

The Birmingham Formula is not a Reliable Instrument for Preoperative Prediction of Thyroid Malignancy (Prospective Study)

Ayman Mismar^{1*}, Walter Kunz², Gabriele Materazzi³, Paolo Miccoli³

Abstract

Objective: Thyroid nodules are very common in population all around the world. Preoperative anticipation of malignancy in a thyroid nodule is crucial for optimum management. There are limitations for all available tests especially in follicular tumors. The Birmingham formula was described to predict the malignancy risk in each individual with a thyroid nodule depending on phenotyping association. We tried to explore prospectively the predictivity of this formula in comparison with the postoperative pathological report.

Methods: Data for 512 patients with a thyroid nodular disease, planned for surgery, were collected 399 of which were included in the study. Malignancy risk was calculated for each patient and compared to postoperative pathology report.

Results: Statistical analysis showed a low sensitivity and a positive predictive value of around 50%, the performance of the formula did not improve when we excluded microcarcinoma.

Conclusions: Data did not support the efficacy of Birmingham formula as a reliable instrument in predicting malignancy in thyroid nodules.

Keywords: Birmingham formula, Thyroid, Malignancy, Phenotyping Association.

(*J Med J* 2014; Vol. 48 (1):40- 48)

Received

Nov. 12, 2012

Accepted

Aug. 8, 2013

Introduction

Thyroid nodular disease is a common, widespread phenomenon with a reported incidence that ranges between 19-76%¹. The importance of revealing these nodules resides in the capability to differentiate benign from malignant ones (< 5%) and to treat them in a less invasive and safer manner². Ultrasonography is the first line of imaging studies to visualize the gland; combined with color flow-doppler and elastographic analysis of thyroid nodules has yielded promising

results to preoperatively determine the nature of the thyroid nodules^{3,4}. Still it is limited by the lack of a standardized technique, inter-observer variation and small series of patients.⁵ Fine needle aspiration (FNA) is still the best single test for discriminating malignant thyroid nodules. However the diagnostic efficacy of FNA declines sharply in the diagnosis of follicular patterned lesions. Thus, the search for molecular markers (HBME1, Galectin, CK19, CITED1, Cyclin D1, HMWCK) has yielded promising results, reaching diagnostic accuracy of 95.3% with positive predictive

1- Department of General Surgery, Faculty of Medicine/University of Jordan. (Amman, Jordan).

2- Department of General Surgery, ABC Medical Centre. (Mexico City, Mexico).

3- Department of General Surgery², Faculty of Medicine/University of Pisa. (Pisa, Italy).

* Correspondence should be addressed to:

Ayman Mismar, E-mail: aymanmismar@yahoo.com

P. O. Box 13024

Department of Surgery, Faculty of Medicine, University of Jordan, Amman, Jordan 11942.

values of 100%⁴ but at higher expenses and requiring Large needle aspiration biopsies instead of fine needle aspiration. Until now, none of the preoperative investigations have satisfactorily identified the subset of patients to be electively investigated by FNA nor are able to accurately predict the presence of malignant disease in thyroid nodules.

Recently, a formula was designed in order to determine the preoperative risk of malignancy in a thyroid nodule for each individual patient; this formula considers clinical variables (age, gender, goiter type) and TSH concentration⁵. The existence of such a formula would mean: optimized treatment for those patients presenting with malignant disease; unnecessary surgical risks and stress in patients with benign disease and judicious use of human and economic resources, the importance of this formula is augmented in the subset of patients with follicular neoplasm due to the limitations of fine needle aspiration.

We set out to explore the possibility that this formula could be used within a clinical setting of a high volume center thyroid surgery center, by describing sensitivity, specificity and predictive values based not only on physical examination to describe goiter type, but on ultrasonographic findings; and utilizing only the final histology report as an endpoint to define malignant and benign disease.

Subjects and Data Recorded

The Department of Surgery at the University of Pisa performs more than 3000 thyroid procedures/year we considered a sample of 600 patients to be sufficient to mimic a 1 year practice in other centers and enough to externally validate a formula designed using 553 patients submitted to open biopsy or surgery. A 3 month period would be enough to collect 600 patients.

We included all patients submitted to

hemithyroidectomy or thyroidectomy between April 1st, and June 30th 2008; whose clinical charts clearly indicated the following variables: gender, age, thyroid ultrasound findings and final histopathology diagnosis. Other variables were also recorded like: US appearance of regional lymph nodes, thyroid volume, FT3, FT4, Tg, AbTg, AbTPO, TRAb, Calcitonin, FNAC, Frozen section, and vocal cord assessment.

The information was recorded inside the operating theater, before the intervention, by three researchers separately.

The pathology result was recorded by a single researcher blinded to the predicted possibility of malignancy in the patient.

Statistical Analysis

The risk of malignancy was calculated for each patient through the Birmingham formula: $P = 1/(1 + e^{-x})$, with e being the antilogarithmic transformation ($e=2.71828$) and x representing a calculation taking into account the patient's age and gender, the goiter type, and serum TSH. For the probability of malignancy, a numeric value for x was obtained through the following calculation: $x = -1.266 + 1.029$ (type) - 0.662 (gender) - 0.085 (age) + 0.00089 (age²) + 0.247 (TSH concentration). The goiter type was coded as 1 for diffuse or multinodular goiter and 2 for solitary nodule. The patient's gender was represented by 1 for males and 2 for females.

The final diagnostic outcome was defined as the presence or absence of malignancy. The two groups were analyzed to describe their characteristics; and the correlation between the predicted probability of malignancy expressed in a percentage against the postoperative pathology report were also compared; we also explored the subgroup: Malignant vs Benign + Occult malignancy (undiagnosed microcarcinomas measuring less than 8mm). SPSS 15 software was used.

Results

694 patients were operated during this period of time. 182 patients were operated within another facility, thus the surgical specimens were processed by a different pathology department and were excluded to reduce variables. Of the remaining 512 patients, only 417 patients had all the information required at the time of surgery. 18 of these patients' final pathology reports were untraceable. In order to maintain a scientific rigor, we accepted a 399

patient cohort and proceeded with the study.

The goiter type was defined as in the original article as: diffuse, multinodular and solitary nodules. We also included and considered dominant nodules as those presenting within a multinodular gland with nodules of less than 1cm (impalpable) and those of a diameter of at least 2 times the largest nodule described at US. For validating purposes the dominant nodule was considered a solitary nodule since the rest of the gland would appear normal during physical examination.

Table 1. Patient Demographic Characteristics

	Cohort (n=399)	Benign (n=230)	Malignant (n=169)	P<0.05	
Male	99 (24.8%)	54 (23.5%)	45 (26.6%)	NS	Chi 2
Female	300 (75.2%)	176 (76.5%)	124 (73.4%)	NS	Chi 2
Age	12-89 (48.8)	14-89 (50.5)	12-84 (46.5)	0.0046	T test
TSH	0-22.4 (1.21)	0-21.6 (1.03)	0-22.4 (1.45)	0.0285	T test
Diffuse/Nodular Goiter	197 (49.4%)	125 (54.3%)	72 (46.6%)	NS	Chi 2
Single Nodule	202 (50.6%)	105 (45.7%)	97 (53.4%)	NS	Chi 2
% predicted malignancy	10.09 (9.21- 10.97 CI)	9.23 (8.15-10.30 CI)	11.27 (9.79- 12.74 CI)	0.0243	T test

Discussion

Age and Sex

Thyroid nodules are more often found in women and the prevalence tends to increase with age: 0.05-1,8% in children with no risk factors; > 50% in adults aged over 60 years (80% women and 65% men over 80 years)². Our results agree with previous observations (median age 49 years with 95.5% confidence interval 47-51 y). In our series of patients 185 (46%) were older than 50 years of age. The observed OR for the prevalence of malignant disease in this group was 0.50 with RR of 0.51.

The prevalence of malignant disease in men was 45% with OR of 1.18 and RR of 1.09 (p>0.05).

The risk of malignancy in children is higher than in adults, with a prevalence of 18-21%⁶.

The sex distribution in children below 15 is female:male 1.5:1, while in patients aged 15-20 the ratio is 3:1⁷ and finally, in adults the ratio reaches 4:1. Sex distribution seen in our series was 3:1 female:male ratio. We found only 5 patients (4 females) younger than 18 years old, 1 of them with follicular carcinoma and the other with a papillary microcarcinoma of 4mm; one male.

Mean age at diagnosis of thyroid malignancy is 41.9 years, with metastatic involvement appearing at a more advanced age 54+ 16.9 compared to 37.7+ 12.3 years in patients with no metastatic disease⁸. Although autopsy studies demonstrated thyroid microcarcinomas occurring at the same rate in each decade in adults⁹.

In our series of patients, unlike previously reported, the mean age for malignancy diagnosis was 46.5 years whereas the benign

group was slightly older (50.5 years).

TSH

Measurement of serum TSH is the most useful test in the initial evaluation of thyroid nodules because of the high sensitivity of the TSH assay in detecting early or subtle thyroid dysfunction. The diagnostic and prognostic value of mildly elevated TSH levels has been reported previously¹⁰.

Serum TSH in 95.5% of the patients was between 0.67 and 0.87. There was a defined trend towards higher serum TSH in patients with malignant disease (Median 1.11 95.5% CI 0.823-1.25 OR 2.31 RR 1.61) while patients with benign nodules showed lower levels (Median 0.64 95.9%CI 0.54-0.717) (P=0.0285). This observation was according to previously reported association of higher TSH concentration with malignant thyroid disease.

Goiter

Neither goiter nor the nodule size is predictive of malignancy. Recent findings determined that nodule size is not predictive of malignancy and the risk for cancer and an aggressive behavior is not associated to size; tumors measuring less than 10mm have a similar prevalence of extracapsular and metastatic growth as those measuring more than 1cm¹¹. Thus, an arbitrary diameter cutoff of 10 or 15mm for cancer risk should be discouraged in clinical practice. Single versus multiple nodules.

The risk of cancer is not significantly higher for solitary nodules than for palpable or impalpable multinodular goiters.

A prospective study showed that 46% of the nodules >1cm detected by ultrasound escape detection by clinical examination¹², meaning there is a need for US screening in patients considered at higher risk for thyroid malignancy.

Our results are consistent with literature findings, we didn't observe an association between single or dominant nodules and thyroid malignancy (p>0.05).

Formula Performance

The initial report by Boelaert and colleagues, stated the limitations to their study such as the lack of histological confirmation; the different assays used for TSH determination and the clinical characterization of the goiter type by the physical examination alone.

We tried to surpass some of the limitations that are clearly drawbacks: first, the goiter type was defined based upon ultrasonographic findings and not on clinical examination and second, every patient had histological diagnosis performed by the same pathology department. TSH assays are, theoretically, standardized within the Italian health system, but we do acknowledge it could be methodology limitation to assume so.

We analyzed the overall accuracy of the formula in predicting malignant conditions of the thyroid gland, we found a statistically significant difference among the 9.25% and the 11.27% mean scores given to these groups of patients.

To determine the test performance we set a cut off value in order to help clinical decision making. We arbitrarily explore cut off values of: 10%, 25%, 50% and 75% defining sensitivity, specificity, positive and negative predictive values as well as likelihood ratios as shown in Table (2).

The specificity shown at higher percentages is remarkable but, given the low sensitivity and the positive predictive value of 50% we consider that the formula is useless in the clinical decision making process. We tried to establish a role for this formula in the setting of *clinically relevant* malignant disease as follows.

Table 2

	Cut off >10%	Cut off >25%	Cut off >50%	Cut off >75%
Sensitivity (TP)	43.8%	5.9%	00.6%	00.6%
Specificity (TN)	67%	97.7%	99.6%	99.6%
Positive Predictive Value	49.3%	58.82%	50%	50%
Negative Predictive Value	61.84%	58.37%	57.68%	57.68%
Likelihood ratio (+)	1.33	1.94	1.36	1.36
Likelihood ratio (-)	0.84	0.97	1.00	1.00

Undiagnosed microcarcinomas + Benign versus Malignancy

Thyroid microcarcinomas are considered as malignancies measuring < 10mm although some authors consider undiagnosed tumors measuring up to 15mm. In the past, this term defined cancers with local metastases without a definite presurgical diagnosis and those detected at histologic examination¹³.

Microcarcinomas have an autopsy prevalence ranging from 0.01-35.6%. Incidental prevalence ranges from 3.1-21%.

We now present the data comparing the following groups of clinical importance: benign disease plus patients diagnosed with thyroid microcarcinomas of less than 6 mm in diameter versus patients diagnosed with thyroid malignancies as shown in table (3).

Table 3

	Cohort (n=399)	Benign + Microcarcinomas (n=254)	Malignant ≥ 6mm (n=145)	P <0.05	
Male	99 (24.8%)	63 (24.8%)	36 (24.8%)	NS	Chi 2
Female	300 (75.2%)	191 (75.2%)	109 (75.2%)	NS	Chi 2
Age	12-89 (48.8)	50.1 (14-89)	46.6 (12-84)	0.0129	T test
TSH	7.7 (95.5 CI 0.67-0.87)	0.66 (96.2 CI 0.57-0.76)	1.1 (95.4 CI 0.8-1.23)	0.0670	T test
Diffuse/Nodular Goiter	197 (49.4%)	138 (54.3%)	59 (40.7%)	0.0359	Chi 2
Single Nodule	202 (50.6%)	116 (45.7%)	86 (59.3%)	NS	Chi 2
% predicted malignancy	10.09 (9.21-10.97 CI)	7.4% (96.2CI 6.72-8.46)	9.4% (95.6CI 8.48-10.46)	0.0423	T test

The size prevalence of the thyroid microcarcinomas and the current ability to diagnose smaller foci (as small as 2mm) of carcinoma pre operatively; mean a new working definition of thyroid microcarcinomas is needed. Microcarcinomas are more often papillary (65-99%) and less than 9 mm (77.7%); with the sclerosing variant more

often found in the smaller tumors (5-11.7%) suggesting defensive mechanism preventing tumor growth¹⁴. Other groups have adopted a conservative conduct toward the incidentally detected papillary microcarcinoma consisting of observation and determination of growth. They confirmed that 70% of the microcarcinomas did not change in size during

follow up and only 6.7% showed enlargement compared to baseline findings during a 5 year follow-up¹⁵.

treatment because, when compared to incidence rates for clinically apparent papillary carcinomas remains unchanged¹⁶.

Recently, authors proposed that papillary microcarcinomas < 5mm should be regarded as “normal finding” and do not require

We determined the test performance as shown in Table (4).

Table 4

	Cut off >10%	Cut off >25%	Cut off >50%	Cut off >75%
Sensitivity (TP)	44.8%	6.2%	00.7%	00.6%
Specificity (TN)	66.5%	96.9%	99.6%	99.6%
PPV	43.3%	52.94%	50%	50%
NPV	67.87%	64.39%	63.72%	63.72%
Likelihood ratio (+)	1.34	1.97	1.75	1.75
Likelihood ratio (-)	0.83	0.97	1.00	1.00

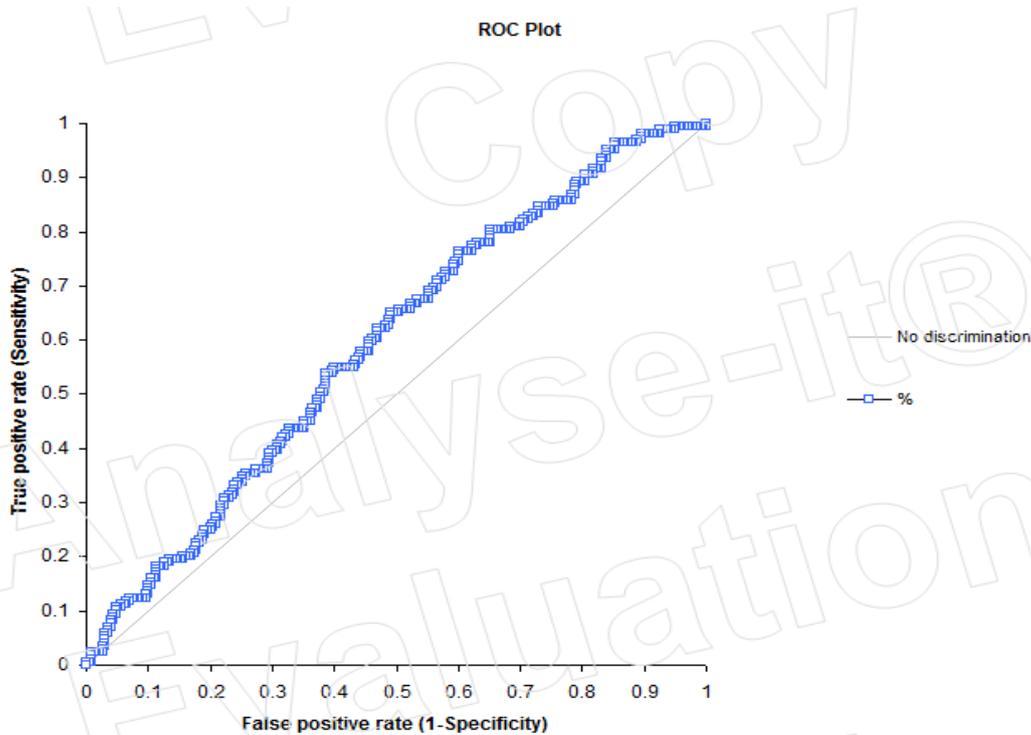


Figure 1. The ROC showed an association between higher predicted percentages of malignancy and the actual diagnosis of carcinoma (p=0.0006, 95% CI 0.54-0.65)

When plotting the ROC for the predicted percentage of malignancy we could not state that the predicted percentage of malignancy strongly correlates with the actual diagnosis, as shown in Figure (2).

characteristics, the predicted percentage of malignancy did not show a better performance than age (p=0.9407) or TSH concentration (p=0.4429) to determine whether a thyroid nodule is malignant or not. As shown in Figure (3).

Given our sample size and patient

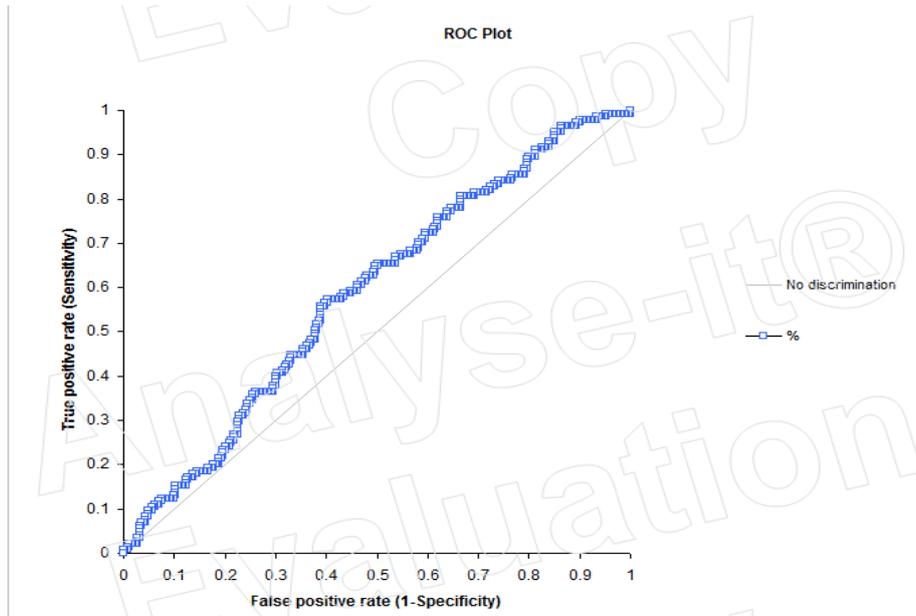


Figure 2

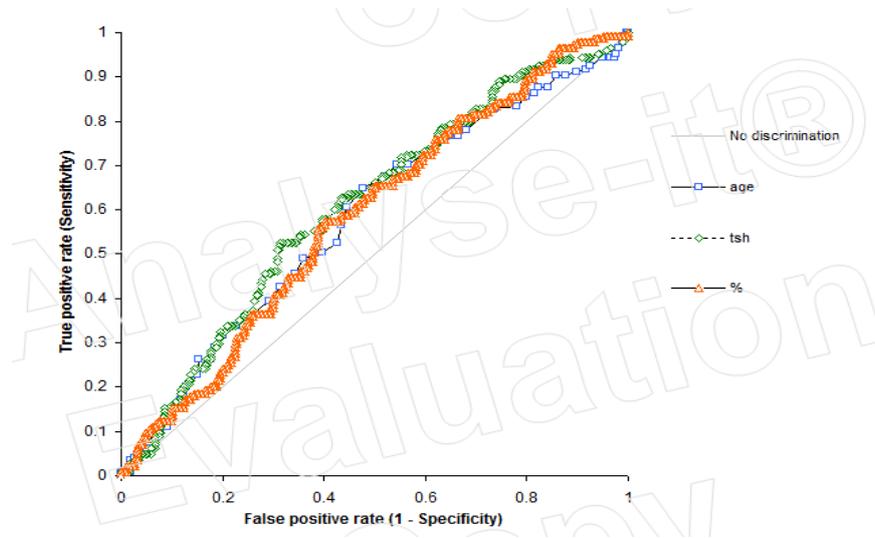


Figure 3

Conclusion

Based on our findings, we have failed to validate the formula that predicts a diagnosis of thyroid malignancy, taking into account patient’s age, gender, goiter type evaluated ultrasonographically, and the TSH concentration. We believe that, excluding serum TSH concentrations, these epigenetic

phenomena might be, in fact, associations as described by others and unrelated to the disease process. Therefore, the assumption that these phenotypic associations may predict thyroid malignancy needs to be further evaluated.

As for now, FNAC/B still remains the cornerstone for decision making in the evaluation of thyroid nodules. In the setting of

indeterminate or suspicious findings in the FNAC molecular fingerprinting of the samples or postgenomic spectroscopic analysis of the

thyroid in situ will probably determine the next step in evaluating and effectively treating thyroid lesions.

References

1. GHARIB H, PAPINE E. Thyroid nodules: clinical importance, assessment and treatment. *Endocrinol Metab Clin N Am* 2007; 36:707-35.
2. De Martino E, Pirola I, Gandossi A, Delbarba A, Cappelli C. Thyroid nodular disease: an emerging problem. *Minerva Endocrinol* 2008; 33:15-25.
3. RAGO T, VITTI P, CHIOVATO L, MAZZEO S, DE LIPERI A, MICCOLI P, VIACAVA P, BOGAZZI F, MARTINO E, PINCHERA A, Role of conventional ultrasonography and color flow-doppler sonography in predicting malignancy in "cold" thyroid nodules". *Eur J Endocrinol* 1998; 138:41-6.
4. JOO PARK, Y, HEON KWAK S, CHUI KIM D, KIM H, CHOE G, PARK DJ, JANG HC, PARK SH. Diagnostic value of Galectin 3, HBME-1, Cytokeratin 19, High molecular Weight Cytokeratin, Cyclin D1, and p27 in the differential diagnosis of thyroid nodules. *J Korean Med Sci* 2007; 22:621-8.
5. Boelaert K, Horacek R, Holder L, Watkinson J C, Sheppard MC, Flanklyn JA. Serum Thyrotropin Concentration as a Novel Predictor of Malignancy in Thyroid Nodules Investigated by Fine-Needle Aspiration. *J Clin Endocrinol Metab* 2006; 91 (11):4295-4301.
6. BENTLEY AA, GILLESPIE C, MAILS D. Evaluation and management of a solitary thyroid nodule in a child. *Otolaryngol Clin N Am*. 2003; 36:117-28.
7. NIEDZIELA M Pathogenesis, diagnosis and management of thyroid nodules in children. *Endocrine-Related Cancer* 2006; 13:427-453
8. Lin KD, Lin JD, Huang J, Huang HS, Jeng LB & Chao TC. Clinical presentations and predictive variables of thyroid microcarcinoma with distant metastasis. *International Surgery* 1997; 82: 378-381.
9. Harach HR, Franssila KO and Wasenius VM. Occult papillary carcinoma of the thyroid. A "normal" finding in Finland. A systematic autopsy study. *Cancer* 1985; 56: 531-538.
10. Hovens GC, Stokkel MP, Kievit J, Corssmit EP, Pereira AM, Romijn JA, Smit JW. Associations of serum thyrotropin concentrations with recurrence and death in differentiated thyroid cancer. *J Clin Endocrinol Metab*. 2007; 92 (7): 2610-5.
11. YEUNG MJ, SERPELL JW. Management of the Solitary Thyroid Nodule. *Oncologist* 2008; 13:105-112.
12. Brander A, Viikinkoski P, Tuuha J, Voutilainen L & Kivisaari L. Clinical Versus Ultrasound Examination of the Thyroid Gland in Common Clinical Practice. *J. Clin. Ultrasound* 1992; 20: 37-42.
13. Roti E, degli Uberti EC, Bondanelli M, Braverman LE. Thyroid Papillary Microcarcinoma. *Eur Soc Endocrinol* 2008 EJE-07-0896 ahead of print.
14. Fukunaga FH and Yatani R. Geographic pathology of occult thyroid carcinomas. *Cancer* 1975; 36: 1095-1099.
15. ITO Y, MIYAUCHI A. Prognostic Factors and Therapeutic Strategies for Differentiated Carcinomas of the Thyroid. *Endocr J* doi:10.1507/endocrj.K08E-166.
16. GHARIB H, PAPINI E, PASCHKE R. Thyroid Nodules: A Review of Current Guidelines, Practices and Prospects. *Eur Soc Endocr* 2008 EJE-08-0135 ahead of print.

عدم اعتماد معادلة برمنغهام كأداة يمكن الاعتماد عليها لتوقع وجود الورم الخبيث في الغدة الدرقية

أيمن مسمار،¹ والتر كنز،² جابريل ماتيراتزي،³ باولو ميكولي³

1- قسم الجراحة العامة، كلية الطب، الجامعة الأردنية، الأردن.

2- قسم الجراحة العامة، مركز (ABC) الطبي، المكسيك.

3- قسم الجراحة العامة، كلية الطب، جامعة بيزا، إيطاليا.

الملخص

الهدف: ان عقيدات الغدة الدرقية هي مرض شائع في أنحاء العالم جميعه، حيث إن المعالجة المثلى لها تتطلب المعرفة المسبقة لاحتمالية احتواء هذه العقيدات على ورم خبيث. وبالنظر إلى محدودية الفحوص المتاحة حالياً وبالذات في تشخيص ورم الغدة الدرقية الحربي، فقد قام فريق عمل جامعة برمنغهام بوصف معادلة للتنبؤ بخطر وجود مرض خبيث في كل فرد يعاني من العقيدات الدرقية. وفي هذه الدراسة حاولنا استكشاف دقة هذه المعادلة وفعاليتها.

المواد والطرق: جُمعت بيانات عن 512 مريضاً يعانون من مرض الغدة الدرقية عقيدية قبيل خضوعهم لعملية جراحية لاستئصال الغدة الدرقية، منهم 399 مريضاً تم تضمينهم في الدراسة، تحسباً بخطر وجود مرض خبيث لكل مريض، ومقارنتهم بتقرير الأنسجة النهائي بعد العملية الجراحية.

النتائج: التحليل الإحصائي أظهر حساسية منخفضة وقيمة تنبؤية إيجابية نحو 50٪. لتحسين أداء المعادلة عندما استبعدنا السرطانات الميكروسكوبية.

الخلاصة: النتائج لا تدعم فعالية معادلة برمنغهام كأداة يمكن الاعتماد عليها فيتوقع وجود الورم الخبيث في الغدة الدرقية.

الكلمات الدالة: معادلة برمنغهام، ورم خبيث، الغدة الدرقية.