

PP 406 - CLASSIFICATION OF THYROID NODULES BY ULTRASOUND IN CLINICAL PRACTICE: THE ADDED VALUE OF THE JUDGMENT OF THE SKILLED ENDOCRINOLOGIST

B. Madeo¹, G. Brigante², A. Ansaloni², E. Taliani², S. Kaleci³, M. L. Monzani², M. Simoni², V. Rochira²

¹Department of Medical Specialties, Azienda Ospedaliero-Universitaria di Modena Modena, ²Dept. Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia Modena,

³Department of Diagnostic Medicine, Clinics and Public Health, Azienda Ospedaliero-Universitaria di Modena Modena

Background and aim. Several ultrasound (US) classifications for thyroid nodules have been proposed. Since most of them are hardly applicable in clinical practice, we set up the Modena US Thyroid Classification (MUT) that stratifies the risk of malignancy considering both knowledge derived from scientific literature and clinician subjective impression. The aim of the present study was to test the diagnostic accuracy of different thyroid US classification systems, AACE/ACE-AME, American Thyroid Association (ATA), British Thyroid Association (BTA), and MUT, and to evaluate inter-classification agreement. **Methods.** We prospectively enrolled 111 patients (33M, 78F; age 19-75) candidate for surgery because of indeterminate, suspicious or malignant cytology. All patients underwent neck US before surgery and a score according to MUT was assigned: 1 not certainly nodular; 2 not suspect; 3 indeterminate; 4 suspect; 5 very suspect. Then, we retrospectively classified nodules according to AACE/ACE-AME, ATA and BTA, thanks to the detailed collection of each nodule US features in a preformed checklist. US pattern was related to histology. Sensitivity, specificity, diagnostic cut-off value and accuracy of each classification were calculated. The overall agreement between classifications was quantified by Bland-Altman test. The agreement between single nodule analysis by different classifications was evaluated considering Weighted Cohen's Kappa. **Results.** Fifteen patients had uninodular and 96 multinodular goiter, for a total of 457 nodules. MUT has the highest accuracy (AUC 0.808) and specificity (89%), followed by ATA and BTA, and finally by AACE/ACE-AME. ATA and BTA are highly interchangeable and MUT is comparable to both of them. AACE/ACE-AME is the least interchangeable with all the other classifications. Considering agreement between single nodule analyses by different classifications, ATA and BTA had the best ($\kappa=0.723$); AACE/ACE-AME showed slight agreement with BTA ($\kappa=0.177$) and MUT ($\kappa=0.183$), and fair agreement with ATA ($\kappa=0.282$); MUT had fair agreement with both ATA ($\kappa=0.291$) and BTA ($\kappa=0.271$). **Conclusions.** Our data analysis to quantify the agreement between different classification systems confirms the reliability and reproducibility to classify malignancy. However, results bring out the limit in specificity of the current reference classifications, which improves when the subjective impression of the clinician is considered.