Retrospective Analysis by Multifactor Regression in the Evaluation of the Results of Fine-needle Aspiration Biopsy of Thyroid Nodules

Askold Kucher[a], Oksana Boyko[a], Kateryna Ilkanych[a], Andriy Fechan[a,b], Natalya Shakhovska[b]

[a] *Danylo Halytsky Lviv National Medical University, 69 Pekarska str., Lviv, 79010, Ukraine*
[b] *Lviv Polytechnic National University, 12 Bandera str., Lviv, 79013,*

Abstract

A retrospective study of the results of the examination of 160 patients who underwent fine-needle aspiration biopsy of thyroid nodules has been conducted. Possibilities of using the TI-RADS (Thyroid Imaging Reporting and Data System) scale as a standardized USG thyroid examination protocol have been considered. Combinations of indicators (sex, age, ultrasound characteristics) were analyzed with the help of multifactor regression in order to recommend the thyroid fine-needle aspiration biopsy. Measures to increase the ultrasound database, which will allow further use of neural networks for their analysis, have been developed.

Keywords 1

thyroid nodules, TI-RADS, fine-needle aspiration biopsy, multifactor regression

1. Introduction

Nodal thyroid formation occurs in 30-40% of the general population and is the most common pathology among endocrine diseases. Malignant thyroid tumors account for about 3-4% of the total number of human tumors [1]. Recently, however, due to the Chernobyl accident and radioactive contamination of the territories, the number of thyroid tumors has increased markedly and the tendency to increase their frequency remains [2.3].

The use of modern ultrasonic technologies allows to detect nodular formations with a size of 0.3 cm. Fine-needle aspiration biopsy (FNAB) under ultrasonographic (USG) navigation has become widespread in the diagnosis of thyroid diseases. Cytological examination is the fastest method of preoperative diagnosis [3].

The purpose of this study was to conduct a comparative analysis of ultrasonographic (USG) examination of the thyroid gland with the results of cytological examination of biopsies of thyroid nodules based on multifactorial linear regression analysis to determine indications for thyroid FNAB.

2. Materials and Methods

We have conducted a retrospective study of the results of the examination of 160 patients who underwent FNAB of nodal thyroid formations. Data on age, sex of patients, thyroid ultrasound protocols and cytological examination of bioptats have been analyzed. FNAB procedures of thyroid nodules have been performed under ultrasound control (transduser10 MHz). The obtained samples

IDDM'2020: 3rd International Conference on Informatics & Data-Driven Medicine, November 19–21, 2020, Växjö, Sweden
EMAIL: kuchermd@gmail.com (A. 1); oxana_bojko@ukr.net (A. 2); ilkanych.katja@gmail.com (A. 3); andrii.v.fechan@lpnu.ua (A. 4); nataliya.b.shakhovska@lpnu.ua (A. 5)
ORCID: 0000-0002-7511-7057 (A. 1); 0000-0002-8810-8969 (A. 2); 0000-0002-6536-7802 (A. 3); 0000-0001-9970-5497 (A. 4); 0000-0002-6875-8534 (A. 5)

were fixed in 95% ethanol solution followed by Romanosky-Giemza staining with May-Grunwald fixation. Cytological examination of the material has been performed by a qualified cytologist.

The ultrasonographic protocol was descriptive in accordance with the TI-RADS (Thyroid Imaging Reporting and Data System) scale:

TI-RADS 1 –normal thyroid gland, absence of focal changes in the thyroid parenchyma;

TI-RADS 2 –formations with a low level of malignancy (benign nodules);

TI-RADS 3 – formations with an average level of malignancy (probably benign nodules);

TI-RADS 4A – formations with a high level of malignancy without changes in regional lymph nodes;

TI-RADS 4B – formations with a high level of malignancy with changes in regional lymph nodes;

TI-RADS 5 – formations with a pre-established malignant process in it.

The cytological conclusion was descriptive in nature, which included a category from I to VI according to the Bethesda system:

category I – nondiagnostic/ unsatisfactory bioptat;

category II – benign process;

category III – atypia of undetermined significance or follicular lesion of undetermined significance;

category IV – follicular neoplasm or suspicious for a follicular neoplasm;

category V – suspicious for malignancy;

category VI – malignant.

Determining the relationship between ultrasonographic (USG) examination of the thyroid gland and the results of cytological examination of biopsies of thyroid nodules was carried out by multivariate regression analysis, which allows to make reasoned conclusions about the development of the studied processes based on and supported by specific mathematical calculations [4].

Linear regression involves the construction of such a straight line, in which the values of the indicators lying on it will be as close as possible to the actual, then extrapolating this line can predict the result.

In the process of constructing multifactor regression models, the following stages can be distinguished:

1. Selection and analysis of all possible factors that affect the process (or indicator) being studied.
2. Mathematical and statistical analysis of factors - the main assumptions of classical regression analysis are checked.
3. Selection the type of regression multifactor model.
4. Verifying the significance of the found parameters of the model and assessing it for the adequacy of real reality.

The following indications have been analyzed: age, sex, shape, margin, composition, echogenicity, conglomerate, vascularization and number of nodules.

To conclude about the feasibility of using the found model, the analysis was conducted in the following areas:

1) The Fisher test was calculated and the found model was checked for adequacy with the source data;
2) Variance of indicators was calculated and analyzed;
3) The correlation coefficient was calculated and analyzed;
4) The coefficient of elasticity was calculated and analyzed;
5) The confidence interval for predicted measures was calculated.

Fisher's F-statistic is calculated with m and (n-m-1) degrees of freedom:

$$F = \frac{\dfrac{\sum_{i=1}^{n}\left(y_{ip} - y_{c}\right)^2}{m}}{\dfrac{\sum_{i=1}^{n}\left(y_{i} - y_{ip}\right)^2}{n-m-1}} \quad (1)$$

where m– the number of factors included into the model;

n – total number of observations;

y_{ip} – the calculated value of the dependent variable in the *i*-th observation;

y_c – the average value of the dependent variable;

y_i – the value of the dependent variable in the *i*-th observation.

According to Fisher's F-tables, the critical value Fcr with m and (n-m-1) degrees of freedom and given confidence level is found.

The procedure for finding the regression equation between different numerical sets, usually includes the following:

1. establishing the significance of the correlation between them;
2. the possibility of representing this dependence in the form of a mathematical expression (regression equation).

The second stage at creation of regression model is the choose of a formula of the regression equation. Linear multifactor regression, which describes the linear relationship between the studied data, is written as:

$$y = a_0 + a_1 x_1 + ... + a_n x_n \quad (2)$$

where y – dependent variable, function;

$a_0, \ldots, a_n$ – regression coefficients;

$x_1, \ldots, x_n$ – independent variables.

Dependent variables in linear multifactorial regression were: sex, nodule size, presence of conglomerates, shape, contours, nodule structure, presence of calcifications, echogenicity, vascularization, elastography, presence of lymph nodes, number of nodules.

To establish adequate coefficients, a logistic regression model with the same set of predictors was created, but with introducing L1-regularization (least absolute shrinkage and selection operator, LASSO) into the construction process. The essence of L1-regularization is in addition to the target regression function a fine for the complexity of the model, proportional to the norm of the coefficient vector. As a result, the coefficients were obtained: a0= 0,3699845; a1= 0,4097208; a2= -0,135977; a3= -0,5495012; a4= -0,0949706; a5= 0,154647; a6= 0,0502302; a7= 0,6084696; a8= -0,2844752; a9= 0,4817886; a10= 0,0574823; a11= 0,4017387; a12= 1,3212888.

3. Results

The results of ultrasonography (USG).

The results of USG in 160 patients have been analyzed. The material from 226 thyroid nodules was obtained by FNAB method. According to USG, 4 nodules are assigned to the TI-RADS 2 category; 145 nodules to the TI-RADS 3 category; 77 nodules to category TI-RADS 4.

The results of FNAB.

According to the results of cytological examination of 226 biopsies, 5 samples were assigned to category I, 119 to category II, 44 to category III; 24 to category IV, 15 to category V, 19 to category VI.

Comparison of USG results and cytological examination of thyroid nodules is presented in Table.1.

4. Discussion

Among 145 nodules that were assigned to TI-RADS 3 (formations with an average level of malignancy - probably benign nodules) in accordance with the TI-RADS scale, the benign process (Bethesda II) was confirmed in 88 (60.7%) cases. Given the large percentage (60.7%) of benign nodules that were suspected of being malignant on the TI-RADS scale, it can be assumed that there is a certain combination of indicators (sex, age, ultrasound characteristics) that will allow monitoring of the nodule and abandoning FNAB in first detected nodule.

Table 1

Comparison of USG results and cytological examination of thyroid nodules

Cytological conclusion	TI-RADS 2	TI-RADS 3	TI-RADS 4 (summarily 4A and 4B)
Bethesda I (unsatisfactory bioptat)	0	4 (2,7%)	1(1,3%)
Bethesda II (benign process)	3(75,0%)	88 (60,7%)	28 (36,4%)
Bethesda III (atypia of undetermined significance or follicular lesion of undetermined significance)	1 (25,0%)	29 (20,0%)	14 (18,2%)
Bethesda IV (follicular neoplasm or suspicious for a follicular neoplasm)	0	13 (8,9%)	11 (14,3%)
Bethesda V (suspicious for malignancy)	0	9 (6,2%)	6 (7,8%)
Bethesda VI (malignant)	0	2 (1,4%)	17 (22,1%)
Total	4 (100%)	145 (100%)	77 (100%)

Unequivocally, it can be argued that it is very important to find certain combinations, which is typical for nodules that correspond to the TI-RADS 3, but were classified as Bethesda III, IV, V and VI.

Among the 77 nodules that were assigned to TI-RADS 4 (formations with a high level of malignancy) on the TI-RADS scale, the benign process was confirmed in 28 (36.4%) cases. This once again confirms the importance of finding a certain combination of indicators in nodules corresponding to the cytological Bethesda II. In turn, if the nodule corresponds to the TI-RADS 4 in according with the USG, it is mandatory to perform FNAB, and not immediately perform surgery.

5. Neural Network Applications Perspectives

Nowadays, neural networks are a promising area of research in medicine, especially in the field of oncology diagnostics [5-6]. Consider the prospects for the application of neural networks in the medical field of oncology diagnosis in a preclinical study. In carcinogenesis, artificial neural networks are successfully applied to the problems of both preclinical and post-clinical diagnoses. The primary goal of medical diagnostic research is to develop more cost-effective and easy-to-use systems, procedures, and methods to support clinicians. In our case, neural networks can be used to analyze ultrasound results in order to develop diagnostic algorithms that can improve data sorting practices.

The neural network plays an important role in the decision support system. Preliminary data analysis suggests that the optimal system for our analysis would be a 5-layer model with 6 outputs and 37 inputs. However, nowadays, there is a critical lack of data for this analysis. To solve this problem, we propose to create a database of ultrasound results with remote access.

The specifics of its work is that third-party users after a preliminary verification will be able to enter their own data into the general database. To facilitate our work, we use a browser application that easily adapts to work with stationary and mobile devices. Figer 1 shows the interface view of the first user login.

The process of filling in the results of studies looks like a consistent selection of options from the drop-down windows. The first entry involves answering several narrowly professional questions to avoid clogging registration by unauthorized persons. In the future, the user is provided with a personal password and a code that will accompany his data in the database.

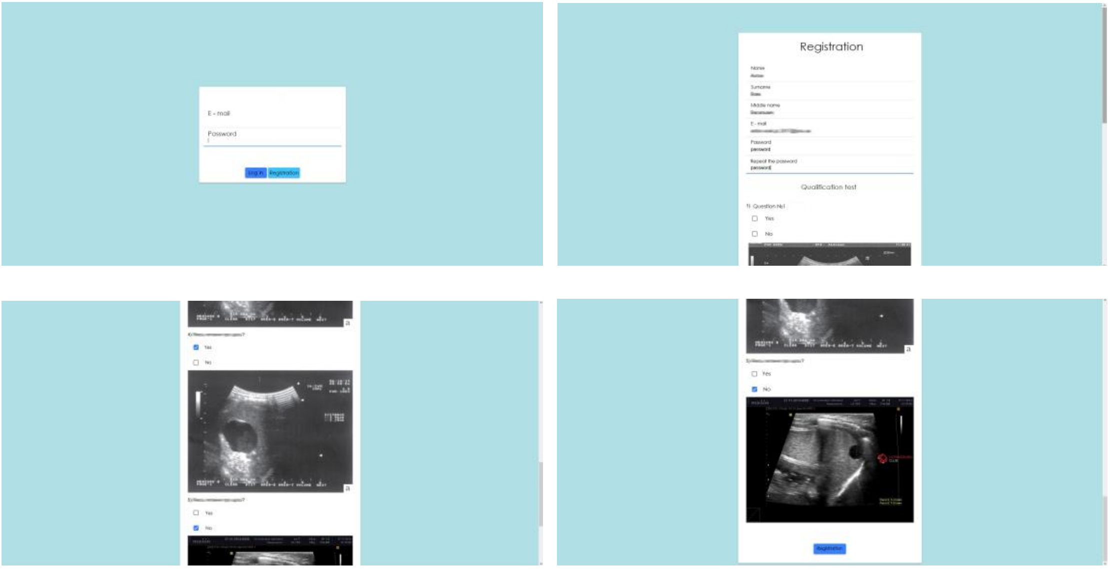

Figure 1: The interface of the first input to fill the ultrasound database.

6. Conclusions

• The use of the TI-RADS scale is recommended for mandatory use in order to standardize the USG thyroid examination protocol.

• Cytological examination of biopsies of thyroid nodules with assessment by the Bethesda system is an informative method for diagnosing cancer.

• In order to determine the indications for thyroid FNAB, it is recommended to continue searching for combinations of criteria that could clearly recommend this manipulation.

• Measures to increase the ultrasound database, which will allow further use of neural networks for their analysis have been developed.

[1] Cristina Romei, Raffaele Ciampi and Rossella Elisei "A comprehensive overview of the role of the RET proto-oncogene in thyroid carcinoma". Nature Reviews Endocrinology, 12.4, (2016): 192-202.

[2] M. Tronko, T. Bogdanova, L. Voskoboynyk, L. Zurnadzhy, V. Shpak and L. Gulak. "Radiation induced thyroid cancer: fundamental and applied aspects". Experimental oncology. 32/3 (2010): 200-204.

[3] Gabriella Pellegriti, Francesco Frasca, Concetto Regalbuto, Sebastiano Squatrito and Riccardo Vigneri. "Worldwide increasing incidence of thyroid cancer: update on epidemiology and risk factors". Journal of cancer epidemiology, 2013 (2013): 965212 https://doi.org/10.1155/2013/965212.

[4] Guiwen Rong, Xu Wang, Hui Xu and Baiqing Xiao. Multifactor Regression Analysis for Predicting Embankment Dam Breaching Parameters. Journal of Hydraulic Engineering, 146(2) (2020): 04019051

[5] Athanasios V. Vasilakos, Yu Tang and Yuanzhe Yao. "Neural networks for computer-aided diagnosis in medicine: A review". Neurocomputing. 216. (2016): 700-708. https://doi.org/10.1016/j.neucom.2016.08.039

[6] Benjamin H. Kann, Reid Thompson, Charles R. Thomas, Jr, Adam Dicker and Sanjay Aneja. Artificial intelligence in oncology: current applications and future directions. Oncology, 33(2). (2019).