White Paper: Reduced Inter-observer Variability of Thyroid Volume Measurements with PIUR tUS Infinity in Comparison to 2D Ultrasound

PIUR IMAGING GmbH, Vienna, Austria

ABSTRACT

Thyroid volumetry is an essential component in thyroid diagnosis and treatment, particularly in post-radiofrequency ablation surveillance. The 2D ultrasound is a conventional method for thyroid assessment, however it is associated with high inter-observer variability. The use of 3D ultrasound has the potential to improve agreement between medical doctors (MDs). This study aims to evaluate inter-observer variability in thyroid and thyroid nodule volume measurements using 3D and 2D ultrasound workflows. A cross-sectional study was conducted with 196 patients. The inter-observer variability for thyroid lobes and nodules was -3.09% ± 16.75% and -1.94% ± 31.03%, respectively, using PIUR tUS Infinity (3D ultrasound), compared to -15.57% ± 49.40% and -9.34% ± 71.53% with conventional 2D ultrasound. The use of 3D ultrasound reduced inter-observer variability by 80% for lobes and 79% for nodules. These findings demonstrate that 3D ultrasound significantly reduces inter-observer variability in thyroid volumetry compared to conventional 2D ultrasound.

Keywords: 3D Ultrasound, PIUR tUS Infinity, Thyroid diagnosis, Thyroid volume, Inter-observer variability.

1. INTRODUCTION

Thyroid and thyroid nodule volume measurements play a crucial role in thyroid diagnosis and treatment.

The thermal ablation, such as radiofrequency ablation (RFA), is used to treat cosmetic or pressure-related problems associated with nodules. This requires post-RFA surveillance of the nodule volume. [1] A difference of more than 50% in nodule volume during follow-up is considered clinically significant for thyroid nodule evaluation and might indicate nodule regrowth. [2], [3]

Accurate nodule volume assessment is also essential for malignancy risk stratification.[2]

Conventional thyroid volume assessment relies on 2D ultrasound using the ellipsoid formula. However, this method is associated with high inter-observer variability [4]. Because the ellipsoid formula merely approximates the thyroid's true shape, it contributes to higher observer variability. [5]

The use of 3D ultrasound has been shown to improve agreement between medical doctors [6], [7].

This study aimed to assess whether PIUR tUS Infinity (3D ultrasound) reduces inter-observer variability compared to conventional 2D ultrasound.

2. METHODS AND MATERIAL

Data Collection

All patients who participated in the study had given their consent to participate in the study in writing. The collected data was part of a 510(k) FDA submission and was used for internal purposes. The study was conducted in various clinical settings across Europe, Brazil, and the United States. Data collection, which included volume assessment of lobes and nodules, occurred between the 15th of November 2023 and the 7th of August 2024.

The inclusion criteria were following: 1) Patients were referred to the thyroid sonography as a part of clinical routine; 2) Patients who are 22 years old or older.

Patients whose scans were not readable or did not contain objects of interest were excluded.

Data was collected and evaluated by ten different health professionals using the following ultrasound systems and probes: Alpinion ECUBE 8 equipped with probes L3-12HWD and L3-12T, XCUBE 70 (SL3-19H), GE LS7 Exp (ML6-15), GE Logiq S8 (ML6-15), Mindray MX7 (L13-3Ns), GE Logiq P9 (ML6-15).

Study Design

The study followed cross-sectional study design. The volume of thyroid lobe and nodules (if present) were evaluated by two different medical doctors (MDs) using two distinct workflows: 2D conventional ultrasound workflow and PIUR tUS Infinity workflow. Inter-observer variability was assessed separately for each workflow by comparing the volume measurements recorded by the two MDs. The inter-observer variabilities of two different workflows were compared.

2D Conventional Ultrasound (2D) Workflow

2D conventional ultrasound workflow were based on 2D scans acquired by linear 2D-caliper measurements using B-Mode ultrasound of two orthogonal planes. The method used for calculation of the thyroid volume was the ellipsoid formula: V[ml] = π/6 × Width × Height × Length [8].

PIUR tUS Infinity (3D) Workflow

PIUR tUS Infinity is a computer-aided solution, illustrated on Figure 1. One of its applications is to aid in the diagnostic workflow of thyroid and thyroid nodules. The MD acquired 3D scans using the PIUR tUS Infinity by sweeping a standard probe of 2D ultrasound, equipped with a PIUR sensor, across the region of interest. Importantly, the PIUR sensor is attached directly to the probe of the same compatible third-party ultrasound device that acquired the scans for 2D conventional ultrasound workflow, so there is no separate or independent scanning step. This ensures that the same ultrasound machine and patients are used for both 2D and 3D data acquisition, allowing a consistent and direct comparison between the two protocols.

2D ultrasound images and position data, generated by the system integrated PIUR Sensor, are the basis for 3D image reconstruction. Computer vision and machine learning algorithms are applied to a sequence of ultrasound images. The solution identifies the thyroid lobe and proposes its margins to the MD who then can make adjustments to the segmentation and approve the final result. Based on this, a lobe volume is being calculated. With additional user input, thyroid nodules can be marked, quantified, and visualized as multiplanar reconstructions or 3D volume renderings.

Statistics

The data was collected in Excel software, and statistical analysis was performed using Python.

The inter-observer variability (IOV) of the 2D conventional ultrasound workflow between the first and second medical doctors (2D MD1 vs. 2D MD2) and the PIUR tUS Infinity workflow (3D MD1 vs. 3D MD2) was calculated using the following formula, as proposed by Popović et al. [9]:

$$IOV = \frac{V_1 - V_2}{\frac{V_1 + V_2}{2}} \times 100\% \tag{1}$$

where V_1 represents a volume measurement performed by first MD; V_2 represents a volume measurement performed by second MD.

The Bland-Altman plots [10] were used to illustrate inter-observer variability by plotting the percentage difference between the two MDs' measurements against their mean. The mean difference (bias) and the 95% limits of agreement, calculated as the mean difference ± 1.96 × SD, were determined to quantify the range within which most differences lie.

To test whether there was a significant difference between the IOVs of the 2D conventional ultrasound workflow and the PIUR tUS Infinity workflow, the Wilcoxon signed-rank test was applied with a significance level of $\alpha = 0.05$. The alternative hypothesis (H_1) was that IOVs of PIUR tUS Infinity were lower than those of 2D conventional ultrasound. The distribution of thyroid lobe and nodule volumes did not follow a normal distribution.

Figure 1 PIUR tUS Infinity interface. The screenshot is a generic image of PIUR tUS Software used in a marketing purpose.

3. RESULTS AND DISCUSSION

The study included 196 patients with 203 nodules. The patient cohort was 84.13% female and 15.87% male, with a mean age of 54.35 ± 13.57 years. The ethnic distribution was 67.55% White, 13.83% African American, 6.91% Hispanic, and 2.59% Asian, with the remaining patients classified as "other" or absent ethnicity data.
The results for inter-observer variability evaluation for PIUR tUS Infinity workflow and 2D conventional ultrasound workflow are presented in the Table 1.

Table 1 Volume measurement results for PIUR tUS Infinity and 2D conventional ultrasound workflow

Object	2D Conventional Ultrasound	PIUR tUS Infinity
Lobes Mean IOV ± SD [%]	-15.57 ± 49.40	-3.09 ± 16.75
Lobes LOA [%]	[81.42, -112.56]	[29.79, -35.97]
Nodules Mean IOV ± SD [%]	-9.34 ± 71.53	-1.94 ± 31.03
Nodules LOA [%]	[131.20, -149.87]	[59.02, -62.90]

Abbreviations: IOV – Inter-observer Variability, SD – Standard Deviation, LoA – Limits of Agreement.

Mean thyroid lobe and nodule volumes measured by the first MD were 7.79 ± 7.39 ml and 2.89 ± 5.98 ml, respectively, using 2D ultrasound, and 10.74 ± 9.67 ml and 2.76 ± 6.11 ml using 3D ultrasound. Corresponding volumes measured by the second MD were 9.08 ± 8.21 ml and 3.20 ± 6.33 ml (2D) and 11.08 ± 10.08 ml and 2.87 ± 6.24 ml (3D).

The p-value of Willcoxon sign-rank test was less than 0.05 for both, lobes (p-value = 2.85 $\times 10^{-15}$) and nodules (p-value = 3.12 $\times 10^{-7}$), indicating that the IOV of PIUR tUS Infinity is significantly lower than IOV of 2D conventional ultrasound.

The Bland-Altman plots for 2D Conventional Ultrasound lobes and nodules analysis are presented in Figure 5 and Figure *3*, while for PIUR tUS Infinity results are shown in Figure 2 and Figure 4.
The results indicate a narrower vertical spread in the differences between the two medical doctors using the same workflow for PIUR tUS Infinity compared to the 2D conventional ultrasound workflow.

The study conducted by Lee et al. [4] reported an interobserver variability of 15.3% ± 16.6% based on 73 patients and 122 nodules, using 2D conventional

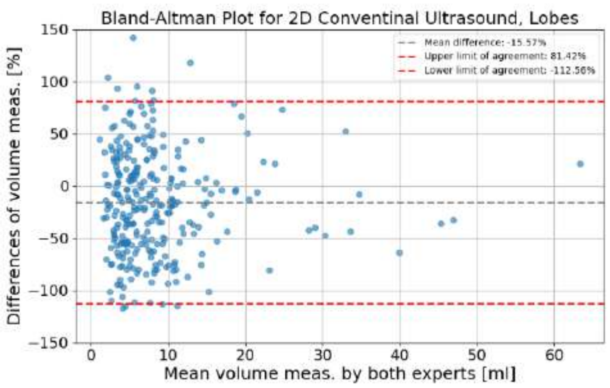

Figure 3 Bland-Altman Plot Using 2D Conventional Ultrasound for Lobes

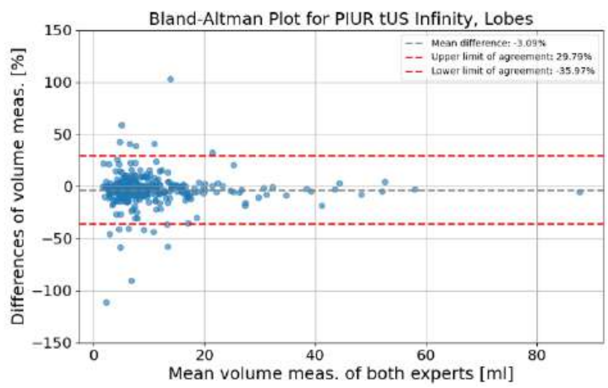

Figure 2 Bland-Altman Plot Using PIUR tUS Infinity for Lobes

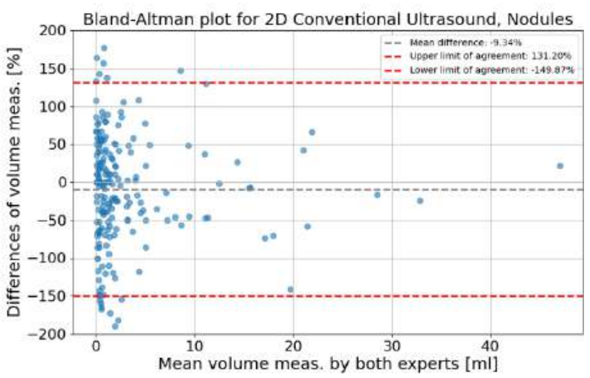

Figure 5 Bland-Altman Plot Using 2D Conventional Ultrasound for Nodules

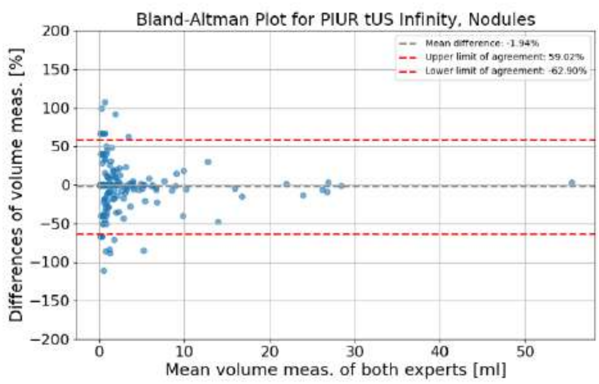

Figure 4 Bland-Altman Plot Using PIUR tUS Infinity for Nodules

ultrasound, with medical doctors' experience ranging from 6 months to 10 years.

The inter-observer variability in the study by Kroenke et al. [11] was reported for thyroid lobes in millilitres using 2D ultrasound and 3D ultrasound (acquired using PIUR tUS Infinity with image processing and thyroid segmentation volume segmentation using ImFusion Suite Version 2.9.7) ultrasound and based on 28 patients. The inter-observer variability in percentage was derived using formula (1). For the 2D ultrasound inter-observer variability was 10.75%, 16.75% and 27.37%, while 3D ultrasound showed IOV 12.85%, 0.92% and 11.93%.

The study by Brauer et al. [12] demonstrated an inter-observer variability of 23.69% for lobes and 48.96% for nodules using 2D ultrasound with the ellipsoid formula, based on 40 patients and 38 nodules.

The inter-observer variability reported in the literature for 2D ultrasound lobe measurements aligns with our findings. However, the variability for 2D ultrasound nodule measurements in previous studies is considerably higher than in our results. The inter-observer variability for 3D ultrasound lobe measurements observed in our study falls within the range reported in the literature (0.92–12.85%).

4. CONCLUSION

This study demonstrates that PIUR tUS Infinity (3D ultrasound) significantly reduces inter-observer variability compared to conventional 2D ultrasound.

PIUR tUS Infinity reduced inter-observer variability by 80% for lobes and 79% for nodules. These findings highlight the potential of 3D ultrasound with semi-automatic segmentation to improve thyroid volumetry by reducing operator dependency.

5. DISCLOSURE

*The data for this study was initially collected as part of a 510(k) FDA submission for internal use. Informed consent was obtained from all patients prior to data collection. The study was conducted and prepared by the internal employees of PIUR IMAGING GmbH.

6. ACKNOWLEDGMENTS

Robert Bauer, Frederik Bender, Dilnaz Sugralina and Magdalena Vogel actively participated in the study.

7. REFERENCES

[1] H. J. Kim, J. H. Baek, W. Cho, and J. S. Sim, "Long-term follow-up of the radiofrequency ablation of benign thyroid nodules: the value of additional treatment," *Ultrasonography*, vol. 41, no. 4, pp. 661–669, Oct. 2022, doi: 10.14366/usg.21231.

[2] B. R. Haugen *et al.*, "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*, vol. 26, no. 1, pp. 1–133, Jan. 2016, doi: 10.1089/thy.2015.0020.

[3] W.-J. Moon *et al.*, "Ultrasonography and the Ultrasound-Based Management of Thyroid Nodules: Consensus Statement and Recommendations," *Korean J. Radiol. Off. J. Korean Radiol. Soc.*, vol. 12, pp. 1–14, Jan. 2011, doi: 10.3348/kjr.2011.12.1.1.

[4] H. J. Lee *et al.*, "Intraobserver and Interobserver Variability in Ultrasound Measurements of Thyroid Nodules," *J. Ultrasound Med.*, vol. 37, no. 1, pp. 173–178, Jan. 2018, doi: 10.1002/jum.14316.

[5] J. Brunn, U. Block, G. Ruf, I. Bos, W. P. Kunze, and P. C. Scriba, "Volumetrie der Schilddrüsenlappen mittels Real-time-Sonographie*1," *DMW - Dtsch. Med. Wochenschr.*, vol. 106, no. 41, pp. 1338–1340, Mar. 2008, doi: 10.1055/s-2008-1070506.

[6] S. Tong, H. N. Cardinal, R. F. McLoughlin, D. B. Downey, and A. Fenster, "Intra- and inter-observer variability and reliability of prostate volume measurement via two-dimensional and three-dimensional ultrasound imaging," *Ultrasound Med. Biol.*, vol. 24, no. 5, pp. 673–681, Jun. 1998, doi: 10.1016/S0301-5629(98)00039-8.

[7] P. Seifert *et al.*, "Optimization of Thyroid Volume Determination by Stitched 3D-Ultrasound Data Sets in Patients with Structural Thyroid Disease," *Biomedicines*, vol. 11, no. 2, p. 381, Jan. 2023, doi: 10.3390/biomedicines11020381.

[8] J. Sotos and N. Tokar, "Testicular volumes revisited: A proposal for a simple clinical method that can closely match the volumes obtained by ultrasound and its clinical application," *Int. J. Pediatr. Endocrinol.*, vol. 2012, no. 1, Jun. 2012, doi: 10.1186/1687-9856-2012-17.

[9] Z. Popovic and J. Thomas, "Assessing observer variability: a user's guide," *Cardiovasc. Diagn. Ther.*, vol. 7, no. 3, Jun. 2017, doi: 10.21037/cdt.2017.03.12.

[10] J. M. Bland and D. G. Altman, "Statistics Notes: Measurement error proportional to the mean," *BMJ*, vol. 313, no. 7049, pp. 106–106, Jul. 1996, doi: 10.1136/bmj.313.7049.106.

[11] M. Krönke *et al.*, "Tracked 3D ultrasound and deep neural network-based thyroid segmentation reduce interobserver variability in thyroid volumetry," *PLOS ONE*, vol. 17, no. 7, p. e0268550, Jul. 2022, doi: 10.1371/journal.pone.0268550.

[12] V. F. H. Brauer, P. Eder, K. Miehle, T. D. Wiesner, H. Hasenclever, and R. Paschke, "Interobserver Variation for Ultrasound Determination of Thyroid Nodule Volumes," *Thyroid*, vol. 15, no. 10, pp. 1169–1175, Oct. 2005, doi: 10.1089/thy.2005.15.1169.