

D5: Factors Impacting the Performance and Heterogeneity of Studies Assessing the Reliability of 2-Dimensional Shear Wave Elastography for Thyroid Malignancy Diagnosis

Abubakar Nattabi H, Yahya N, Ahmad R, Mohamad M, Yussof AN, Mohamed Sharif N, Mohd Zaki F

Diagnostic Imaging and Radiotherapy Programme, School of Diagnostic & Applied Health Sciences, Faculty of Health Sciences, Universiti Kebangsaan Malaysia

ABSTRACT

Introduction: 2-dimensional shear wave elastography (2D-SWE) is a quantitative technique used in the evaluation of thyroid nodule stiffness to diagnose malignancy. High heterogeneity has been reported among the available 2D-SWE studies but the possible causes have remained unsubstantiated. This study aims to analyze the factors causing the high heterogeneity in diagnostic performance between studies and whether these factors impact on reliability of 2D-SWE for thyroid malignancy classification. **Methods:** The literature search was carried out in PubMed, Scopus and Web of Science databases. The final 14 studies were reviewed and information on study design was retrieved including level of pre-compression (LoP) [none-minimal, unclear], plane of measurement (PoM) [sagittal, transverse, unclear], region of interest placement (ROIP) [whole lesion, stiffest, unclear], region of interest (ROI) size (ROIS) [largest, 2mm/less, unclear, variable], elasticity index (EI) parameters used [mean, maximum, minimum] and optimal EI parameter (OEIP). Pooled sensitivity and specificity estimates plus area under receiver operating characteristic curve and their respective 95% confidence intervals (CI) were used to assess 2D-SWE diagnostic performance among subgroups. Subgroup heterogeneity was significant at I²=50% and marked at I²=75%. A difference of ≥5 units in subgroup diagnostic performance and I² ≥ 25% in heterogeneity was deemed considerable. Univariate meta-regression of subgroup models and multivariate regression of methodological [LoP, ROIP, ROIS, OEIP] and spectrum composition [country, patient inclusion bias, malignancy prevalence] models was performed, p-value < 0.05 showed significance. Variability in sensitivity and specificity was evaluated through the threshold effect using Spearman's correlation coefficient between logit true and false positive ratios. **Results:** We analyzed 14 studies that included 2851 thyroid nodules (1092 malignant, 1759 benign) from 2139 patients. Subgroup analysis revealed considerably higher diagnostic performance for unclear vs none-minimal [AUC 0.88 (CI=0.81-0.96) vs 0.83 (CI=0.80-0.86)] and for stiffest vs whole lesion [AUC 0.83 (CI=0.79-0.86) vs 0.78 (CI=0.73-0.84)]. Considerable disparity in sensitivity existed among all subgroups, however, a sizeable specificity estimate difference was observed only for mean vs maximum OEIP subgroups [0.84 (0.81-0.86) vs 0.64 (0.59-0.70)], China vs other countries [0.82 (0.79-0.85) Vs 0.76 (0.73-0.78) and low vs high bias [0.77 (0.74-0.79) vs 0.82 (0.79-0.84)]. Between subgroup heterogeneity was significant (p<0.005, I² >50%) for both sensitivity and specificity estimates. ROIP, ROIS and study population were found to be potential independent predictors of DOR on univariate meta-regression (p=0.002, <0.001 and <0.001 respectively). On multivariate meta-regression, ROIS emerged as the only significant covariate affecting 2D-SWE diagnostic performance (p=0.022). There was a moderate threshold effect noted (r=0.494, p=0.032). **Conclusion:** ROIS was a consistently predominant covariate for 2D-SWE diagnostic performance in this study. However, the influence of LoP cannot be understated as the disparity in applied LoP was ubiquitous. Broadly, 2D-SWE study design standardization is critical in pursuit of a reference thyroid malignancy threshold EI. Otherwise the accuracy, reliability and reproducibility of 2D-SWE in the thyroid gland shall remain a medical quandary.

KEY WORDS:

2-Dimensional Shear Wave Elastography, heterogeneity, thyroid Ma