ULTRASOUND

Non-enhanced ultrasound is not a satisfactory modality for measuring necrotic ablated volume after radiofrequency ablation of benign thyroid nodules: a comparison with contrast-enhanced ultrasound

Lin Yan [1,2] · Yukun Luo [1] · Jing Xiao [1] · Lin Lin [1]

Abstract

Objective To investigate the intra- and inter-observer reliability and agreement between gray-scale and Doppler ultrasound (US) and contrast-enhanced ultrasound (CEUS) in measuring ablated volume (Va) after radiofrequency ablation (RFA) for benign thyroid nodules.

Methods A total of 173 patients with 190 benign thyroid nodules who underwent RFA were included in this study. After RFA, the total volume of ablated nodule was divided into Va and the incompletely treated vital volume. Patients were followed up at 1, 3, 6, 12 months, and every 12 months thereafter. Two independent observers measured Va using US and CEUS during the same follow-up visit. The intra- and inter-observer reliability of the two measurement modalities was assessed using intraclass correlation coefficient (ICC) with 95% confidence interval. The Bland-Altman analysis was used to evaluate agreement, which was expressed as a mean difference with 95% limits of agreement (LOA).

Results The mean follow-up time was 23.17 ± 12.70 months. Va measured by US was significantly larger than by CEUS ($p < 0.001$). The intra- and inter-observer reliability decreased over the follow-up period and became moderate in both subgroups at 12 months (all ICC < 0.75). The mean difference and LOA became larger and wider during the follow-up. The best agreement was found in nodules < 10 ml at 1 month with a mean difference of 1.166 and LOA between 0.413 and 3.294.

Conclusions The intra- and inter-observer reliability and agreement of US and CEUS in measuring Va were unsatisfactory. CEUS should be considered when Va was needed for further evaluation or in the case of nodules with suspected regrowth.

Key Points

• *Va measured by gray-scale and Doppler US was significantly larger than that by CEUS.*
• *Va measured by gray-scale and Doppler US lacked intra- and inter-observer reliability and agreement with CEUS.*
• *CEUS should be preceded to gray-scale and Doppler US for the measurement of Va.*

Keywords Radiofrequency ablation · Thyroid nodule · Ultrasonography · Reproducibility of results

✉ Yukun Luo
lyk301@163.com

[1] Department of Ultrasound, First Medical Center, Chinese PLA General Hospital, No. 28 Fuxing Road, Haidian District, Beijing, China

[2] Health Management Center, The Second Hospital of Dalian Medical University, No. 467 Zhongshan Road, Shahekou District, Dalian, China

Abbreviations

CEUS	Contrast-enhanced ultrasound
CIs	Confidence intervals
LOA	95% limits of agreement
ICC	Intraclass correlation coefficient
RFA	Radiofrequency ablation
US	Gray-scale and Doppler ultrasound
Vt	Total volume
Va	Ablated volume
Vv	Vital volume
VRR	Volume reduction rate

Introduction

Radiofrequency ablation (RFA) is a safe and effective treatment modality to patients with benign thyroid nodules by guidelines [1, 2]. Many studies have shown that the volume reduction after ablation is significant with clinical improvement of local symptoms or cosmetics [3–9]. Although the anatomy of the thyroid gland and neck is complex, the complication rate after RFA for benign nodules is only 2.11% and no life-threatening complications occur [10].

There are several parameters for post-ablation evaluation, like volume reduction rate (VRR), therapeutic success rate, and cosmetic/symptom scores, most of which mainly based on gray-scale and color Doppler ultrasound (US). With recent extensive studies, some novel parameters have emerged. When dividing the total volume (Vt) of ablated nodule after RFA into the necrotic ablated volume (Va) and the incompletely treated vital volume (Vv) [11, 12], Sim et al [11] found that increased Vv could be an early sign of nodule regrowth after RFA. A quantitative index, the initial ablation ratio (IAR) calculated by the ratio of Va to Vt at the first follow-up period, predicted therapeutic success after RFA [13]. All these parameters needed Va for calculation, identified on US as a decreased hypoechoic zone without vascularity in the treated nodule [11, 13]. However, the boundary between the ablated and vital zone is not easily differentiated on US, making Va measurement potentially inaccurate.

Contrast-enhanced ultrasound (CEUS) is a contrast harmonic imaging technique that allows the detection and characterization of focal lesions by assessing the microvascularization with a second-generation contrast agent [14–17]. Compared with US, CEUS is a superior method for the detection of microvasculature circulation dynamics and is useful for precise definition of the size and margins of the necrotic zone induced by ablation [18]. CEUS has been used to evaluate the complete ablation immediately after the procedure. It has not been used to measure Va after thyroid ablation. To our knowledge, no study has compared the measurement of Va using US with CEUS during the follow-up of RFA for benign thyroid nodules.

Therefore, the purpose of this study was to investigate the intra- and inter-observer concordance and agreement between US and CEUS in measuring Va after RFA for benign thyroid nodules.

Materials and methods

This retrospective study was approved by the Institutional Review Board of Chinese PLA Hospital. Written informed consent was obtained from all the patients prior to RFA and CEUS.

Patients

All the enrolled patients fulfilled these inclusion criteria: (1) confirmation of benign nodule status on two separate fine-needle aspiration (FNA) or core-needle biopsy (CNB); (2) no suspicious malignant features on US examination; (3) patients with solid or predominantly solid nodules; (4) indication to treat, due to cosmetics, clinical symptoms, or rapid growth; (5) serum thyroid hormone and thyrotropin levels within normal ranges; (6) refusal or ineligibility for surgery; (7) follow-up time of $\geq$ 12 months; (8) underwent CEUS during the follow-up. Exclusion criteria were as follows: (1) malignancy findings or follicular neoplasm on FNA or CNB; (2) nodules with benign result on FNA or CNB had suspicious of malignancy in US; (3) follow-up time of < 12 months; (4) refused CEUS during the follow-up.

From August 2014 to December 2018, 517 patients with benign thyroid nodules underwent RFA in this institution. Among them, patients who refused CEUS ($N = 295$) or follow-up time of < 12 months ($N = 103$) were excluded. At last, 173 patients with 190 benign thyroid nodules were enrolled in this study. The flowchart of patient enrolment is shown in Fig. 1.

Pre-ablation assessment

US and CEUS before and after RFA, as well as during follow-up, were performed using an Acuson Sequoia 512 (Siemens Healthineers) with a 15L8W linear array transducer or a iU22 (Philips Medical Systems) with a L12-5 linear array transducer or an M9 (Mindray) with a L12-4 linear array transducer.

US-guided RFA was always performed using the Acuson Sequoia 512 with a 6L3 linear array transducer.

CEUS evaluated the ablated zone of the nodule immediately after RFA and in the follow-up. CEUS was performed after bolus injection of 2.4 ml of SonoVue (Bracco), followed by a 5 ml of normal saline flush.

Before treatment, the volume of thyroid nodules was calculated by the ellipsoid formula: $V = \pi abc/6$ (V is the volume, while a is the largest diameter, b and c are the other two perpendicular diameters). The nodules were divided into two subgroups (< 10 ml and $\geq$ 10 ml) based on the initial volume. Symptom score was self-measured by patients using a 10-cm visual analogue scale (grade 0–10) [2]. The cosmetic score was assessed by a physician (1, no palpable mass; 2, no cosmetic problem but palpable mass; 3, a cosmetic problem on swallowing only; and 4, a readily detected cosmetic problem) [2].

Fig. 1 Flowchart of patient enrolment
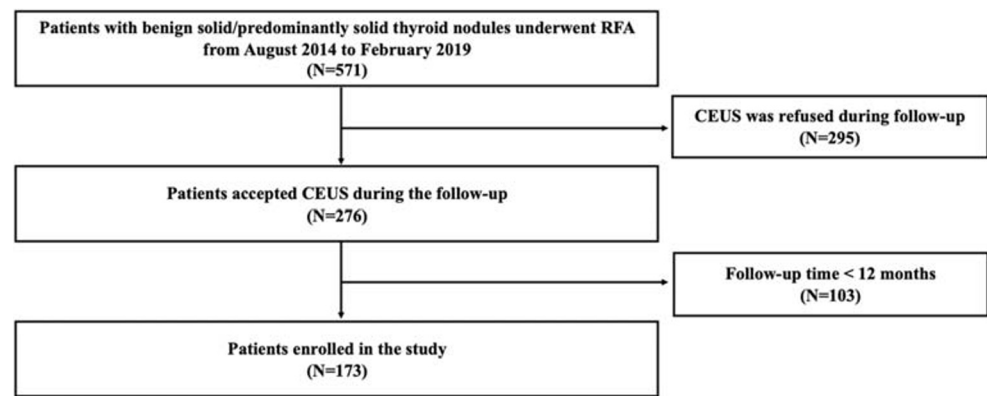
Ablation procedure

All RFA procedures were performed by an experienced US physician with more than 20-year experience in thyroid US and interventional US (Y.K.L.). A bipolar RFA generator (CelonLabPOWER, Olympus Surgical Technologies Europe) and an 18-gauge bipolar RF electrode with 0.9-cm active tip were used (CelonProSurge micro 100-T09, Olympus Surgical Technologies Europe) in this study.

Patients layed on an operating table in the supine position with the neck extended. Local anesthesia with 1% lidocaine was administered. RFA was performed using the trans-isthmic approach, hydrodissection technique, and moving-shot technique. CEUS was performed immediately after the RFA procedure to evaluate the ablation area. If any enhancement existed, a complementary ablation could be performed. Each patient was observed for 1–2 h in the hospital while any adverse event including complication and side effect occurring during and immediately after ablation was evaluated [19].

Post-ablation measurement of Va and assessment

Two physicians (observer A, Y.L. with more than 10 years' experience in thyroid US and CEUS; observer B, X.J. with 3 years' experience in thyroid US and CEUS) performed all the measurements. Before this study, the two observers standardized the measurement method. Va was defined as a decreased hypoechoic zone in the treated nodule without Doppler signal on US [11, 13]. On CEUS, it presented as a non-enhancement zone within the treated nodule during both the arterial phase and venous phase. The anteroposterior and transverse diameters of Va were measured on the transverse US image with the largest dimensions, and the longitudinal diameter was measured on the longitudinal US image with the largest dimensions. Va was measured with the calipers placed outside of the halo [20].

Patients were scanned consecutively by the observers during the same visit. Only one observer was present in the ultrasound room at any time. For each patient, each observer performed a complete new set of scans for measurement, consisting of US and CEUS, without knowledge of the other one's results. During the examination, US was performed first. The three diameters of the hypoechoic zone in the treated nodule were measured twice to calculate the means of each observer. Then the CEUS mode was switched. The real-time microbubble perfusions within ablated nodule and surrounding tissues were observed for a minimum of 2 min and recorded digitally for further analysis. After CEUS images were reviewed, the three diameters of the non-enhancement zone during both phases were measured twice to calculate the means of each observer. When two observers finished examinations, the means of each measurement modality were calculated based on the means of the two observers. Thus, a total of 6 measurements were obtained for each nodule at each follow-up period. The cost and measurement time of each modality were also recorded.

After RFA, patients were followed up at 1, 3, 6, 12 months, and every 12 months thereafter. The nodule volume, Va, VRR, cosmetic, and symptom scores were evaluated during the follow-up period. The volume reduction was calculated as follows: VRR = ([initial volume - final volume] × 100%)/ initial volume. Therapeutic success was defined as a > 50% volume reduction at the last follow-up point [19].

Statistical analysis

Statistical analysis was performed using the SPSS statistical software (version 25.0) and GraphPad Prism (version 8.0.0) software. Continuous data were expressed as mean ± SD (range). Wilcoxon signed-rank tests were used for pairwise comparisons. Reliability was defined as the extent to which measurements can be replicated, which reflects not only the degree of correlation but also an agreement between measurements [21]. The intra- and inter-observer reliability was assessed using intraclass correlation coefficient (ICC) with

95% confidence intervals (CIs) based on absolute agreement, two-way random effects model. Reliability was classified as follows: excellent (ICC > 0.90), good (ICC = 0.75–0.90), moderate (ICC = 0.5–0.74), and poor (ICC < 0.50) [21].

The Bland-Altman analysis was used to evaluate the pairwise agreement of the two measurement modalities. An agreement was expressed as the mean difference with 95% limits of agreement (LOA, mean difference ± 1.96 SD). The mean difference also called bias was the tendency for one modality to underestimate or overestimate the measurement relative to the other [22]. LOA was the range within which 95% of the differences between measurements by the two modalities would lie [23] and expressed the absolute magnitude of the agreement between the two modalities. The width of LOA varied with the precision of the measurements. LOA was wider when measurements were imprecise and vice versa [24]. Before the Bland-Altman analysis, the Kolmogorov-Smirnov test was used to assess the normality of the distribution. If a non-normal distribution was shown, a logarithmic transformation was performed, and the Bland-Altman analysis was applied to the transformed data. Antilogarithm was performed to obtain values relating to the ratios of measurements by the two modalities to fully understand the LOA and easily interpret the results of the Bland-Altman analysis [23, 24]. Moreover, the conclusion on agreement should be made based on the width of LOA in comparison to *a priori*–defined clinical criteria [24, 25]. Because no study has evaluated the agreement of the two modalities on measurements, the clinical criteria of thyroid nodule volume by the ellipsoid formula were used as a reference, which was reported between ± 13.1 and ± 48.6% [26–28]. Therefore, the acceptable agreement in this study should be a LOA ranged from 0.5 to 1.5. The intra- and inter-observer reliability and agreement analysis were performed on the total number of nodules, and then on the subgroups defined by the initial volume before RFA, namely < 10 ml and ≥ 10 ml. The Wilcoxon signed-rank tests were used for pairwise comparisons. Differences with $p < 0.05$ were considered statistically significant.

Results

The clinical characteristics of patients are presented in Table 1. A total of 173 patients (mean age 46.35 ± 12.06 years) with 190 benign thyroid nodules (initial volume of 9.90 ± 12.85 ml) were enrolled in this study. The number of nodules < 10 ml was 129 and ≥ 10 ml was 61. All patients underwent a single session ablation.

Table 1 Clinical characteristics of patients before RFA

Characteristics	Data
No. of patients	173
Age (years)	46.35 ± 12.06 (18–79)
Sex (F/M)	154/19 (80.02/10.98)
No. of nodules	190
No. of patients with number of treated nodules	
One nodule	159 (91.91)
Two nodules	11 (6.36)
Three nodules	3 (1.73)
Largest diameter (mm)	2.95 ± 1.32 (1.10–6.70)
Initial volume (ml)	9.90 ± 12.85 (0.40–71.39)
< 10 ml (N = 129)	3.51 ± 2.73 (0.40–9.41)
≥ 10 ml (N = 61)	23.40 ± 15.20 (10.01–71.39)

Values are presented as mean ± SD (range) or number of tumors (percentages)

During RFA, the mean power was 6.75 ± 2.95 W. The mean RFA time was 385.87 ± 269.32 s and the mean energy was 2390.83 ± 1944.19 J.

The cost of one US and CEUS examination in our country was 16.96 USD and 142.28 USD, respectively. The mean measurement times of each modality by two observers at each follow-up period are showed in Table 2. The measurement times by CEUS were significantly longer than that by US at each follow-up period (all $p < 0.001$).

Efficacy

During a mean follow-up time of 23.17 ± 12.70 months (range 12–67 months), the volume decreased significantly from 9.90 ± 12.85 to 2.20 ± 4.51 ml ($p < 0.001$) with a VRR of 85.63 ± 14.27%. At the last follow-up, the therapeutic success rate was 97.37% (185/190). Symptom score significantly decreased from 2.71 ± 2.15 to 0.94 ± 1.16 ($p < 0.001$). The cosmetic

Table 2 The measurement time of CEUS and conventional US by observers at each follow-up period

Time (months)	CEUS (s)	Conventional US (s)	*p* value
1	649.38 ± 48.45	435.29 ± 51.74	< 0.001
3	650.82 ± 56.20	436.11 ± 52.73	< 0.001
6	650.02 ± 44.45	436.31 ± 46.54	< 0.001
12	652.61 ± 48.48	438.12 ± 45.49	< 0.001
24	653.84 ± 35.33	440.81 ± 38.21	< 0.001

Values are presented as mean ± SD

Conventional US, conventional ultrasound; *CEUS*, contrast-enhanced ultrasound

score significantly decreased from 2.47 ± 1.20 to 1.33 ± 0.56 ($p < 0.001$).

Safety

All the patients were tolerable to the RFA procedure. Side effects like local pain occurred in 16 patients (9.25%) and resolved spontaneously within 3 days. No complications occurred during or after RFA. No patients had side effects or delayed complications related to CEUS.

Intra- and inter-observer reliability

The measurement methods of Va by CEUS and US are shown in Fig. 2 and the results by these two modalities are summarized in Table 3. The measurements by US during the follow-up were all significantly larger than those by CEUS (all $p < 0.001$). Representative cases are shown in Figs. 3 and 4. The intra- and inter-observer reliability of the two measurement modalities are presented in Table 4. The intra- and inter-observer reliability for all the nodules decreased over follow-up time: they were excellent at 1 month, good at 3–6

months, and moderate at 12–24 months, respectively. The inter-observer reliability in nodules < 10 ml was excellent at 1 month (ICC = 0.902), which was only good in nodules ≥ 10 ml (ICC = 0.894). The intra- and inter-observer reliability became moderate in nodules ≥ 10 ml at 6 months (ICC = 0.744) and in nodules < 10 ml at 12 months (ICC = 0.743), respectively.

Agreement

The Bland-Altman analysis of the measurement between the two modalities during the follow-up is shown in Table 5 and Fig. 5. After antilogarithm, the mean differences were all above one and became larger during the follow-up period. LOA also became wider during the follow-up period. The best agreement between the two modalities was in nodules < 10 ml at 1 month with a mean difference of 1.166 and a LOA of 0.413 to 3.294. It means that for approximately 95% of cases, the measurements by US were between 0.413 and 3.294 times the measurements by CEUS, which were larger than the clinical criteria. This was applied to all the reported LOAs hereinafter.
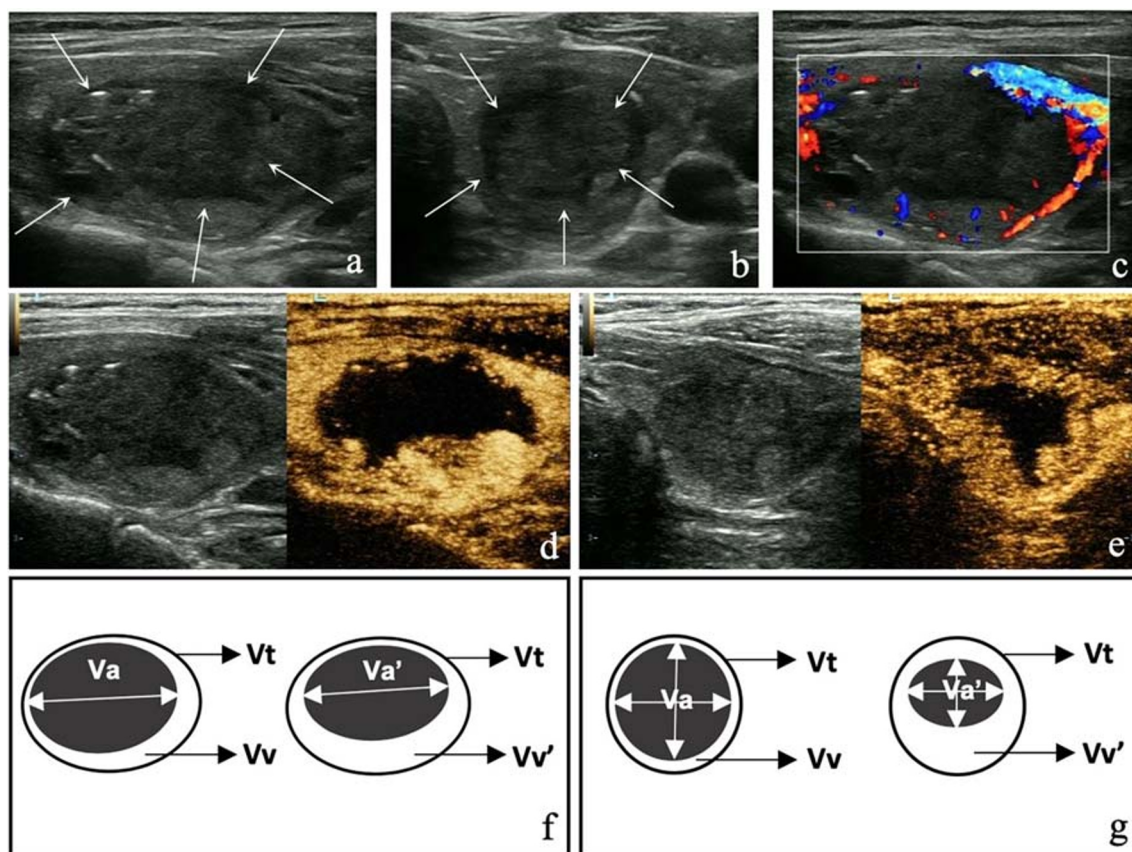
Fig. 2 The measurements of ablated volume (Va) on US and CEUS images of a benign thyroid nodule at 6 months after RFA. **a–c** The longitudinal and transverse US images showed Va present as a decreased hypoechoic zone (arrows) without color signal. **d, e** CEUS showed Va was a non-enhanced zone during both arterial phase and venous phase.

Diagrams (**f, g**) showed the measurement method and the relationship of total volume (Vt), Va, and vital volume (Vv). Va measured by US (arrows) was 2.10 ml and Va measured by CEUS was 1.13 ml. Va was much larger than Va

Table 3 The measurement of Va by CEUS and conventional US

|	Observer A		Observer B		CEUS (ml)	Conventional US (ml)	*p* value
	CEUS (ml)	Conventional US (ml)	CEUS (ml)	Conventional US (ml)			
Total							
1 month	3.13 ± 3.78	3.54 ± 3.96	3.20 ± 3.81	3.47 ± 3.90	3.17 ± 3.80	3.50 ± 3.92	< 0.001
3 months	1.45 ± 2.07	1.98 ± 2.59	1.48 ± 2.09	1.96 ± 2.66	1.47 ± 2.08	1.97 ± 2.61	< 0.001
6 months	1.19 ± 1.89	1.81 ± 3.22	1.17 ± 1.87	1.67 ± 2.88	1.18 ± 1.88	1.77 ± 3.06	< 0.001
12 months	0.73 ± 1.47	1.35 ± 3.00	0.75 ± 1.49	1.47 ± 3.18	0.74 ± 1.48	1.41 ± 3.09	< 0.001
24 months	0.52 ± 0.82	1.21 ± 2.36	0.53 ± 0.79	1.22 ± 2.44	0.53 ± 0.80	1.22 ± 2.40	< 0.001
< 10 ml							
1 month	1.45 ± 1.24	1.66 ± 1.36	1.52 ± 1.30	1.62 ± 1.33	1.49 ± 1.27	1.64 ± 1.34	0.005
3 months	0.64 ± 0.75	0.91 ± 1.05	0.67 ± 0.77	0.87 ± 1.03	0.66 ± 0.76	0.89 ± 1.03	< 0.001
6 months	0.47 ± 0.61	0.64 ± 0.70	0.47 ± 0.61	0.69 ± 0.74	0.47 ± 0.61	0.69 ± 0.72	0.001
12 months	0.30 ± 0.49	0.54 ± 0.68	0.32 ± 0.51	0.59 ± 0.71	0.31 ± 0.50	0.56 ± 0.69	< 0.001
24 months	0.22 ± 0.34	0.44 ± 0.41	0.25 ± 0.37	0.45 ± 0.45	0.23 ± 0.36	0.44 ± 0.42	< 0.001
≥ 10 ml							
1 month	6.67 ± 4.78	7.48 ± 4.69	6.74 ± 4.83	7.35 ± 4.63	6.71 ± 4.81	7.41 ± 4.65	0.012
3 months	3.30 ± 2.82	4.45 ± 3.32	3.35 ± 2.86	4.48 ± 3.46	3.33 ± 2.84	4.46 ± 3.36	0.002
6 months	2.75 ± .2.67	4.36 ± 4.77	2.69 ± 2.64	3.79 ± 4.35	2.72 ± 2.65	4.08 ± 4.55	0.010
12 months	1.53 ± 2.19	2.87 ± 4.66	1.55 ± 2.21	3.12 ± 4.92	1.54 ± 2.20	3.00 ± 4.79	< 0.001
24 months	0.96 ± 1.11	2.37 ± 3.44	0.96 ± 1.04	2.38 ± 3.57	0.96 ± 1.07	2.38 ± 3.50	0.001

Values are presented as mean ± SD (range)

*Conventional US*, conventional ultrasound; *CEUS*, contrast-enhanced ultrasound

Discussion

This study investigated the intra- and inter-observer reliability and agreement between US and CEUS in measuring Va after RFA for benign thyroid nodules. The results showed the Va measured by US were significantly larger than measured by CEUS. The intra- and inter-observer reliability and agreement of the two modalities decreased over follow-up time. The intra- and inter-observer reliability became moderate after 12 months of RFA. In terms of agreements, compared with CEUS, the measurements were overestimated by US. The best agreement was found at 1 month and LOA exceeded the
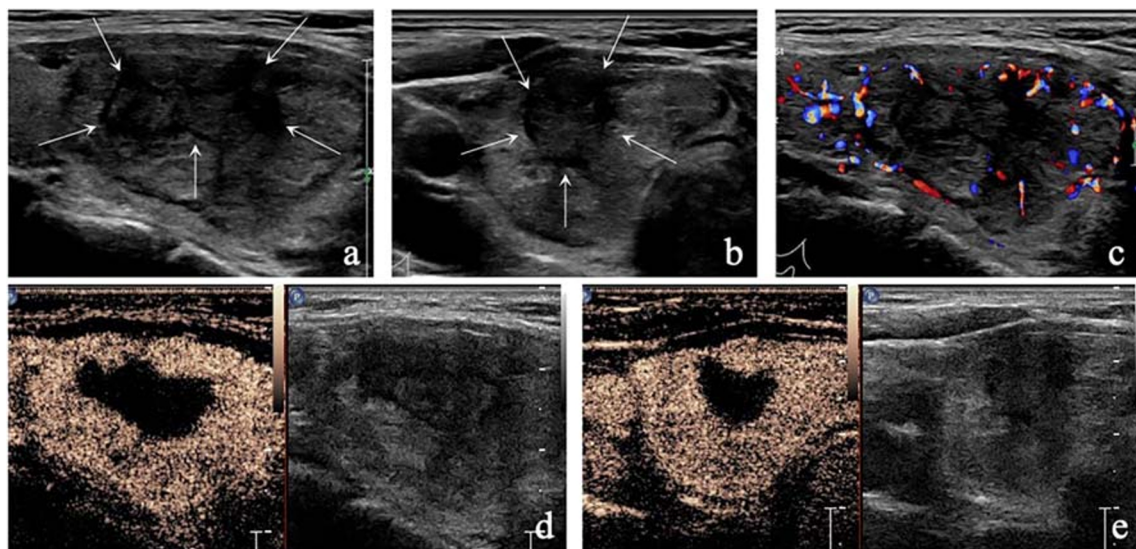
Fig. 3 US and CEUS images of a benign thyroid nodule at 3 months after RFA. **a–c** US showed a relatively clear boundary between Va and Vv. However, Va (1.58 ml) measured by US (arrows) was much larger than Va (0.94 ml) measured by CEUS (**d, e**)
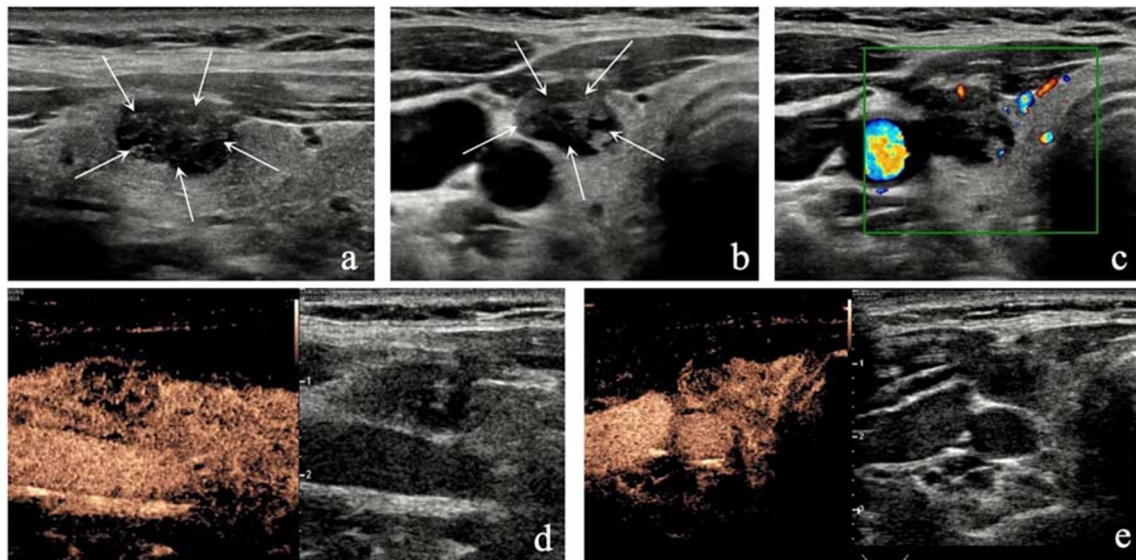
Fig. 4 US and CEUS images of a benign thyroid nodule at 24 months after RFA. **a–c** US showed that the boundary between the Va and Vv was not easily differentiated. Va (0.04 ml) measured by US (arrows) was much larger than Va' (0.01 ml) measured by CEUS (**d**, **e**)

clinical criteria. These results demonstrated US could be neither highly reliable nor provide equivalent results compared to CEUS in the measurement of Va. Therefore, US could not replace CEUS for the measurement of Va.

Although CEUS is a superior method for precise definition of the ablated necrotic zone [18], it has not been widely used to measure Va after thyroid ablation. Some researchers

believed that it could be easy to detect the margin of Va in the treated nodules on US [29]. Other studies found that it was difficult to clarify the boundary between the ablated and vital zone on US, and CEUS should be used [13, 30]. At present, no studies have been investigated the reliability and agreement of the US with CEUS in the measurement of Va after RFA. This study found the intra- and inter-observer reliability of the two

Table 4 The intra-observer and inter-observer reliability of the two measurement modalities

|	Intra-observer reliability		Inter-observer reliability
	Observer A	Observer B	
Total			
1 month	0.943 (0.920–0.960)	0.942 (0.919–0.958)	0.994 (0.921–0.959)
3 months	0.893 (0.829–0.930)	0.886 (0.829–0.922)	0.892 (0.833–0.928)
6 months	0.819 (0.724–0.880)	0.819 (0.730–0.878)	0.820 (0.727–0.880)
12 months	0.747 (0.636–0.823)	0.726 (0.603–0.809)	0.737 (0.620–0.816)
24 months	0.678 (0.385–0.831)	0.610 (0.275–0.792)	0.646 (0.335–0.812)
< 10 ml			
1 month	0.896 (0.842–0.931)	0.898 (0.848–0.931)	0.902 (0.854–0.934)
3 months	0.823 (0.710–0.888)	0.849 (0.766–0.901)	0.838 (0.742–0.896)
6 months	0.811 (0.685–0.884)	0.782 (0.623–0.870)	0.795 (0.648–0.877)
12 months	0.742 (0.547–0.845)	0.739 (0.531–0.845)	0.743 (0.540–0.847)
24 months	0.715 (0.231–0.885)	0.721 (0.324–0.882)	0.726 (0.290–0.888)
≥ 10 ml			
1 month	0.895 (0.810–0.941)	0.891 (0.806–0.939)	0.894 (0.811–0.941)
3 months	0.826 (0.619–0.914)	0.807 (0.608–0.900)	0.821 (0.619–0.910)
6 months	0.740 (0.448–0.874)	0.744 (0.490–0.872)	0.744 (0.477–0.874)
12 months	0.707 (0.468–0.836)	0.681 (0.416–0.823)	0.694 (0.442–0.830)
24 months	0.631 (0.031–0.867)	0.551 (- 0.135–0.835)	0.593 (- 0.430–0.852)

Values are presented as ICC with 95%CIs

Table 5 Bland-Altman analysis of the agreement between conventional US and CEUS

	Log conventional US Vs log CEUS		Antilogarithms	
	Mean difference	LOA	Mean difference	LOA
Total				
1 month	0.145	- 0.792 to 1.081	1.156	0.453–2.948
3 months	0.365	- 0.878 to 1.608	1.440	0.416–4.993
6 months	0.566	- 1.262 to 2.393	1.760	0.283–10.946
12 months	0.792	- 1.181 to 2.765	2.208	0.307–15.879
24 months	0.982	- 0.825 to 2.789	2.669	0.438–16.265
< 10 ml				
1 month	0.153	- 0.885 to 1.192	1.166	0.413–3.294
3 months	0.357	- 0.871 to 1.584	1.428	0.419–4.874
6 months	0.566	- 1.320 to 2.451	1.760	0.267–11.600
12 months	0.761	- 1.187 to 2.709	2.141	0.305–15.014
24 months	0.948	- 0.945 to 2.839	2.581	0.389–17.099
≥ 10 ml				
1 month	0.127	- 0.556 to 0.809	1.135	0.574–2.246
3 months	0.383	- 0.908 to 1.674	1.467	0.404–5.333
6 months	0.566	- 1.165 to 2.296	1.761	0.312–9.934
12 months	0.843	- 1.187 to 2.873	2.323	0.305–17.690
24 months	1.028	- 0.719 to 2.774	2.795	0.487–16.023

Conventional US, Conventional ultrasound; *CEUS*, contrast-enhanced ultrasound; *LOA*, 95% limits of agreement

modalities decreased over the follow-up period. The intra-observer reliability of the two observers was similar, which indicated the experience of observer did not affect the reliability. The intra- and inter-observer reliability were good to excellent in nodules < 10 ml at the first 6 months and were good in nodules ≥ 10 ml at the first 3 months. This indicated the reliability of nodules < 10 ml was better than that of nodules ≥ 10 ml. However, they all became moderate at 12
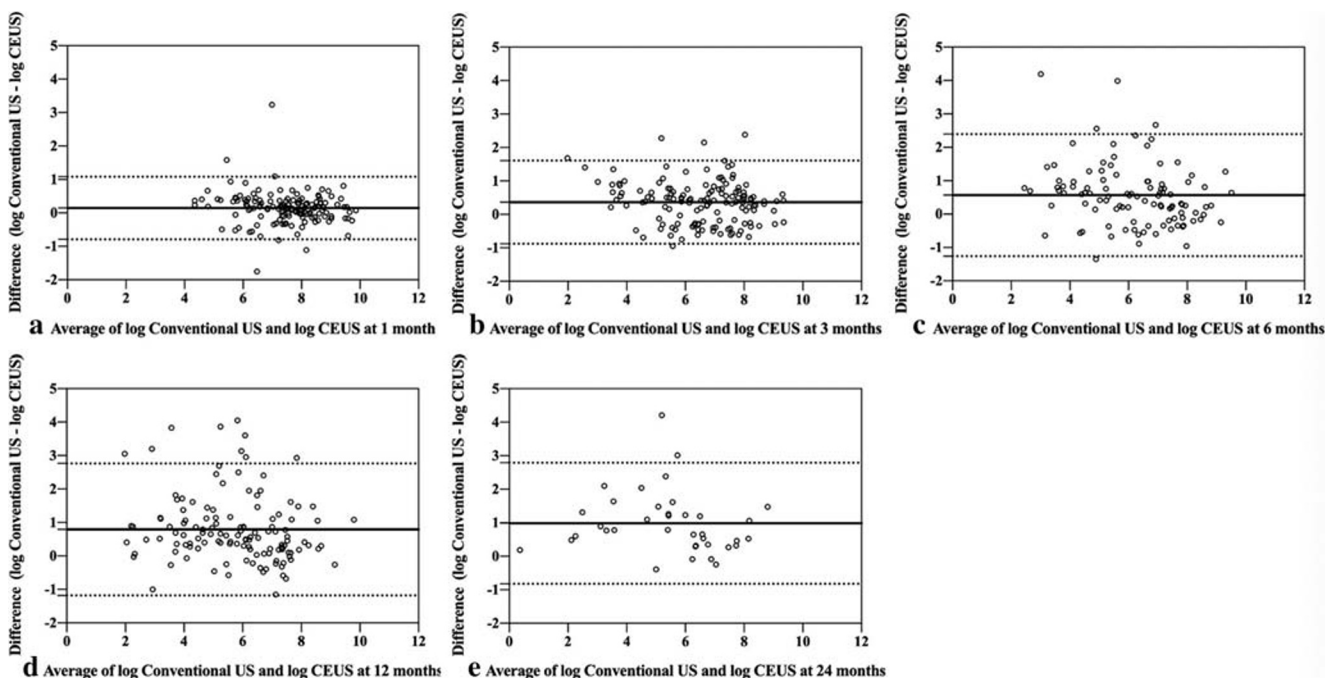
Fig. 5 Bland-Altman plots of agreement of measurements between the two modalities during the follow-up. The Bland-Altman plots of Va by two modalities at 1, 3, 6, 12, and 24 months after RFA were showed from (**a**) to (**e**). Logarithmic transformation was used to show the data. The *x*-axes showed the log means of two measurement modalities of Va. The y-axes showed the log differences between the measurements. Solid lines were the mean difference (bias). Top and bottom dashed lines correspond to upper and lower margins of 95% limits of agreement (LOA)

months after RFA, which suggested that even in the case of nodules with small volume, the intra- and inter-observer reliability was not considerable after 1 year of RFA.

The Bland-Altman analysis was used to evaluate the agreement between the two modalities. This study showed the mean difference was all above one and became larger during the follow-up period. This indicated the measurements by US were all overestimated, which were consistent with the significantly larger measurements by US compared with CEUS. Given that no studies reported the clinical criteria of the two modalities, the conclusion on agreement in this study was established based on the clinical criteria of thyroid nodule volume. Choi et al [27] performed Bland-Altman analysis to evaluate the agreement of thyroid nodule volume and found that the LOA was ± 13.1%. However, Lee et al [26] also used Bland-Altman analysis for evaluation and reported a higher LOA of 44.1%. Brauer et al [28] used a logarithmic method to estimate the inter-observer variation of 48.6% in nodule volume and suggested the volume changes of less than 50% should be considered the measurement variation, which was also recommended by the 2015 American Thyroid Association Guidelines [1]. Therefore, in this study, the LOA ranging between 0.5 and 1.5 was used as the clinical criteria. The best agreement in this study was found in nodules < 10 ml at 1 month and LOA ranged between 0.413 and 3.294, which was much larger than the clinical criteria. Moreover, LOA of all the nodules became much wider and all larger than the clinical criteria during the follow-up period, which indicated that the agreement between the two modalities was unsatisfactory.

Accurate detection and measurements of the true treated volume were essential for a successful evaluation [14]. In recent years, some novel parameters have emerged to evaluate the efficacy of ablation, which were all based on the measurement of Va. Sim et al [11] found that Vv increased occurred about 1 year prior to the nodule regrowth and suggested that Vv increased could be an early sign of regrowth. However, because Vv was equal to Vt minus Va, the overestimated Va with lack of reliability and agreement could have an inevitable impact on the evaluation of nodule regrowth. In addition, a quantitative index IAR was determined as the ratio of Va to Vt at the first month after RFA to predict the therapeutic success of ablation [13]. If IAR was larger than 70%, therapeutic success of ablation could be expected. Although the intra- and inter-observer reliability at 1 month was good to excellent, depending on the initial volume of nodules, US could show an unsatisfactory agreement with CEUS on measurement. Because of the larger Va measured by US, the IAR could be overestimated in some nodules, which could affect the prediction of therapeutic success and the follow-up management. To date, there is no consensus about the indications of CEUS for benign thyroid nodules after thermal ablation. US, as a cost-effective and non-invasive measurement modality, is the most

common method to evaluate thyroid nodules, both in the pre-treatment setting and after treatment [31]. However, its sensitivity and specificity are not susceptible to low-contrast differences between the blood and tissues [32], and color Doppler is not sensitive enough to detect slow and low-volume flow at the level of perfusion [33]. In contrast, despite CEUS being a relatively expensive and time-consuming technique, it can overcome the limitations of US by displaying the parenchymal microvasculature and assess vascularization and tissue perfusion on microbubble contrast agents [14, 16, 34]. CEUS can differentiate the necrotic ablated zone from the ablated nodule clearly [14, 16, 34], which can be an effective modality for the evaluation of efficacy. Therefore, when Va is needed to calculate the novel parameters of efficacy, such as IVR for technique success, or Vv for prediction of regrowth, CEUS should be applied for accurate measurement. Moreover, studies have shown that several factors are related to nodule regrowth after ablation, such as a large initial volume [11, 12, 35, 36], solidity [36], location close to critical structure [12, 35, 37], abundant vascularity [35, 38], low energy applied per volume [35], and 12-month VRR < 50% [39]. If nodules with any of these factors are suspected of regrowth, CEUS should also be considered.

In terms of safety, CEUS has been performed safely in various applications with minimal risk to the patients [17]. The US contrast agent has demonstrated an excellent safety profile with no specific renal, cerebral, or liver toxicity, which allows for repeated administration in the same session when needed [16, 33, 40]. Complications caused by CEUS are rare, and the most frequent adverse events are headache, nausea, chest pain, and chest discomfort [17]. In this study, all the patients were tolerable to CEUS and no complications occurred.

There were some limitations in this study. First, it was a single-center study. Future multicenter studies are needed to confirm our study. Second, the sample size was relatively small. Considering the numbers of each subgroup, the nodules in this study were not divided into three subgroups, which was recommended by the recent reporting criteria of thyroid ablation [16]. Third, the follow-up time was relatively short. Nevertheless, this study showed the intra- and inter-observer reliability and agreement decreased over a mean follow-up time of 23.17 ± 12.70 months. The study will be continued to follow up these patients to obtain more conclusions. Fourth, this study did not compare the intra- and inter-observer reliabilities and agreement between US, CEUS, and microvascular imaging techniques, such as superb microvascular imaging (SMI) in measuring Va. SMI, a recently developed modality, had good sensitivity to differentiate between slow blood-velocity flow signals and movement artifacts within the lesion [41]. SMI not only displayed microvasculature better than US in thyroid nodules but also was a convenient, noninvasive, and cost-effective technique for the patients [41, 42]. Zhao

et al [43] reported that SMI alone was sufficient for evaluation of blood flow in thyroid nodules, and its diagnostic value was comparable to the hypo-enhancement in CEUS for differentiating thyroid cancer. Further investigations for the comparison of the different modalities for the nodule measurements in the follow-up period after ablation are also needed.

In conclusion, the intra- and inter-observer reliability and agreement of US and CEUS in measuring Va were unsatisfactory. CEUS should be considered an effective modality, when Va was needed for further evaluation or in the case of nodules with suspected regrowth.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Yukun Luo.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval Institutional Review Board approval was obtained.

Methodology
• retrospective
• observational
• performed at one institution

References

1. Haugen BR, Alexander EK, Bible KC et al (2015) 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer. Thyroid 26:1–133
2. Kim JH, Baek JH, Lim HK et al (2018) 2017 Thyroid radiofrequency ablation guideline: Korean Society of Thyroid Radiology. Korean J Radiol 19:632–655
3. Deandrea M, Sung JY, Limone P et al (2015) Efficacy and safety of radiofrequency ablation versus observation for nonfunctioning benign thyroid nodules: a randomized controlled international collaborative trial. Thyroid 25:890–896
4. Cesareo R, Pasqualini V, Simeoni C et al (2015) Prospective study of effectiveness of ultrasound-guided radiofrequency ablation versus control group in patients affected by benign thyroid nodules. J Clin Endocrinol Metab 100:460–466
5. Guang Y, He W, Luo Y et al (2019) Patient satisfaction of radiofrequency ablation for symptomatic benign solid thyroid nodules: our experience for 2-year follow up. BMC Cancer 19:147
6. Hamidi O, Callstrom MR, Lee RA et al (2018) Outcomes of radiofrequency ablation therapy for large benign thyroid nodules: a Mayo Clinic case series. Mayo Clin Proc 93:1018–1025
7. Cheng Z, Che Y, Yu S et al (2017) US-guided percutaneous radiofrequency versus microwave ablation for benign thyroid nodules: a prospective multicenter study. Sci Rep 7:9554
8. Radzina M, Cantisani V, Rauda M et al (2017) Update on the role of ultrasound guided radiofrequency ablation for thyroid nodule treatment. Int J Surg 41(Suppl 1):S82–s93
9. Rabuffi P, Spada A, Bosco D et al (2019) Treatment of thyroid nodules with radiofrequency: a 1-year follow-up experience. J Ultrasound 22:193–199
10. Chung SR, Suh CH, Baek JH, Park HS, Choi YJ, Lee JH (2017) Safety of radiofrequency ablation of benign thyroid nodules and recurrent thyroid cancers: a systematic review and meta-analysis. Int J Hyperthermia 33:920–930
11. Sim JS, Baek JH, Lee J, Cho W, Jung SI (2017) Radiofrequency ablation of benign thyroid nodules: depicting early sign of regrowth by calculating vital volume. Int J Hyperthermia 33:905–910
12. Sim JS, Baek JH (2019) Long-term outcomes following thermal ablation of benign thyroid nodules as an alternative to surgery: the importance of controlling regrowth. Endocrinol Metab (Seoul) 34:117–123
13. Sim JS, Baek JH, Cho W (2018) Initial ablation ratio: quantitative value predicting the therapeutic success of thyroid radiofrequency ablation. Thyroid 28:1443–1449
14. Albrecht T, Blomley M, Bolondi L et al (2004) Guidelines for the use of contrast agents in ultrasound. January 2004. Ultraschall Med 25:249–256
15. Lekht I, Gulati M, Nayyar M et al (2016) Role of contrast-enhanced ultrasound (CEUS) in evaluation of thermal ablation zone. Abdom Radiol (NY) 41:1511–1521
16. Claudon M, Dietrich CF, Choi BI et al (2013) Guidelines and good clinical practice recommendations for contrast enhanced ultrasound (CEUS) in the liver - update 2012: A WFUMB-EFSUMB initiative in cooperation with representatives of AFSUMB, AIUM, ASUM, FLAUS and ICUS. Ultrasound Med Biol 39:187–210
17. Sidhu PS, Cantisani V, Dietrich CF et al (2018) The EFSUMB guidelines and recommendations for the clinical practice of contrast-enhanced ultrasound (CEUS) in non-hepatic applications: update 2017 (short version). Ultraschall Med 39:154–180
18. Ma JJ, Ding H, Xu BH et al (2014) Diagnostic performances of various gray-scale, color Doppler, and contrast-enhanced ultrasonography findings in predicting malignant thyroid nodules. Thyroid 24:355–363
19. Mauri G, Pacella CM, Papini E et al (2019) Image-guided thyroid ablation: proposal for standardization of terminology and reporting criteria. Thyroid 29:611–618
20. Frates MC, Benson CB, Charboneau JW et al (2005) Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. Radiology 237:794–800
21. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15:155–163
22. Anvari A, Halpern EF, Samir AE (2018) Essentials of statistical methods for assessing reliability and agreement in quantitative imaging. Acad Radiol 25:391–396
23. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. Stat Methods Med Res 8:135–160
24. Chhapola V, Kanwal SK, Brar R (2015) Reporting standards for Bland-Altman agreement analysis in laboratory research: a cross-sectional survey of current practice. Ann Clin Biochem 52:382–386
25. Mantha S, Roizen MF, Fleisher LA, Thisted R, Foss J (2000) Comparing methods of clinical measurement: reporting standards for bland and altman analysis. Anesth Analg 90:593–602

26. Lee HJ, Yoon DY, Seo YL et al (2018) Intraobserver and interobserver variability in ultrasound measurements of thyroid nodules. J Ultrasound Med 37:173–178

27. Choi YJ, Baek JH, Hong MJ, Lee JH (2015) Inter-observer variation in ultrasound measurement of the volume and diameter of thyroid nodules. Korean J Radiol 16:560–565

28. Brauer VFH, Eder P, Miehle K, Wiesner TD, Hasenclever H, Paschke R (2005) Interobserver variation for ultrasound determination of thyroid nodule volumes. Thyroid 15:1169–1175

29. Baek JH (2017) Factors related to the recurrence of benign thyroid nodules after thermal ablation. Int J Hyperthermia 33:957–958

30. Wang B, Han ZY, Yu J, Liang P (2017) Factors associated with recurrence of BTN after ablation. Int J Hyperthermia 33:959–960

31. Papini E, Pacella CM, Solbiati LA et al (2019) Minimally-invasive treatments for benign thyroid nodules: a Delphi-based consensus statement from the Italian minimally-invasive treatments of the thyroid (MITT) group. Int J Hyperthermia 36:376–382

32. Kang ST, Yeh CK (2012) Ultrasound microbubble contrast agents for diagnostic and therapeutic applications: current status and future design. Chang Gung Med J 35:125–139

33. Alvarez-Sánchez MV, Napoléon B (2014) Contrast-enhanced harmonic endoscopic ultrasound imaging: basic principles, present situation and future perspectives. World J Gastroenterol 20:15549–15563

34. Chong WK, Papadopoulou V, Dayton PA (2018) Imaging with ultrasound contrast agents: current status and future. Abdom Radiol (NY) 43:762–772

35. Wang B, Han ZY, Yu J et al (2017) Factors related to recurrence of the benign non-functioning thyroid nodules after percutaneous microwave ablation. Int J Hyperthermia 33:459–464

36. Lim HK, Lee JH, Ha EJ, Sung JY, Kim JK, Baek JH (2013) Radiofrequency ablation of benign non-functioning thyroid nodules: 4-year follow-up results for 111 patients. Eur Radiol 23:1044–1049

37. Zhao CK, Xu HX, Lu F et al (2017) Factors associated with initial incomplete ablation for benign thyroid nodules after radiofrequency ablation: first results of CEUS evaluation. Clin Hemorheol Microcirc 65:393–405

38. Dossing H, Bennedbaek FN, Hegedus L (2011) Long-term outcome following interstitial laser photocoagulation of benign cold thyroid nodules. Eur J Endocrinol 165:123–128

39. Negro R, Greco G, Deandrea M, Rucco M, Trimboli P (2020) Twelve-month volume reduction ratio predicts regrowth and time to regrowth in thyroid nodules submitted to laser ablation: a 5-year follow-up retrospective study. Korean J Radiol 21:764–772

40. Hudson JM, Williams R, Tremblay-Darveau C et al (2015) Dynamic contrast enhanced ultrasound for therapy monitoring. Eur J Radiol 84:1650–1657

41. Jiang ZZ, Huang YH, Shen HL, Liu XT (2019) Clinical applications of superb microvascular imaging in the liver, breast, thyroid, skeletal muscle, and carotid plaques. J Ultrasound Med 38:2811–2820

42. Zhu YC, Zhang Y, Deng SH, Jiang Q (2018) A prospective study to compare superb microvascular imaging with grayscale ultrasound and color doppler flow imaging of vascular distribution and morphology in thyroid nodules. Med Sci Monit 24:9223–9231

43. Yongfeng Z, Ping Z, Hong P, Wengang L, Yan Z (2020) Superb microvascular imaging compared with contrast-enhanced ultrasound to assess microvessels in thyroid nodules. J Med Ultrason (2001) 47:287–297