Discrepancies between the ultrasonographic and gross pathological size of papillary thyroid carcinomas

Soo Yeon Hahn¹, Jung Hee Shin¹, Young Lyun Oh², Young-Ik Son³
¹Department of Radiology and Center for Imaging Science, Thyroid Center, ²Department of Pathology, Thyroid Center, ³Department of Otorhinolaryngology-Head and Neck Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Purpose: The goal of this study was to investigate the level of agreement between tumor sizes measured on ultrasonography (US) and in pathological specimens of papillary thyroid carcinomas (PTCs) and to identify the US characteristics contributing to discrepancies in these measurements.

Methods: We retrospectively reviewed the US findings and pathological reports of 490 tumors in 431 patients who underwent surgery for PTC. Agreement was defined as a difference of <20% between the US and pathological tumor size measurements. Tumors were divided by size into groups of 0.5–1 cm, 1–2 cm, 2–3 cm, and ≥3 cm. We compared tumors in which the US and pathological tumor size measurements agreed and those in which they disagreed with regard to the following parameters: taller-than-wide shape, infiltrative margin, echogenicity, microcalcifications, cystic changes in tumors, and the US diagnosis.

Results: The rate of agreement between US and the pathological tumor size measurements was 64.1% (314/490). Statistical analysis indicated that the US and pathological measurements significantly differed in tumors <1.0 cm in size (P=0.033), with US significantly overestimating the tumor size by 0.2 cm in such tumors (P<0.001). Cystic changes were significantly more frequent in the tumors where US and pathological tumor size measurements disagreed (P<0.001).

Conclusion: Thyroid US may overestimate the size of PTCs, particularly for tumors <1.0 cm in size. This information may be helpful in guiding decision making regarding surgical extent.

Keywords: Thyroid cancer, papillary; Ultrasonography; Pathology, surgical; Dimensional measurement accuracy

Introduction

Ultrasonography (US) is widely accepted as the technique of choice for the preoperative staging of papillary thyroid carcinoma (PTC). US examination is useful for assessing the size, location, number, and characteristics of thyroid nodules [1–3]; however, US is highly dependent on both the instrument and the operator [4–6].

Previous guidelines have recommended total thyroidectomy as the primary initial surgical treatment option for nearly all differentiated thyroid cancers greater than 1 cm with or without evidence of loco-
Materials and Methods

The Institutional Review Board of our institution approved this study, and the requirement for informed consent was waived.

Patients

A total of 655 PTCs were surgically resected from 539 patients at our institution between March 2006 and December 2006. Among these 539 patients, 12 patients with 43 multiple PTCs were excluded from this study due to confusion in correlations between tumor size by preoperative US and tumor size by postoperative pathologic measurement. Thirty-one patients with 31 PTCs were excluded due to a lack of preoperative US imaging, and 10 patients with 10 PTCs were excluded due to a delay longer than 12 months between US and surgery (mean delay of the subjects who were included, 4.7 months; range, 16 days to 12 months). In addition, 55 patients with 81 PTCs were also excluded because the tumors measured <0.5 cm. Ultimately, 431 patients with 490 PTCs were included in the analysis. Data were obtained for the study through a review of pathological and US reports.

Ultrasonography

All lesions were examined using a 7–15 MHz linear array transducer (HDI 5000, Advanced Technology Laboratories, Bothell, WA, USA) or a 5–12 MHz linear array transducer (LOGIQ700, GE Medical Systems, Milwaukee, WI, USA). US examination was performed by one of three board-certified radiologists or one of two senior residents, all aware of the patients’ clinical records.

The US findings were described, including shape, margin, internal echogenicity of the mass, the presence of microcalcifications, cystic changes in tumors, and US diagnosis. The shape was categorized as being wider-than-tall or taller-than-wide. Taller-than-wide shape was defined as a mass that was greater in its anteroposterior dimension than its transverse dimension. The margin was classified as either infiltrative or circumscribed. Internal echogenicity was classified as marked hypoechogenicity or other. Marked hypoechogenicity was defined as a lower level of echogenicity than the surrounding strap muscle. The presence or absence of microcalcifications and cystic changes in tumors (none, <50%, and ≥50%) was also evaluated.

The US diagnosis was suspicious for malignancy if a thyroid mass had at least one malignant US feature, including microcalcifications, taller-than-wide shape, an infiltrative margin, or marked hypoechogenicity [12], and as benign if suspicious US features were absent. If more than one malignant mass was detected by US examination, the US findings of each tumor were evaluated separately.

Statistical Analysis

In order to simplify the statistical analysis of the data, only the maximum diameter was measured in centimeters using US. The formalin-fixed specimen was measured, and thyroid tumor size was based on the largest diameter. Although several methods exist for determining tumor size, the pathological measurement is considered to be the gold standard and is recognized as such by the American Joint Committee on Cancer (AJCC) [13]. Therefore, we used the pathological tumor size as the gold standard in our study.

Intraclass correlation coefficient (ICC) analysis was used to evaluate the concordance between tumor size as measured by US and the size that was confirmed on pathology. ICC values between 0.4 and 0.75 may be taken to represent fair to good reliability, and values above 0.75 represent excellent reliability [14]. For the purposes of this study, tumors were divided into four groups based on their size: (1) 0.5–1 cm, (2) 1–2 cm, (3) 2–3 cm, and (4) ≥3 cm. Tumor size agreement was defined as a difference of less than 20% between the US and pathological tumor size measurements. Using Pearson’s chi-square test, we compared the rates of agreement and disagreement between US and pathological tumor size in each of these four groups. In the tumors where the measurements disagreed, correlations between US and pathological size were assessed using the Wilcoxon signed-rank test with continuity correction.

We compared the US findings in tumors for which the measurements agreed and those that demonstrated disagreement. US findings were compared using the Pearson’s chi-square test and Fisher’s exact test as appropriate. Statistical software SAS ver. 9.1.3 (SAS Institute Inc., Cary, NC, USA) was used for all data analyses.
P-values <0.05 were considered to indicate statistical significance.

**Results**

All patients were surgically treated through total thyroidectomy (n=416) or lobectomy (n=15). The mean tumor size was 1.20 cm (range, 0.3 to 6.9 cm) as determined by preoperative US and 1.16 cm (range, 0.5 to 7.5 cm) as determined by postoperative pathology. The patients ranged in age from 12–80 years, with a median of 47 years. Of the 431 patients, 354 (82.1%) were female. Single tumors were found in 371 patients (86.1%), and two tumors were found in a single patient in 13.9% of the cases.

The histopathological diagnoses were classic PTC in 93.9% of cases (460/490), follicular variant PTC in 4.7% of cases (23/490), diffuse sclerosing variant PTC in 0.6% of cases (3/490), oncocytic variant PTC in 0.4% of cases (2/490), and solid variant PTC in 0.4% of cases (2/490).

The ICC between the US and pathological tumor size measurements was 0.957, indicating excellent reliability. The rate of agreement between US and pathological tumor size measurements was 64.1% (314/490). In the tumors where the measurements disagreed, US overestimated the tumor size in 115 cases (23.5%) and underestimated it in 61 cases (12.4%). The tumors in which the measurements agreed are compared with those in which the measurements disagreed according to pathological tumor size in Table 1. The agreement rates between US and pathological tumor size were 58.5% (159/272) in tumors between 0.5 cm and 1.0 cm, 70.5% (110/156) in tumors between 1.0 cm and 2.0 cm, 75.0% (27/36) in tumors between 2.0 cm and 3.0 cm, and 69.2% (18/26) in tumors larger than 3.0 cm. The tumor size displayed significant variation depending on the measurement modality in tumors <1.0 cm (P=0.033). In this subgroup (≥0.5 cm and <1.0 cm), a total of 113 tumors were overestimated (77 tumors, 68.1%) or underestimated (36 tumors, 31.9%) by US. In this group, statistical analysis indicated that the use of US led to significant overestimation of the tumor size by an average of 0.2 cm in tumors <1.0 cm (P<0.001) (Table 2). Among the 272 pathologically measured subcentimeter tumors, 51 tumors (18.8%) were classified as tumors ≥1.0 cm by US measurements. In contrast, 18 of the 218 tumors with a pathological size ≥1.0 cm (8.3%) were classified as subcentimeter tumors by US measurements.

Among the US findings, only cystic changes were significantly more frequent in tumors for which the US and pathological measurements disagreed (P<0.001) (Table 3). However, no significant differences were found in the prevalence of taller-than-wide shape (P=0.598), an infiltrative margin (P=0.519), marked hypoechoigenicity (P=0.918), microcalcifications (P=0.533), and a US diagnosis of suspicious for malignancy (P=0.590) between the two groups.

**Discussion**

The 2015 American Thyroid Association Management Guidelines

---

**Table 1.** Comparison of tumors showing agreement and disagreement between ultrasonographic and pathological size measurements according to tumor size on pathology

| Group        | Tumor size on pathology (cm) | Total | P-value |
|--------------|-----------------------------|-------|---------|
|              | 0.5–0.9                     | 1.0–1.9 | 2.0–2.9 | ≥3.0 |
| Agreement    | 159 (58.5)                  | 110 (70.5) | 27 (75.0) | 18 (69.2) | 314 | 0.033* |
| Disagreement | 113 (41.5)                  | 46 (29.5) | 9 (25.0) | 8 (30.8) | 176 |
| Overestimation| 77 (28.3)                   | 29 (18.6) | 6 (16.7) | 3 (11.5) | 115 |
| Underestimation| 36 (13.2)                   | 17 (10.9) | 3 (8.3) | 5 (19.3) | 61 |
| Total        | 272                         | 156    | 36      | 26     | 490 |

The data are numbers of lesions. The numbers in the parentheses are percentages (%).

*P-values of <0.05 were regarded as indicating statistical significance.

**Table 2.** Comparison of tumor size differences between ultrasonography (US) and pathology in tumors for which these measurements disagreed

| Tumor size on pathology (cm) | Median tumor size difference between US and pathology (cm) | P-value |
|------------------------------|----------------------------------------------------------|---------|
| 0.5–0.9                      | 0.2                                                      | <0.001* |
| 1.0–1.9                      | 0.3                                                      | 0.079   |
| 2.0–2.9                      | 0.6                                                      | 0.260   |
| ≥3.0                         | −1.0                                                     | 0.232   |

*P-values of <0.05 were regarded as statistically significant.
for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer contain revised indications for thyroid lobectomy as the initial treatment for papillary and follicular carcinomas [8]. These indications include small (<1 cm), low-risk, unifocal, intrathyroidal papillary carcinomas in the absence of prior head and neck irradiation or radiologically or clinically involved cervical nodal metastases [8]. Therefore, the accurate preoperative determination of T and N categories on US is significant for ensuring the appropriate treatment of patients with PTC [15–17] and avoiding unnecessary total thyroidectomy.

In this study, we evaluated the agreement (i.e., concordance defined as a difference less than 20%) between papillary carcinoma size as measured by US and the size confirmed on pathology. We found that the agreement rate between US and pathological tumor size measurements was 64.1% (314 of 490 tumors), and that the tumor size between US and pathology was significantly discordant in tumors <1.0 cm. The accuracy of preoperative US for assessing tumor size has been previously evaluated using various methods in a small set of studies. However, this is the only study to include sufficient numbers of subcentimeter tumors when evaluating the agreement between US and pathological tumor size measurements of PTCs. Bachar et al. [11] reported significant discrepancies between the pathological size of solitary PTCs and their estimated size on the preoperative US scan for tumors measuring larger than 1.5 cm on US. The percentage of tumors ≤1.0 cm was 11.6% (34/292) in that study [11], corresponding to 59.4% (291/490) in our study. Deveci et al. [9] demonstrated that the agreement in the size of thyroid nodules measured by US and surgical pathology examination was ≤50%, except in the ≤1.0 cm size range (78.5%). In their study, however, both benign and malignant thyroid nodules with diverse histologic diagnoses were included in the study population. Yoon et al. [10] reported that papillary carcinoma sizes measured postoperatively were consistently significantly smaller than the US-estimated sizes, with a mean percentage difference of 9.9%. However, although they evaluated differences in the size of the individual tumors measured by preoperative US and postoperative pathology, they did not assess the significance of agreement between the measurement methods.

According to our analysis, the use of US led to the significant overestimation of the tumor size by an average of 0.2 cm in tumors <1.0 cm. In other words, there is a low probability that papillary microcarcinomas would be incorrectly estimated to be larger than 1.0 cm, thereby leading to a more extensive operation than necessary. According to our study, 51 of 272 pathologically measured subcentimeter tumors (18.8%) were classified as tumors with a size ≥1.0 cm by US measurement. Previously, Deveci et al. [9] reported that the rate of agreement between US and pathology measurements for thyroid nodules decreased as the tumor size increased. In their study, the agreement rate between the US and pathology size was 78.5% in tumors ≤1.0 cm and ≤56.0% in tumors ≥1.1 cm. However, as previously stated, their study population incorporated both benign and malignant thyroid nodules, including 61 benign (58.0%) and 44 malignant (42.0%) thyroid nodules ≤1.0 cm in size.

In the present study, we identified differences in US characteristics between the tumors for which US and pathological size
measurements agreed and those for which they did not. Cystic changes were significantly more frequent in the group that exhibited disagreement (14.2% vs. 4.1%), but this was the only US finding that demonstrated significant variation between these two groups (Table 3). Compression of a thyroid nodule during US, aspiration of the cystic fluid by preoperative fine needle aspiration biopsy, and tissue shrinkage during fixation could be reasons for the size discrepancy between the US and pathological measurements of thyroid nodules with cystic changes [9]. Other factors, such as tumor histology and growth pattern, tissue fixation and processing, prior biopsy, and prior therapy, can likewise affect the final measured tumor size.

Although US is the most sensitive imaging modality available for the examination of the thyroid gland and thyroid-associated abnormalities, the major limitations of US in thyroid imaging are its operator dependency and interobserver and intraobserver variability [18-20]. According to previous studies, interobserver and intraobserver variability can affect US measurements of thyroid nodules [21–23]. Moreover, size measurement by US may be of limited accuracy in nodules with a vague margin, irregular shape, small size, and also in conglomerated masses of small nodules [6]. US measurements may also be of limited validity in patients with a short neck, a large goiter, or a thyroid nodule located in the lower portion of the neck [4,24]. Since our institution is a tertiary referral center for thyroid surgery, a considerable number of patients with thyroid nodules in our study had undergone fine needle aspiration prior to visiting our institution. Prior fine needle aspiration biopsy can alter thyroid nodule size due to subsequent hemorrhage and/or scarring [4,25]. Therefore, US evaluation prior to fine needle aspiration biopsy can reduce the error rate in the US measurements of thyroid nodules [26,27].

According to a previous report [9], the presence of calcifications or coexisting thyroiditis was not found to significantly influence the accuracy of nodule size measurements on US. In this study, we found that the prevalence of microcalcifications was not significantly different between the groups. However, we did not analyze the role of coexisting thyroiditis.

Our study had several limitations. First, we analyzed imaging findings and pathological data retrospectively. This retrospective approach may have prevented us from identifying accurate US findings in real time, which could have influenced the evaluation and the size measurement of the tumors on US. However, this is well representative of clinical practice and was not manipulative. Second, data were obtained for the study through a review of pathological and US reports, and we were unable to assess interobserver and intraobserver variability. Third, we did not consider the experience of the US operator, although US examination of the thyroid is highly operator-dependent. However, all of the US examinations in our institution were performed by radiologists specializing in thyroid imaging. Finally, the mean tumor size in our study was small (≤1.2 cm), which may have limited our evaluation of tumor characteristics on preoperative thyroid US.

In conclusion, the ultrasonographic tumor size agreed with the pathological tumor size in 64.1% of PTCs, allowing for a difference of less than 20%. Thyroid US could lead to a significant overestimation of the tumor size of PTCs, particularly for tumors <1.0 cm. Among the US findings of PTCs, cystic changes in tumors were associated with overestimation of the size in US imaging study. These findings may be helpful in guiding decision-making about surgical extent.

ORCID: Soo Yeon Hahn: http://orcid.org/0000-0002-4099-1617; Jung Hee Shin: http://orcid.org/0000-0001-6435-7357; Young Lyun Oh: http://orcid.org/0000-0002-9127-4642; Young-Ik Son: http://orcid.org/0000-0003-0570-6953

Conflict of Interest
No potential conflict of interest relevant to this article was reported.

Acknowledgments
This study was supported in part by the Research Fund of the Korean Society of Ultrasonin in Medicine.

References
1. Miskin M, Rosen IB, Walfish PG. Ultrasonography of the thyroid gland. Radiol Clin North Am 1975;13:479-492.
2. Ramsay I, Meire H. Ultrasonics in the diagnosis of thyroid disease. Clin Radiol 1975;26:191-197.
3. Rosen IB, Walfish PG, Miskin M. The ultrasound of thyroid masses. Surg Clin North Am 1979;59:19-33.
4. Gallo M, Pesenti M, Valcavi R. Ultrasound thyroid nodule measurements: the "gold standard" and its limitations in clinical decision making. Endocr Pract 2003;9:194-199.
5. Nafisi Moghadam R, Shajari A, Afkhami-Ardekani M. Influence of physiological factors on thyroid size determined by ultrasonict. Acta Med Iran 2011;49:302-304.
6. Knudsen N, Bols B, Bulow I, Jorgensen T, Perrild H, Ovesen L, et al. Validation of ultrasonography of the thyroid gland for epidemiological purposes. Thyroid 1999;9:1069-1074.
7. American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer, Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid 2009,19:1167-1214.
8. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, 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 2016;26:1-133.

9. Deveci MS, Deveci G, LiVolsi VA, Gupta PK, Baloch ZW. Concordance between thyroid nodule sizes measured by ultrasound and gross pathology examination: effect on patient management. Diagn Cytopathol 2007;35:579-583.

10. Yoon YH, Kwon KR, Kwak SY, Ryu KA, Choi B, Kim JM, et al. Tumor size measured by preoperative ultrasonography and postoperative pathologic examination in papillary thyroid carcinoma: relative differences according to size, calcification and coexisting thyroiditis. Eur Arch Otorhinolaryngol 2014;271:1235-1239.

11. Bachar G, Buda I, Cohen M, Hadar T, Hilly O, Schwartz N, et al. Size discrepancy between sonographic and pathological evaluation of solitary papillary thyroid carcinoma. Eur J Radiol 2013;82:1899-1903.

12. Moon WJ, Jung SL, Lee JH, Na DG, Baek JH, Lee YH, et al. Benign and malignant thyroid nodules: US differentiation--multicenter retrospective study. Radiology 2008;247:762-770.

13. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. AJCC cancer staging manual. 7th ed. New York: Springer, 2010.

14. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30-46.

15. Gonzalez HE, Cruz F, O'Brien A, Goni I, Leon A, Clare R, et al. Impact of preoperative ultrasonographic staging of the neck in papillary thyroid carcinoma. Arch Otolaryngol Head Neck Surg 2007;133:1258-1262.

16. Stulak JM, Grant CS, Farley DR, Thompson GB, van Heerden JA, Hay ID, et al. Value of preoperative ultrasonography in the surgical management of initial and reoperative papillary thyroid cancer. Arch Surg 2006;141:489-494.

17. Nam SY, Shin JH, Han BK, Ko EY, Ko ES, Hahn SY, et al. Preoperative ultrasonographic features of papillary thyroid carcinoma predict biological behavior. J Clin Endocrinol Metab 2013;98:1476-1482.

18. Brauer VF, Eder P, Miehle K, Wiesner TD, Hasenclever H, Paschke R. Interobserver variation for ultrasound determination of thyroid nodule volumes. Thyroid 2005;15:1169-1175.

19. Jarlov AE, Nygard B, Hegedus L, Karstrup S, Hansen JM. Observer variation in ultrasound assessment of the thyroid gland. Br J Radiol 1993;66:625-627.

20. Choi SH, Kim EK, Kwak JY, Kim MJ, Son EJ. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. Thyroid 2010;20:167-172.

21. Wienke JR, Chong WK, Fielding JR, Zou KH, Mittelstaedt CA. Sonographic features of benign thyroid nodules: interobserver reliability and overlap with malignancy. J Ultrasound Med 2003;22:1027-1031.

22. Hegedus L. Thyroid size determined by ultrasound. Influence of physiological factors and non-thyroidal disease. Dan Med Bull 1990;37:249-263.

23. Lyschik A, Drozd V, Schloegl S, Reiners C. Three-dimensional ultrasonography for volume measurement of thyroid nodules in children. J Ultrasound Med 2004;23:247-254.

24. Sannazzari P, Menozzi PG, Belloni L. Relation between the parathyroids and the other endocrine glands. Note 9. Parathyroid histomorphological changes after administration of TSH, desiccated thyroid and methlythiouracil. Arch Maragliano Patol Clin 1959;15:1295-1305.

25. Baloch ZW, LiVolsi VA. Post fine-needle aspiration histologic alterations of thyroid revisited. Am J Clin Pathol 1999;112:311-316.

26. Pandit AA, Vaideeswar P, Mohite JD. Infarction of a thyroid nodule after fine needle aspiration biopsy. Acta Cytol 1998;42:1307-1309.

27. Gordon DL, Fisak M, Fisher SG. Changes in thyroid nodule volume caused by fine needle aspiration: a factor complicating the interpretation of the effect of thyrotropin suppression on nodule size. J Clin Endocrinol Metab 1999;84:4566-4569.