## Poster 213

*Thyroid Cancer Friday Poster Clinical*

# REVIEW OF STUDIES EVALUATING THE DIAGNOSTIC ACCURACY OF THE AFIRMA® GENE EXPRESSION CLASSIFIER IN INDETERMINATE THYROID NODULES

N. Busaidy[3], Q. Duh[4], C. Rahilly-Tierney[1,2], G. Randolph[5], H. Gharib[6]

[1]Strategic Research Partners LLC, Boston, MA; [2]Medicine, Harvard Medical School, Boston, MA; [3]Endocrine Neoplasia and Hormonal Disorders, MD Anderson Cancer Center, Houston, TX; [4]Surgery, University of California San Francisco, San Francisco, CA; [5]Otolaryngology, Massachusetts Eye and Ear Institute, Boston, MA; [6]Medicine, Mayo Clinic, Rochester, MN

Afirma® Gene Expression Classifier (GEC) determines malignancy risk for Bethesda System of Reporting Thyroid Cytopathology (BSRTC) III/IV (indeterminate) thyroid nodules, and has been validated in a blinded, multicenter, prospective clinical trial. Several authors have purported to examine the diagnostic accuracy of GEC in observational studies. We use a customized Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS2) tool to evaluate the quality of these studies. We searched Medline and EMBASE from January 1, 2010 until March 15, 2016, for original, full-length studies that evaluated the diagnostic accuracy of GEC. We developed a customized QUADAS2 tool consisting of signaling questions that evaluated 4 domains of each study: patient population, index test, reference standard diagnosis, and flow and timing. For each study, these questions will used by a panel of reviewers to identify whether methodologic flaws exist that introduce bias into the study's findings and/or limit the applicability of the study's conclusions. Each of the reviewers will independently apply the tool to each included study, and disagreements will be discussed in conference. Once reviewers agree on answers for all questions, these selected answers will be presented in tables according to the QUADAS2 format. We have identified 11 studies that have purported to evaluate diagnostic accuracy of GEC. Four panelists have been recruited to evaluate these studies using the customized QUADAS2 tool. In studies evaluated to date, the reviewers have identified potential sources of bias, including lack of follow-up of unoperated, GEC-benign subjects who are subsequently excluded from the calculation of specificity and negative predictive value (NPV). In a blinded, multi-center,

prospective study in which all subjects were assigned reference standard diagnoses and included in the analysis, GEC was demonstrated to have a high sensitivity and NPV. Subsequent studies have not fully evaluated the diagnostic accuracy of the test as they have suffered from bias introduced by disproportionate exclusion of GEC-benign subjects from the analyses, and therefore do not substantially add to current understanding of the performance of the test.