_{ik}, (BC)_{jk}, (ABC)_{ijk}$, and ε_{ijk} are jointly independent normal random variables with means of zero and variances $\sigma_A^2, \sigma_B^2, \sigma_C^2, \sigma_{AB}^2, \sigma_{AC}^2, \sigma_{BC}^2, \sigma_{ABC}^2$, and σ_E^2, respectively. The terms $A_i, B_j, C_k, (AB)_{ij}, (AC)_{ik}, (BC)_{jk}, (ABC)_{ijk}$, and ε_{ijk} represent the effects of factor A, factor B, factor C (e.g., A = sonographer, B = transducer, and C = probe cover), the interaction between factors A and B, the interaction between factors A and C, the interaction between factors B and C, the interaction between the three factors, and the error term (repeatability effect). The terms n_A, n_B, and n_C represent the number of conditions for factors A, B, and C, respectively (e.g., $n_A = 3$ sonographers, $n_B = 4$ transducers, and $n_C = 3$ probe covers), and $n_{\text{acquisition}}$ represents the number of measurements made with each combination of factors A, B, and C.

III. Results

The estimated AC and BSC curves show good agreements among 108 measurements for each phantom. The frequency-dependent (3.7–7.5 MHz) AC and BSC curves of phantom UIC1 are shown in Fig. 5. The AC variances are slightly larger than the BSC variances because the AC estimation requires a calculation that the BSC estimation does not require, that is, the AC estimation uses a slope of a straight line that fits the natural log ratio of the sample power spectrum to the reference power spectrum at different depths [13]. Due to different scales of measurement, the variances of the AC and BSC measurements are on different scales; thus, the factor variances were compared with response variances separately for the two types of measurements.

The boxplots of AC and logBSC for each phantom over the 5.4–5.8 MHz bandwidth are shown in Figs. 6 and 7, respectively. The red horizontal lines in Figs. 6 and 7 are the calibrated values of AC and logBSC, respectively, for each phantom. The estimated AC and logBSC values (in Figs. 6 and 7, respectively) are close to the calibrated AC and logBSC for each phantom (see Fig. 10). Overall, the AC and logBSC values within the small bandwidth are consistent and agree with calibrated values.

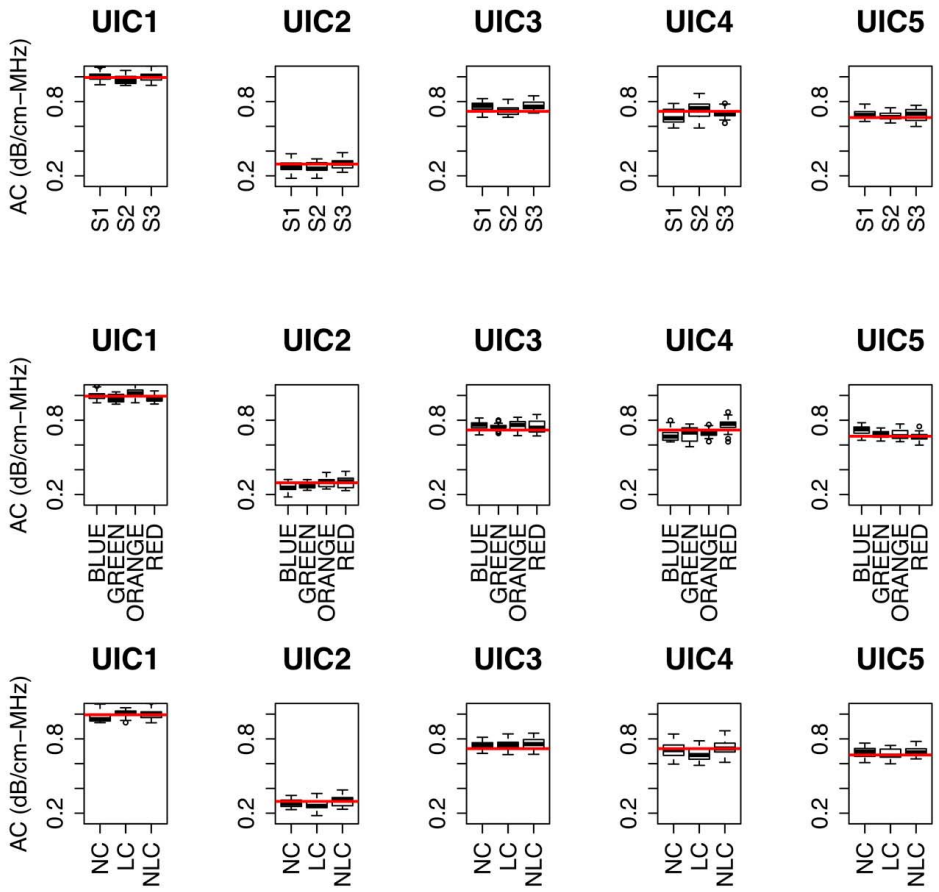

Fig. 6. Boxplots of AC of each phantom. The plots in the top row are drawn with respect to sonographers, the plots in the middle row are drawn with respect to transducers, and plots in the bottom row are drawn with respect to transducer probe covers. The red horizontal lines are the calibrated values of AC (see Fig. 10) for each phantom. The size of the box indicates the upper quartile and lower quartile. The inner lines are the median of estimated AC, the top whiskers are the maximum value except the outliers, and the low whiskers are the minimum value except the outliers, and open circles (if any) represent outliers.

The R&R results of AC and logBSC of all phantoms are shown in Figs. 8 and 9, respectively, and are expressed in terms of the standard deviation. For R&R estimates of AC, the repeatability shows no correlation with phantoms and is consistent for each phantom. Except for the reproducibility of phantom UIC4, the reproducibility shows no correlation with phantoms and is consistent for the other four phantoms. In general, the repeatability (0.020–0.023 dB/cm-MHz) is smaller than the reproducibility (0.035–0.061 dB/cm-MHz) for each phantom. There are interaction effects between different factors for each phantom. The AC uncertainty caused by the transducer probe effect (0.028–0.053 dB/cm-MHz) is larger than the uncertainty caused by the sonographer effect by 1%–23% or the probe cover effect by 2%–111%.

For R&R estimates of logBSC, the repeatability and the reproducibility also show no correlation with phantoms and are consistent for each of phantom. The repeatability (0.23–0.30 dB) is better than the reproducibility (0.42–0.61 dB) for each phantom. The logBSC uncertainty caused by the transducer probe effect (0.39–0.60 dB) is larger than the uncertainty caused by the sonographer effect by 2%–31% or the probe cover effect by 16%–72%.

The R&R results show that the total R&Rs are consistent among all phantoms. For each phantom, the total R&R of AC and logBSC is much smaller compared to the mean value for that phantom. The total R&R of AC and logBSC is small (AC: 0.042–0.065 dB/cm-MHz; logBSC: 0.50–0.68 dB) compared to the ranges of AC (0.28–0.99 dB/cm-MHz) and logBSC (−33.16−−20.35 dB). The ratio of the highest total R&R to the AC range is 9.8% and the ratio of the highest total R&R to the logBSC range is 7.6%.

IV. Discussion

This study examined the R&R of AC and BSC biomarkers in transvaginal QUS phantoms for obstetric applications. The total R&R variabilities are low for both AC and logBSC for all five phantoms and indicate high precisions of AC and log-BSC results using the reference phantom technique. The total R&R variabilities of AC and logBSC are 0.042–0.065 dB/cm-MHz and 0.50–0.68 dB, respectively, and are small compared to the AC range (0.28–0.99 dB/cm-MHz) and the logBSC range (−33.16−−20.36 dB). The repeatability of AC (0.020–0.023 dB/cm-MHz) and logBSC (0.23–0.30 dB) is smaller than the reproducibility of AC (0.035–0.061 dB/cm-MHz) and logBSC (0.42–0.61 dB), respectively. The variabilities of AC (0.001–0.003 dB/cm-MHz) and logBSC (0.35–0.52 dB) measured in the calibration are smaller than the total R&R variability for AC and logBSC.

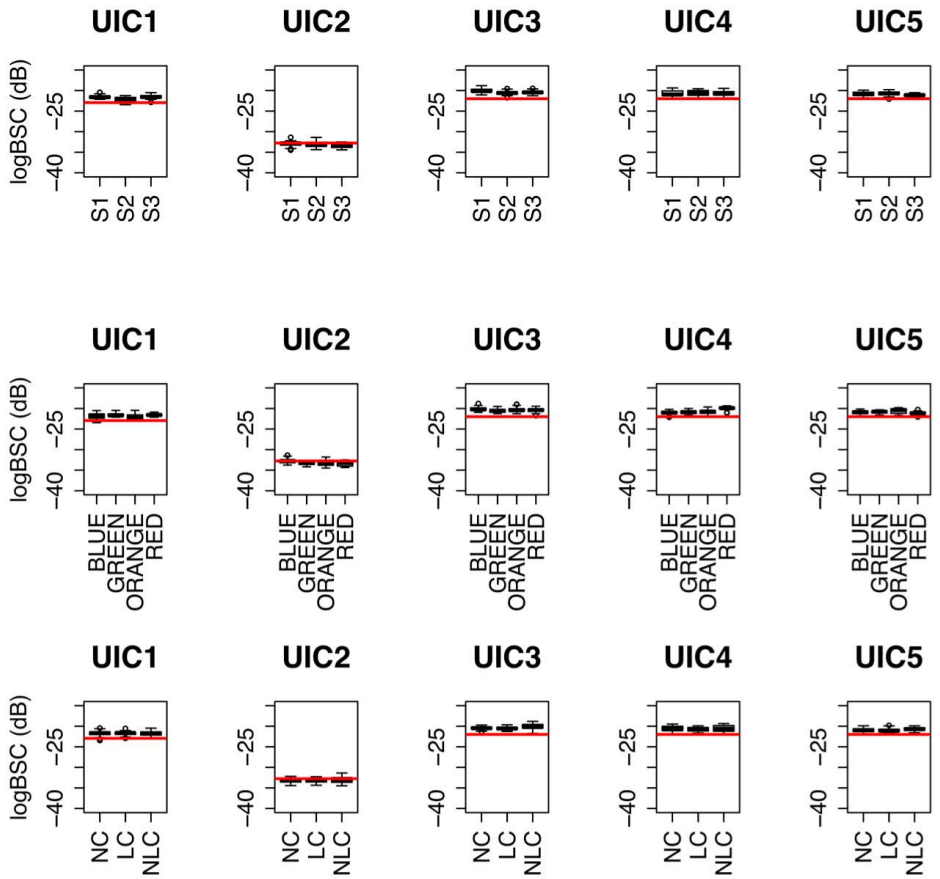

Fig. 7. Boxplots of logBSC of each phantom. The plots in the top row are drawn with respect to sonographers, the plots in the middle row are drawn with respect to transducers, and plots in the bottom row are drawn with respect to transducer probe covers. The red horizontal lines are the calibrated values of logBSC (see Fig. 10) for each phantom. The size of the box indicates the upper quartile and lower quartile. The inner lines are the median of estimated logBSC, the top whiskers are the maximum value except the outliers, and the low whiskers are the minimum value except the outliers, and open circles (if any) represent outliers.

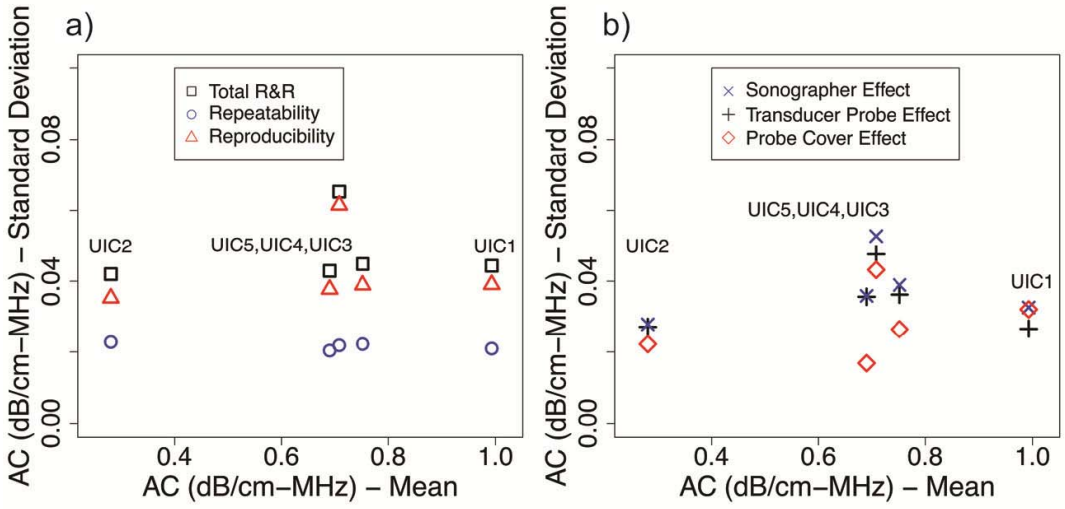

Fig. 8. R&R summary of AC of all phantoms. (a) Total R&R, repeatability, and reproducibility. (b) Sonographer effect, transducer probe effect, and probe cover effect.

Compared with the results of a previous liver-mimicking phantom study [9], the repeatability of AC and logBSC in this study is close to the repeatability of AC and logBSC in the previous study [9]. The reproducibility of AC and logBSC is larger than the reproducibility of AC and logBSC in the previous liver-mimicking study due to several reasons. The first reason is that this study has an additional probe cover effect, which the previous study did not have [9]. Compared to

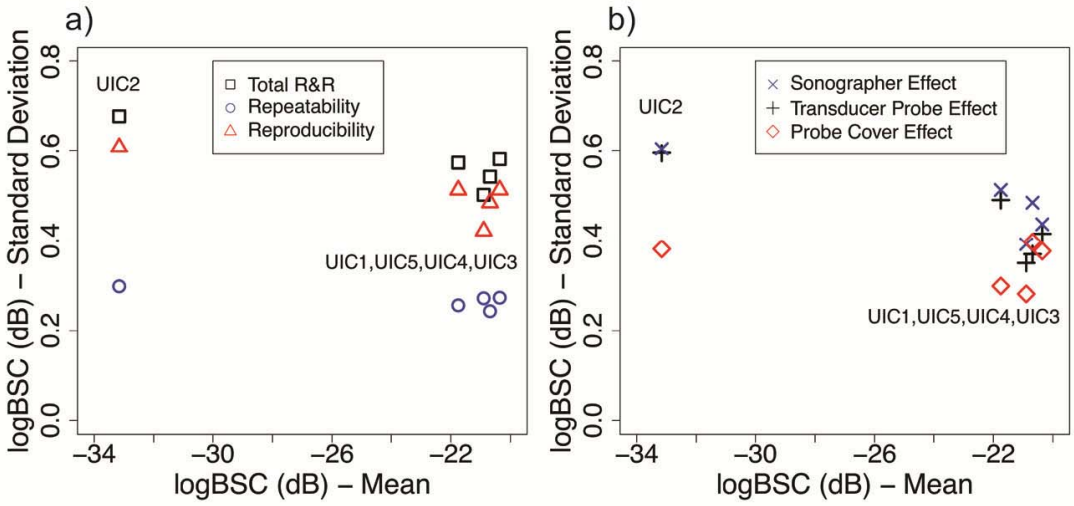

Fig. 9. R&R summary of logBSC of all phantoms. (a) Total R&R, repeatability, and reproducibility. (b) Sonographer effect, transducer probe effect, and probe cover effect.

the two factors (sonographers and transducer) in the previous liver-mimicking study, there were three factors (sonographers, transducers, and probe covers) in this study that resulted in increased reproducibility variability. The second reason is that the sample variations of probe covers introduced additional variability in the measurements. In this study, the same set of transmission coefficients was used to compensate the loss of all probe covers for each type. However, the same type probe covers might physically differ from each other and have slightly different transmission losses. Therefore, using the same set of transmission coefficients for all probe cover types led to larger reproducibility variability for both AC and logBSC.

There are some other factors which contribute to the differences in reproducibility between the two phantom-based studies. For example, the type of transducer used in the study can influence reproducibility differently. The transducers used in this study were transvaginal probes with a small foot print (31 mm) and a wide field of view (176°), while the transducers used in [9] were larger curvilinear arrays (e.g., foot print of 4C1 = 61 mm) with a narrower field of view (66°). The frequency can also be a factor influencing the reproducibility. The center frequency used in this study (5.6 MHz) was greater than that in the previous study (2.8 MHz) [9]. However, the study was not designed to determine how the transducer type and the frequency range influenced reproducibility of AC and logBSC.

This study shows good R&R of AC and logBSC in phantoms and suggests good QUS precision in obstetric applications. Further studies should determine the precision of QUS parameters *in vivo*. From the previous liver studies [10]–[12], the R&R *in vivo* are lower than the R&R in phantoms due to additional variability caused by biological tissues. The required precision ranges of AC and logBSC will depend on their clinical use in obstetric ultrasound.

There were some limitations in this phantom-based study. We only considered effects of sonographers, probes, and probe covers. There were other factors that could have contributed to the reproducibility that were not considered here such as the transducer type, the operating frequency, the imaging platform, and the FOI size. Future studies may evaluate the effects of these factors on reproducibility of AC and BSC. Also, we used the spectrum difference method for AC calculations and reference phantom method for BSC calculations. Other AC and BSC computing algorithms exist but were not evaluated [17], [18].

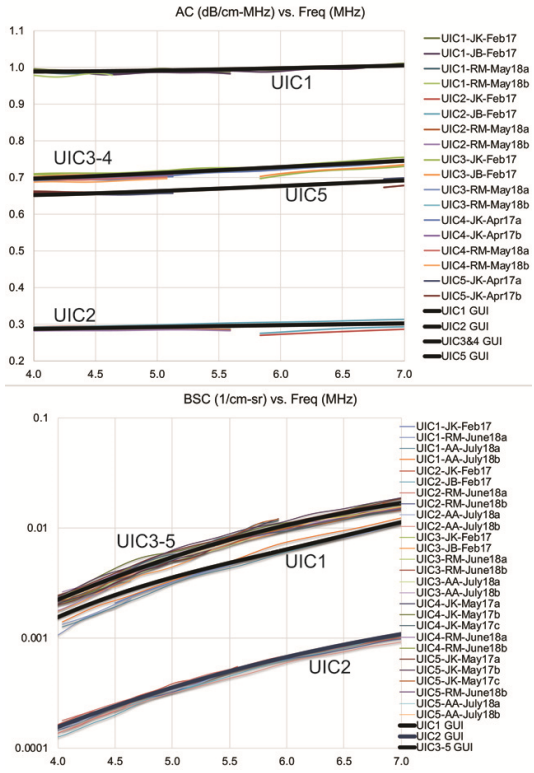

Fig. 10. Calibrated AC (top) and BSC (bottom) versus frequency for the five UIC reference phantoms along with each calibration curve (denoted by GUI). The calibration data were acquired between February 2017 and July 2018 and bracketed the dates of the R&R phantom study that was conducted with the five phantoms and three sonographers during April 2018 and May 2018.

V. Conclusion

The AC and logBSC biomarkers measured with transvaginal QUS phantoms using the reference phantom method are repeatable, and reproducible among sonographers, transducer probes, and probe covers.

Appendix
Phantom Calibration Results

The five Zerdine Hydrogel phantoms (CIRS, Inc., Norfolk, VA, USA) were calibrated multiple times before and after the R&R phantom data acquisition by the three sonographers. The detailed QUS calibration procedures are discussed in [9]. The AC and BSC calibration results are shown in Fig. 10. The sound speeds we measured for all phantoms were 1538 $\pm$ 4 m/s at 20 $\pm$ 1 °C. Multiple additional calibration results were averaged to yield the phantom calibrations. The precision of each of the phantom's calibrations around the 5.6-MHz center frequency is represented as the ratio of the frequency-averaged (5.4–5.8 MHz) standard deviation to the frequency-averaged mean, yielding 0.087%–1.41% for AC and 1.36%–2.49% for BSC. In addition, coded into the QUS processing methodologies were the respective round-trip pressure transmission coefficients (Fig. 4) for both the latex (LC) and nonlatex (NLC) probe covers.

Acknowledgment

The authors would like to thank R. J. Miller, A. Ahmad, J. Kelly, and J. Berndt for the dedicated contributions for the phantom calibration expertise.

References

[1] L. G. Kessler *et al.*, "The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions," *Stat. Methods Med. Res.*, vol. 24, no. 1, pp. 9–26, Feb. 2015.

[2] L. A. Wirtzfeld *et al.*, "Techniques and evaluation from a cross-platform imaging comparison of quantitative ultrasound parameters in an in vivo rodent fibroadenoma model," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 60, no. 7, pp. 1386–1400, Jul. 2013.

[3] M. P. Andre *et al.*, "Accurate diagnosis of nonalcoholic fatty liver disease in human participants via quantitative ultrasound," in *Proc. IEEE Int. Ultrason. Symp.*, vol. 2014, pp. 2375–2377.

[4] S. C. Lin *et al.*, "Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat using a new quantitative ultrasound technique," *Clin. Gastroenterology Hepatology*, vol. 13, no. 7, pp. 1337–1345, Jul. 2015.

[5] J. S. Paige *et al.*, "A pilot comparative study of quantitative ultrasound, conventional ultrasound, and MRI for predicting histology-determined steatosis grade in adult nonalcoholic fatty liver disease," *Amer. J. Roentgenology*, vol. 208, no. 5, pp. W168–W177, May 2017.

[6] B. L. McFarlin *et al.*, "Beyond cervical length: A pilot study of ultrasonic attenuation for early detection of preterm birth risk," *Ultrasound Med. Biol.*, vol. 41, no. 11, pp. 3023–3029, Nov. 2015.

[7] B. L. McFarlin *et al.*, "Development of an ultrasonic method to detect cervical remodeling *in vivo* in full-term pregnant women," *Ultrasound Med. Biol.*, vol. 41, no. 9, pp. 2533–2539, Sep. 2015.

[8] D. C. Sullivan *et al.*, "Metrology standards for quantitative imaging biomarkers," *Radiology*, vol. 277, no. 3, pp. 813–825, 2015.

[9] A. Han, M. P. Andre, J. W. Erdman, R. Loomba, C. B. Sirlin, and W. D. O'Brien, Jr., "Repeatability and reproducibility of a clinically based QUS phantom study and methodologies," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 1, pp. 218–231, Jan. 2017.

[10] A. Han *et al.*, "Repeatability and reproducibility of the ultrasonic attenuation coefficient and backscatter coefficient measured in the right lobe of the liver in adults with known or suspected nonalcoholic fatty liver disease," *J. Ultrasound Med.*, vol. 37, no. 8, pp. 1913–1927, Aug. 2018.

[11] A. Han *et al.*, "Inter-sonographer reproducibility of quantitative ultrasound outcomes and shear wave speed measured in the right lobe of the liver in adults with known or suspected non-alcoholic fatty liver disease," *Eur. Radiol.*, vol. 28, no. 12, pp. 4992–5000, Dec. 2018.

[12] A. Han *et al.*, "Inter-platform reproducibility of ultrasonic attenuation and backscatter coefficients in assessing NAFLD," *Eur. Radiol.*, vol. 29, no. 9, pp. 4699–4708, Sep. 2019.

[13] L. X. Yao, J. A. Zagzebski, and E. L. Madsen, "Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors," *Ultrason. Imag.*, vol. 12, no. 1, pp. 58–70, 1990.

[14] S. S. Brunke, M. F. Insana, J. J. Dahl, C. Hansen, M. Ashfaq, and H. Ermert, "An ultrasound research interface for a clinical system," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 53, no. 10, pp. 1759–1771, Oct. 2007.

[15] R. K. Burdick, C. M. Borror, and D. C. Montgomery, *Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models*, (Statistics and Applied Probability). Philadelphia, PA, USA: SIAM, 2005.

[16] Y. Park and D. G. Simpson, "Robust probabilistic classification applicable to irregularly sampled functional data," *Comput. Stat. Data Anal.*, vol. 131, pp. 37–49, Mar. 2019.

[17] A. Haak *et al.*, "Algorithm for estimating the attenuation slope from backscattered ultrasonic signals," in *Proc. IEEE Int. Ultrason. Symp.*, Sep. 2009, pp. 1946–1949.

[18] H. Kim and T. Varghese, "Hybrid spectral domain method for attenuation slope estimation," *Ultrasound Med. Biol.*, vol. 34, no. 11, pp. 1808–1819, Nov. 2008.

Shi Chen was born in Shaanxi, China, in 1997. He received the B.S. degree in electrical engineering from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 2019. He is currently pursuing the M.S. degree in electrical engineering with Stanford University, Stanford, CA, USA.

His research interests are biomedical ultrasound imaging, quantitative ultrasound imaging, and signal processing.

Mr. Chen was awarded as one of the 2018–2019 University Honors Bronze Tablet Scholars in 2019.

Barbara L. McFarlin received the B.S.N., M.S., and Ph.D. degrees from University of Illinois at Chicago (UIC), Chicago, IL, USA.

She conducted her dissertation research at the Bioacoustics Research Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL, USA, and has collaborated with Prof. O'Brien since 2001. She is a Certified Nurse Midwife, who has delivered over 4000 babies, and a Registered Obstetric and Gynecologic Sonographer. She is currently a Professor with the Department of Women Children and Family Health Science, UIC. Her research seeks to understand the mechanisms of cervical remodeling leading to preterm birth; and develop new technologies to noninvasively predict women at risk for spontaneous preterm birth. She is currently collecting data on an NIH R01 funded study to predict spontaneous preterm birth with quantitative ultrasound in human pregnancy.

Dr. McFarlin is a fellow of the American College of Nurse Midwives and the American Academy of Nurses.

Barbara T. Meagher received the bachelor's and M.S. degrees in nursing and the master's degree in the art of pastoral studies from Loyola University of Chicago, Chicago, IL, USA, and the Doctorate degree in nursing practice from the Chamberlain College of Nursing, Downers Grove, IL, USA, in 2015.

She completed her midwifery certification at the University of Illinois at Chicago, Chicago, IL, USA, in 1996, where she has been a registered diagnostic medical sonographer since 1999, and currently an Assistant Professor with the College of Nursing, University of Illinois at Chicago. She has lectured nationally and internationally on a variety of topics related to high risk pregnancy. Her research interests include postpartum hemorrhage, high risk pregnancy, bacterial vaginosis, and preterm labor.

Tara A. Peters received the B.S. degree from The University of Texas at Austin, Austin, TX, USA, in 1997, and the degree in diagnostic medical sonography in Oak Brook, IL, USA, in 2011.

She is currently a Visiting Research Specialist with the Department of Women Children and Family Health Science, University of Illinois at Chicago, Chicago, IL, USA. She has worked with multiple research studies since 2012, including research in preterm birth, strength recovery following stem cell transplantation, and cardiac effects of binge drinking by young adults. In 2017, she received her certificate from SOCRA as a Certified Clinical Research Professional (CCRP).

Douglas G. Simpson (M'12) received the B.A. degree from Carleton College, Northfield, MN, USA, and the M.S. and Ph.D. degrees from the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

From 1982 to 1985, he was a Mathematical Statistician with the Biometry and Risk Assessment Program, National Institute of Environmental Health Sciences. Since1985, he has been with the University of Illinois at Urbana–Champaign, Champaign, IL, USA, where he is currently a Professor with the Department of Statistics and an Affiliated Professor with the Beckman Institute for Advanced Science and Technology. His research interests include applied and computational statistics, quantitative image analysis, machine learning and functional data, and the general theory of robust and semiparametric statistical methods.

Dr. Simpson is a fellow of the American Statistical Association, the Institute of Mathematical Statistics, and the American Association for the Advancement of Science. He served as an Associate Editor for the *Journal of the American Statistical Association* from 1996 to 1999, *Biometrics* from 2000 to 2006, and *Chemometrics and Intelligent Laboratory Systems* from 1999 to 2006. He served as a regular member of the Biostatistical Research and Design (BMRD) Study Section of the National Institutes of Health from 2006 to 2010, and as Chair-Elect, Chair, and Past-Chair of the American Statistical Association Caucus of Academic Representatives from 2007 to 2010.

William D. O'Brien, Jr., (S'64–M'70–SM'79–F'89–LF'08) received the B.S., M.S., and Ph.D. degrees from the University of Illinois at Urbana–Champaign, Urbana, IL, USA.

From 1971 to 1975, he was with the Center for Devices and Radiological Health, U.S. Food and Drug Administration. In 1975, he joined the University of Illinois, as a Faculty Member. He was the Donald Biggar Willet Professor of Engineering, where he is currently a Research Professor of electrical and computer engineering and the Director of the Bioacoustics Research Laboratory. He has authored 423 papers. His research interests involve the many areas of ultrasound-tissue interaction, including biological effects and quantitative ultrasound imaging.

Dr. O'Brien is a fellow of the Acoustical Society of America and the American Institute of Ultrasound in Medicine, and is a Founding Fellow of the American Institute of Medical and Biological Engineering. He was a recipient of the IEEE Centennial Medal in 1984, the AIUM Presidential Recognition Awards in 1985 and 1992, the AIUM/WFUMB Pioneer Award in 1988, the IEEE Outstanding Student Branch Counselor Award for Region 4 in 1989, the AIUM Joseph H. Holmes Basic Science Pioneer Award in 1993, the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society Distinguished Lecturer from 1997 to 1998, the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society's Achievement Award in 1998, the IEEE Millennium Medal in 2000, the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society's Distinguished Service Award in 2003, the AIUM William J. Fry Memorial Lecture Award in 2007, and the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society's Rayleigh Award in 2008. He served as President of the IEEE Sonics and Ultrasonics Group (currently the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society) from 1982 to 1983, an Editor-in-Chief of the IEEE TRANSACTIONS ON ULTRASONICS, FERROELECTRICS, AND FREQUENCY CONTROL from 1984 to 2001, and a President of the American Institute of Ultrasound in Medicine from 1988 to 1991.

Aiguo Han (S'13–M'15) received the B.S. degree in acoustics from Nanjing University, Nanjing, China, in 2008, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 2011 and 2014, respectively.

He is currently a Research Assistant Professor with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign. His current research interests include ultrasonic wave propagation in biological media, biomedical ultrasound imaging, quantitative ultrasound, and signal processing and machine learning for ultrasonic tissue characterization.