IS GALECTIN-3 A GOOD METHOD FOR THE DETECTION OF MALIGNANCY IN PATIENTS WITH THYROID NODULES AND A CYTOLOGIC DIAGNOSIS OF “FOLLICULAR NEOPLASM”? A CRITICAL APPRAISAL OF THE EVIDENCE

Alvaro Sanabria, MD, MS,1,2 André Lopes Carvalho, MD, PhD,2 Victor Piana de Andrade, MD, PhD,3 Juan Pablo Rodrigo, MD,4,5 José Guilherme Vartanian, MD, PhD,2 Alessandra Rinaldo, MD,6 Mauro Kazuo Ikeda, MD, PhD,2 Kenneth Owen Devaney, MD, JD,7 José Magrin, MD, PhD,2 Fernando Augusto Soares, MD, PhD,3 Alfio Ferlito, MD, DLO, DPath, FRCS, FRCSEd, FRCsI, FRCPath, MCAP,6 Luiz Paulo Kowalski, MD, PhD2

1 Department of Surgery, Universidad de La Sabana, Bogotá, Colombia
2 Department of Head and Neck Surgery and Otorhinolaryngology, Hospital do Câncer AC Camargo, São Paulo, Brazil
3 Department of Pathology, Hospital do Câncer AC Camargo, São Paulo, Brazil
4 Department of Otolaryngology, Hospital Universitario Central de Asturias, Oviedo, Spain
5 Surgical Oncology Division, Instituto Universitario de Oncología del Principado de Asturias, Oviedo, Spain
6 Department of Surgical Sciences, ENT Clinic, University of Udine, Udine, Italy. E-mail: a.ferlito@uniud.it
7 Department of Pathology, Foote Hospital, Jackson, Michigan

Accepted 26 January 2007
Published online 24 May 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hed.20642

Abstract: Background. Thyroid nodules are the most common surgical disease of the thyroid. Fine-needle aspiration biopsy (FNAB) is the most commonly employed tool for establishing a diagnosis. However, 15% to 25% of FNAB reports yield inconclusive results. Immunostaining of cytological smears from FNAB with galectin-3 has been proposed as a tool for differentiating between benign and malignant nodules. We performed a systematic review to evaluate the utility of galectin-3.

Methods. Prospective studies of nodules with FNAB reports of “follicular neoplasm” and with a definitive diagnosis confirmed by histopathology were selected. Calculations of individual sensitivity, specificity, and positive and negative likelihood ratios were made.

Results. The articles selected were those with the best methodological quality.

Conclusion. Galectin-3 could be a good tool to guide therapeutic decision in patients with thyroid nodules and FNAB results of follicular neoplasm, but available information has methodological flaws that precludes a definitive answer about galectin-3 utility in the clinical setting. ©2007 Wiley Periodicals, Inc. Head Neck 29: 1046–1054, 2007

Keywords: galectin-3; thyroid neoplasm; meta-analysis; follicular neoplasm; diagnostic test

Thyroid nodules are the most common surgical disease of the thyroid gland. Fine-needle aspiration biopsy (FNAB) is the diagnostic method recommended to establish the pathologic diagnosis of...
Galectin-3 to Diagnose Malignancy in Patients with Thyroid Nodules

HEAD & NECK—DOI 10.1002/hed

November 2007

1047

thyroid nodules, and its operative characteristics have been assessed widely in the medical literature.\(^1\)\(^{-}\)\(^{3}\) FNAB helps to differentiate between benign and malignant nodules and to design treatment plans. Moreover, the routine use of FNAB in the evaluation of thyroid nodules can reduce the need of thyroidectomy for diagnostic purposes by 20\% to 50\%, while increasing the yield of cancer diagnoses in pathologic specimens by 15\% to 45\%. However, based on current criteria, 10\% to 25\% of FNAB specimens are marked by high cellularity, a microfollicular pattern and scant colloid, resulting in inconclusive results, commonly reported as “follicular pattern,” a diagnosis which does not rule out malignancy.\(^4\)\(^{-}\)\(^{7}\)

In these cases, the cytological diagnosis cannot define the nature of the nodule as benign or malignant, because definitive pathologic characterization relies on the finding of a well-defined capsule and its invasion by tumor cells. Almost 75\% to 80\% of patients who underwent surgery based on such a FNAB diagnosis showed an incomplete capsule (hyperplastic nodules) or a well-defined capsule not crossed over by tumors cells (benign neoplastic nodules), ultimately classified as benign lesions.\(^4\)\(^{-}\)\(^{7}\)

Therefore, any measure that could decrease yet further the number of thyroid surgical procedures for diagnostic purposes would represent a significant advance in the management of thyroid nodules.

Previously, some authors have advocated the use of several markers, such as CD44, telomerase, high mobility group I (Y) protein [HMGI(Y)], oncogenic fibronectin, cyclooxygenase-2 (COX-2), the antihuman mesothelial cell mouse monoclonal antibody HBME-1, and galectin-3,\(^8\)\(^{-}\)\(^{9}\) and imaging techniques as ultrasonography with doppler analysis\(^10\)\(^{-}\)\(^{12}\) of thyroid nodules, as useful tools to differentiate benign from malignant lesions. Among the molecular markers described, galectin-3 is a promising marker that is known to be overexpressed in malignant thyroid lesions.\(^9\)

Galectin-3 is a β-galactoside-binding protein normally localized in the cytoplasm that plays a role in a great variety of cellular pathways in different tissues, including embryogenesis, inflammation, and apoptosis inhibition.\(^13\)

Its role in cancer development has been related to a mutation that impairs its ability to modulate apoptosis.\(^14\)

It has been suggested that galectin-3 might serve as a molecular tool for differentiating between malignant and benign thyroid nodules.\(^13\)

Many studies have been carried out to assess the operative characteristics of galectin-3 as a marker to define the histological nature of follicular neoplasms obtained from FNAB, some going so far as to suggest that galectin-3 staining can be found in a substantial proportion of thyroid malignancies (both follicular and papillary carcinomas of the thyroid), but only infrequently in benign processes such as multinodular goiters or follicular adenomas. Unfortunately, some of these studies have not followed the methodological recommendations for a diagnostic test study, which can overestimate the sensitivity and specificity of the test. To resolve these discrepancies, we carried out a systematic review to assess the operative characteristics of galectin-3 as a molecular marker to define treatment approach in patients with thyroid nodules and a FNAB diagnosis of “follicular pattern.”

MATERIALS AND METHODS

We performed a bibliography search on MEDLINE and EMBASE databases for studies published from January 1980 to September 2006 using the terms “thyroid nodule,” “thyroid cancer,” “galectin,” “galectin-3,” “diagnostic test,” “sensitivity,” “specificity,” and “operative characteristics.” An expanded search was used with each relevant article using Boolean operators. References were explored to identify other articles. We included only studies published in the English language.

Study Selection. After a preliminary search, all abstracts were reviewed by the authors, and those that dealt with thyroid nodules and galectin-3 evaluation as a diagnostic test were selected for further analysis. Criteria for inclusion of studies in the review were: use of galectin-3 alone or with other molecular markers in the cytological specimen as an immunocytochemical marker; prospective collection of data (study directionality) meaning that immunostaining results were obtained first and then compared to histopathological results; patients studied first by FNAB with a report of follicular neoplasm and final histopathology results for all cases. Independent evaluation was not considered an important factor in this kind of study, in light of the fact that satisfying all of the conditions described earlier should have ensured that galectin-3 staining result were recorded before the definitive results of histopathology report had been learned.

Investigators acquired data about sample size, characteristics of patients, methods of galectin-3 detection, form of evaluation and absolute number of true-positive, true-negative, false-positive, and
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Sample size</th>
<th>Patients</th>
<th>Method</th>
<th>Evaluation</th>
<th>Flaws</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inohara et al.</td>
<td>1999</td>
<td>14 patients. Not stated if indeterminate sample</td>
<td>Prospective collection obtained by fine-needle aspiration</td>
<td>Cell isolation by immunomagnetic isolation technique and galectin-3 detection made with immunoblotting</td>
<td>Patients submitted to surgery</td>
<td>Small sample size. Indeterminate characteristic of the nodule not stated. Analysis by different grades of staining (+ to ++++) was not made.</td>
</tr>
<tr>
<td>Bartolazzi et al.</td>
<td>2001</td>
<td>226 patients in a multicentric study. 82 patients with indeterminate sample</td>
<td>Prospective collection obtained by ultrasound-guided fine-needle aspiration</td>
<td>Avidin-biotin complex</td>
<td>Patients previously submitted to surgery, reasons unknown. Positivity for galectin assessed by 2 independent pathologists. Positivity was considered when more than 10% of cells stained.</td>
<td>Analysis by different grades of staining (+ to ++++) was not made.</td>
</tr>
<tr>
<td>Collet et al.</td>
<td>2005</td>
<td>55 patients with indeterminate samples. Only 34 patients had a gold standard confirmation with surgery</td>
<td>Prospective collection with 26-27-gauge needle with indeterminate nodules (no definitive cytological diagnosis or slightest atypia)</td>
<td>Two smears used for galectin-3 immunohistochemistry. Avidin-biotin complex. Samples spread on slides and analyzed after air-drying</td>
<td>Surgical decision not modified by galectin-3 immunohistochemistry results. Positivity for galectin assessed by 2 independent investigators. Positivity was considered when more than 20% of cells stained.</td>
<td>Analysis by different grades of staining (+ to ++++) was not made. Small sample size</td>
</tr>
<tr>
<td>Rossi et al.</td>
<td>2005</td>
<td>99 patients with cytological diagnosis of follicular origin. 37 patients with follicular pattern and surgical gold standard</td>
<td>Prospective collection obtained by ultrasound-guided fine-needle aspiration. Clear characteristics stated to define indeterminate nodule.</td>
<td>Two smears used for galectin-3 immunohistochemistry. Avidin-biotin complex</td>
<td>Positivity for galectin assessed by 2 independent pathologists. Positivity was considered when more than 50% of cells stained.</td>
<td>Small sample size. Analysis by different grades of staining (+ to ++++) was not made.</td>
</tr>
<tr>
<td>Saggiorato et al.</td>
<td>2005</td>
<td>125 patients with cytological diagnosis of follicular neoplasm</td>
<td>Prospective collection obtained by fine-needle aspiration. Preparation in alcohol-fixed cell blocks.</td>
<td>Avidin-biotin complex</td>
<td>Positivity was considered when more than 10% of cells stained.</td>
<td>Analysis by different grades of staining (+ to ++++) was not made.</td>
</tr>
<tr>
<td>Kim et al.</td>
<td>2006</td>
<td>146 patients with cytological diagnosis of follicular neoplasm. 55 patients with indeterminate sample lesions diagnosed by cytology</td>
<td>Prospective collection obtained by fine-needle aspiration. Preparation in ethyl alcohol-fixed cell blocks embedded in paraffin</td>
<td>Avidin-biotin complex</td>
<td>Focally positive considered when 10% to 50% of cells stained, and diffusely positive when more than 50% of cells stained</td>
<td>Analysis by different grades of staining (+ to ++++) was not made.</td>
</tr>
</tbody>
</table>
false-negative results from the articles to perform the statistical analyses.

Sensitivity, specificity, and positive and negative likelihood ratios were calculated for each study.15

RESULTS

The primary search identified 37 articles.16–53 Only 6 of these studies satisfied the inclusion criteria delineated earlier (Table 1).17,20,31,45,48,53 The reasons for excluding the other articles were: galectin-3 was studied only in surgical specimens and not in FNAB smears,16,18,19,21–28,30,32–36,38,40,41,43,44,49–52 inclusion of samples for cytological analysis from nodules not classified as follicular pattern,29,37,39,43 and possibility of duplicity or redundancy of results because same authors or periods of time.41,42,46,47 Results of true-positive, true-negative, false-positive, and false-negative results of the selected articles are reported in Table 2.

Global prevalence of malignancy in nodules classified as follicular pattern was 51.6% (range 21.9% to 64.3%).

Operative characteristics of each article are presented in Table 3.

DISCUSSION

Thyroid nodules are a common indication for referral of patients to a head and neck surgery clinic. The estimated prevalence of symptomatic thyroid nodules is close to 5% in the general population, and it increases to approximately 40%, if imaging methods, such as ultrasonography, are used as screening devices.54 As a consequence, this significant number of patients with thyroid nodules will, potentially, require some form of medical and/or surgical investigation of these findings. At the same time, however, it is also known that most of these nodules will prove to be benign, and so as a consequence definitive surgical management would not be required in those patients.54

FNAB is considered by many to be the method of choice for assessment of thyroid nodules. Since the 1990s, it has been repeatedly demonstrated that FNAB has a high sensitivity and specificity when it permits a diagnosis of a clearly benign or malignant entity; indeed, the use of FNAB has been responsible for the better selection of patients for surgical treatment with a resulting decrease in the number of thyroidectomies for diagnostic purposes.3,55,56 However, FNAB is highly dependent on aspiration technique and cytopathologist expertise, with a well-known subjective category in the middle of the spectrum, where cancer cannot be ruled out, namely “follicular pattern.”4–7 Surgical intervention is called for in this context because current diagnostic criteria for recognition of certain types of thyroid malignancies include architectural features not assessable in the cytologic specimen. This category includes benign lesions such as follicular adenomas and goiter nodules, and malignant lesions such as the follicular variant of papillary carcinoma and follicular carcinomas.

Previous studies have demonstrated that only 15% to 25% of pathologic reports of thyroidectomies performed for follicular neoplasm yield malignant diagnoses.5–7 This means that almost two thirds of patients with this FNAB cytologic diagnosis actually harbor benign lesions and undergoes surgery procedures without clinical advantage, but with a predictable risk of complications.57,58 Any method that would serve to decrease the number of unnecessary diagnostic thyroidectomies could help to improve the management of patients with follicular neoplasm.

One of the recently recommended and tested molecular markers touted for use in this context is galectin-3. Galectin-3 is a member of the β-galac-
tosidase-binding family of lectins and has been implicated in various biologic processes such as cell adhesion, cell recognition, proliferation, differentiation, immunomodulation, and organization of the cellular matrix, depending on its subcellular location.\textsuperscript{59,60} Cytoplasmic galectin-3 works as an antiapoptotic molecule, and increased expression of cytoplasmic galectin-3 is known to be involved in the neoplastic progression of human malignancy, including thyroid carcinomas.\textsuperscript{60}

In the latest years, the diagnostic value of the galectin-3 immunodetection in thyroid neoplasms has been widely discussed because several authors have demonstrated a correlation between galectin-3 expression and thyroid cancer in surgical specimens.\textsuperscript{17,26,29,31} Thus, the immunohistochemical analysis of galectin-3 might be an important tool for diagnostic definition. However, Cvejic et al.\textsuperscript{21} observed galectin-3 immunostaining in 5 of 14 follicular adenomas and lack of expression in 4 of 15 follicular carcinomas. These results show that although most studies have associated the galectin-3 expression and thyroid cancer, some authors reported a less significant expression in follicular adenomas, and lack of expression in follicular carcinomas.\textsuperscript{17,29,31} With regard to papillary carcinomas, galectin-3 immunohistochemistry can provide a sensitive and reliable approach in the preoperative diagnosis by FNAB,\textsuperscript{42} although most forms of papillary carcinomas and inflammations can also be recognized with tolerable certainty by their characteristic morphological and/or cytological features. The main problem is the identification of the follicular variant papillary carcinoma. The application of the galectin-3 reaction can be of help in recognizing the follicular variant of papillary carcinoma better than follicular carcinomas, since galectin-3 expression correlates better with those cases referred to as well differentiated tumors with uncertain malignant potential (WDT-UMP) having morphological features related to papillary carcinoma than to those with morphological features related to follicular tumors with uncertain malignant potential (FT-UMP).\textsuperscript{43}

On the basis of these reports, many studies have been published assessing the utility of galectin-3 in the clinical diagnosis of thyroid malignancies. However, methods and results of many of these studies vary. Even more, many of these studies were not specifically designed to test the utility of galectin-3 identification in FNAB smears in particular. Therefore, we decided to make a systematic review and a critical appraisal of the currently available evidence.
Only 6 of 37 studies originally found were considered appropriate to be included in this systematic review. The selected studies, taken together, suggested that galectin-3 is not an absolute marker for identifying malignancies on cytologically indeterminate thyroid nodules (follicular pattern), although it could be used as a useful supplementary marker. Studies not included had important methodological weakness that made their results hard to interpret—inadequate design of the studies made it impossible to prove their utility or to gauge the clinical utility of this potential diagnostic tool.

Operative characteristics values showed that galectin-3 has a sensitivity for detection of neoplasms that ranges from 75% to 100% and a specificity that ranges from 75% to 100%. The 6 selected studies showed some differences in their respective patient populations. The Saggiorato et al population does not include hyperplastic nodules, or poorly differentiated carcinomas. The authors studied 125 follicular lesions, and all of them were shown to be neoplastic. This is a very uncommon cohort since hyperplastic nodules traditionally represent about 50% of the cytologically indeterminate cases and this feature can help to enhance the specificity. On the other side, the poorly differentiated carcinoma is a well-known pitfall because of frequent negative staining for galectin-3. No poorly differentiated carcinomas were studied in the reports by Bartolazzi et al. and Kim et al.

Calculated likelihood ratio (LR) results also proved to be clinically important in the analysis of follicular thyroid lesions. Previously, a value of 10 or more for a positive LR and 0.10 or less for a negative LR were proposed as a means of identifying diagnostic tests which might serve as useful clinical tools. Results from selected studies reviewed here approached these recommended values, with the exception of the studies by Rossi et al., Collet et al., and Kim et al. All these results suggest a potential utility of galectin-3 in FNAB smears to better select patients for surgical treatment. However, to avoid the possibility of false-negative tests based only on the determination of galectin-3, it has been demonstrated that the combination of several markers (ie, galectin-3, HBME-1, and cytokeratin-19) could increase the diagnostic accuracy of the immunohistochemistry.

Nonetheless, we could also identify some flaws from primary studies that might have influenced the results. Methodological principles that guide test evaluation studies were fully developed during the twentieth century, especially after knowledge of Bayesian statistics became widely disseminated. After the publication of the users’ guides to the medical literature by the Journal of the American Medical Association, principles to be applied in critical evaluation and design of these kinds of studies were widely available. Furthermore, the Standards for Reporting of Diagnostic Accuracy Group provided guidelines for standardization of diagnostic test studies and recommended that authors follow them to guarantee a better methodological quality. Briefly, diagnostic test studies must satisfy several conditions in order to yield reliable results.

First, the study must be prospective, or at least, must be designed following a prospective directionality. Retrospective studies and inverse directionality studies (those that go from final pathology result of benign and malignant lesion back to the FNAB results) could artificially inflate certain operative characteristics. In the present systematic review, those studies that violated this condition were excluded.

Second, the study must examine a clinically relevant population to assess the operative characteristics of the test. In clinical practice, patients represent a continuum from those with evident diagnoses to those without disease. In the middle of such a continuum, there are patients with clinical uncertainty, in whom a diagnostic test will help to choose a therapeutic approach for evaluation of their thyroid nodules. Studies that assessed the utility of galectin-3 in all types of nodules could have an important bias, because those with a benign or malignant result on FNAB are located at the opposite extremes of this continuum, and depending on the number of patients included from each extreme, operative characteristic could show an artificially high or low sensitivity and specificity. Ideally, studies should include only patients with thyroid nodules with an FNAB result of follicular pattern, where uncertainty is clearly known. These studies were also excluded from this systematic review. In addition, we found an unusually high pooled prevalence of malignancy, approaching 51% (very different from the more commonly reported values of 15% to 25%). Such a sizable difference between the rates of malignancy suggests that original populations of included studies were very heterogeneous and thus implies that dissimilar inclusion criteria were employed in these series.

Third, the study must always have a gold standard. In the case of thyroid nodules with an
FNAB result of follicular pattern, the surgical specimen is that gold standard. Studies that included patients without gold standards (those that only followed galectin-3 negative nodules) could yield biased results, because nobody can with certainty define the benign nature of the nodule in the absence of a thorough pathologic examination. Moreover, in the specific case of thyroid tumors, the long preclinical stage of malignant lesion could artificially increase the specificity of the test. In this systematic review, included studies used pathologic analysis as a gold standard, but some of them had to be reanalyzed excluding those cases without that examination. Therefore, originally informed operative characteristics are not similar to those calculated and reported here.

Fourth, independent and blind evaluation of the new test and the gold standard must be guaranteed. In the case of thyroid nodules, knowing that benign or malignant diagnosis was made in final histopathology could influence the cytopathologist to find characteristics that support these histologic findings. The independent analysis will be satisfied if a prospective study design is chosen. The included studies did not report clearly the achievement of this condition, but we believe that their forward directionality designs served to decrease the effect of this potential bias.

Fifth, intraobserver and interobserver agreement should be always assessed. This condition guarantees that test application is reliable and results are not dependent on the observer. Different methods have been designed to test cytopathological staining of cells obtained by FNAB. Studies included in this systematic review are heterogeneous regarding immunohistochemistry methods, and it is known that these differences could affect operative characteristics. Collet et al. used air-dried smears instead of fixed smears. Rossi et al. used thin-layered specimens to overcome the background unspecific staining difficulties, while Inohara et al. applied Ber-Ep4 labeling to tumor cells with an immunomagnetic technique; this labeling was applied to isolated neoplastic cells facilitating staining interpretation. Saggiorato et al., Bartolazzi et al., and Kim et al. used cell-block preparations instead of smear preparations. Collet et al., Saggiorato et al., and Kim et al. used enzyme labeled polymer (biotin-free system) as a detection system, which can enhance the intensity of the staining and reduce unspecific background and oxyphilic cell staining. Antibody dilution varied from 1:100 to 1:500 among the 5 studies. Finally, in the case of galectin-3 test, a positive result (as recognized by identification of a set percentage of marked cells) can be a very subjective variable that could influence the conclusions. None of the included studies reported agreement, so it is hard to assess the reproducibility of the test.

Sixth, in cases with multiple cutoff points, a stratified analysis with each cutoff point should be tried. Tests with a continuum of possible results, or more than 2 results (as is the case of galectin-3 immunohistochemistry evaluation using grades) should be analyzed using a receiver operating characteristics (ROC) curve. Changes in operative characteristics should be reported with each selected cutoff point, because it helps to clarify the global test behavior and to allow a better selection of a cutoff point (depending on whether the researcher is seeking a more sensitive or specific test). Studies included in the current systematic review defined positive results using increased percentage of marked cells, but none of them reported ROC curve analysis. Because the selected cutoff points were different between different studies, this could affect the results because studies considered positive when more than 20% of cells showed immunoreactivity could increase specificity but decrease sensitivity in comparison with studies that designated positive results by employing a 10% threshold.

Systematic reviews have been used most often as a tool for summarizing results from studies that have low precision because of small sample size. However, another important application of these sorts of study designs is to assess the methodological quality of current studies and to identify the flaws and knowledge vacuums about a specific subject. This last function could help researchers to design better studies and to focus on relevant areas presently marked by insufficient knowledge. Results from this systematic review attempt to fulfill these expectations and thus to make recommendations for improving the design of diagnostic test studies.

In conclusion, results of this review suggest that galectin-3 may well prove to be a useful tool for refining therapeutic approaches in patients with thyroid nodules and FNAB results of “follicular pattern.” Nevertheless, the studies reviewed here were based upon differing methodologies, an observation which should prompt the wary reader to hesitate to accept uncritically all such published results. The prospect of employing improved/better standardized staining protocols and more consistent cutoff points of intensity or
frequency of positive tumor cells to routine analysis of galectin-3 on cytologic specimens remains an attractive one, and appears to be a worthy goal; however, the actual details of such a standardized methodology remain to be established. As it has proven to be the case in most other areas of application of immunohistochemistry to diagnostic surgical pathology, this may be an area in which the use of a panel of antibodies will yield the maximum quantity of diagnostic information. For example, galectin-3 might be united with other antibodies (such as cytokeratin 19 and/or CD44 variant 6) in the construction of a panel which might have sufficient power to reliably discriminate between neoplastic and non-neoplastic thyroid lesions sampled by FNAB.

REFERENCES


1054 Galectin-3 to Diagnose Malignancy in Patients with Thyroid Nodules


