

DOI: 10.21767/2472-1158.100094

Epidemiological Study of Thyroid Carcinoma Using Principal Component Analysis

MR Bricha^{1*}, EM Hamzaoui¹, Y Aboussaleh², A Mesfioui², A Soulaymani² and H Aschawa³

Abstract

In this paper, we present a new epidemiological study of thyroid carcinoma, spread over three years (2005-2008), in a sample of 399 Moroccan patients who underwent total thyroidectomy followed by metabolic radiotherapy with Iodine-131. Indeed, in addition to calculating descriptive statistics, we adopted a classification approach, based on the principal component analysis method, to classify our data. The study focused on three types of the thyroid carcinoma: papillary, follicular and undifferentiated. This method allowed us an epidemiological classification according to four criteria: age, sex, type of carcinoma and the region the subject came from. The results obtained show that papillary carcinoma remains the most dominant form among the three histological types of thyroid cancer, with a high incidence in urban coastal areas. Vesicular carcinoma is also present in these areas with a slightly lower impact. Thus, unlike other cancers, thyroid cancer can be developed in cases of a young age. 54.63% of people affected by this disease are between 20 and 45 years old. Also, this study showed that women with thyroid cancer accounted for 87.97% compared to men (12.03%). Of these, 54.13% are between the ages of 20 and 45, followed by women over the age of 45 (44.44%). While among men, we found that 48.63% of cases are older than 45 years, 47.88% are of average age (between 20 and 45 years) and 3.49% are under 19 years old.

Keywords: Descriptive statistics; Iodine-131; Metabolic radiotherapy; Principal Component Analysis; Thyroidectomy; Thyroid carcinoma

Received: March 13, 2018; **Accepted:** March 21, 2018; **Published:** March 26, 2018

Introduction

Thyroid cancer is gaining increasing interest around the world as its incidence has increased since the Chernobyl and Fukushima nuclear accidents [1]. In Morocco, although geographically distant from these areas, a rapid increase in the incidence of thyroid cancer has been clearly observed. Indeed, this increase can be explained by several factors including access to diagnostic means and the increasing performance of these resources that have identified and track subjects developing a given form of thyroid cancer. Effective cancer therapy always necessitates a sound understanding of cancer pathophysiology [2]. Some epidemiological study can give some answers: aims to investigate the causes of diseases and the factors or markers of risk that influence their occurrence in a population

Several epidemiological studies based on the use of first-order statistics have been conducted in Moroccan hospitals. Thus, some studies conducted a retrospective epidemiological study to evaluate the influence of sex, age, tumor size and histological

type [2,3]. Ainahi A. et al., Ainahi Abdelhakim et al. ont étudié un échantillon de 30 individus: 9 patients indexés atteints d'un carcinome médullaire de la thyroïde (MTC) correspondant à 3 sujets avec une évidence clinique de MEN2, 6 avec MTC apparemment sporadique (sMTC), et 21 proches ayant été étudiés pour des mutations RET [4].

This research work has the particularity of introducing a new approach to analyze and to classify our data using the principal component analysis (PCA) method [5,6]. This method is widely used to solve data reduction and classification problems as well as to help formulate hypotheses that will need to be investigated using inferential statistical models and studies [4,6]. The PCA is based on the calculation of certain statistical measures such as mean, variance and correlation [5,6]. This is why we have found it useful to couple it with a simple descriptive statistical analysis of our data [7], in order to make a contribution to the classical epidemiological approaches to thyroid cancer in our country.

- 1 DPR Laboratory, National Centre for Nuclear Energy Sciences and Techniques (CNESTEN), Rabat, Morocco
- 2 Faculty of Sciences of Kenitra, Genetics - Neuroendocrinology and Biotechnology Laboratory, Ibn Tofail University, Morocco
- 3 Department of Nuclear Medicine, IBN-Rochd de Casablanca Teaching Hospital, Casablanca, Morocco

***Corresponding author:** MR Bricha

✉ bricharabie@gmail.com

DPR Laboratory, National Centre for Nuclear Energy Sciences and Techniques (CNESTEN), Rabat, Morocco.

Citation: Bricha AMR, Hamzaoui EM, Aboussaleh Y, Mesfioui A, Soulaymani A, et al. (2018) Epidemiological Study of Thyroid Carcinoma Using Principal Component Analysis. J Clin Epigenet. Vol.4 No.1:9

Material and Methods

Data description

The study is carried out on a sample of 399 Moroccan patients, suffering from a form of thyroid cancer, and having followed metabolic radiotherapy with iodine 131, during the first 3 years of the opening, of the Medical Service. Nuclear University Hospital Ibn Rochd of Casablanca, Morocco (2005 - 2008). This event resulted in a 21% increase in the number of patients treated as shown in the graph in **Figure 1** below [8].

The data collected relate to the patient's age, sex, region of origin and the type of thyroid cancer he has developed. We are interested here in the three types of cancer whose histological classification was published by the World Health Organisation (WHO) in 2004: papillary, vesicular and undifferentiated [9,10-12].

To be able to perform the analysis by the PCA, the data must be quantitative (discrete or ordinal). For this reason, we quantified the patient's sex, region of origin and type of cancer parameters by assigning numerical values [5].

Principal component analysis

The principal component analysis (PCA) is a multidimensional statistical factorial method which allows to obtain, from a matrix of data $X \in E^L$, including quantitative variables p values for n individuals, geometric representations of these units and these variables. When the data space E is large, it is difficult to find an adequate representation to visualize the space of points. The PCA is used to find the best subspace with a reduced dimension ($L = 2$ to 3 for example), in which the cloud of the data contained in X is best represented.

The essential steps of the PCA can be summarized according to the following points [5,6]:

- Presentation of the data: the n rows of the matrix X constitute the individuals (observations) and the p columns represent the variables;
- Calculation of basic descriptive parameters: mean, variance, correlation;
- Calculation of the matrix of correlations: this matrix gives a first idea of the associations existing between the different variables. The calculation of its eigenvalues makes it possible to detect the percentages of inertia.
- Calculation of the eigenvectors of the correlation matrix: these vectors, ranked in descending order of the associated eigenvalues, make it possible to constitute the orthonormal basis of the data projection subspace.
- Principal component analysis: starting from the matrix X of the data, which is normalized so that the average of each variable is null and that its standard deviation is equal to 1, we obtain the coordinates of the projected of the individuals in the previous orthonormal basis. This allows us to represent the projected cloud of the initial cloud of weight.

Results

Analyse descriptive

The descriptive analysis of our data was based on the calculation of percentages of representation relative to the sample (**Figure 2**). Thus, a first analysis consisted of representing the distribution of the percentage of cancer patients according to their regions of origin. The **Figures 3 and 4** shows that the most affected

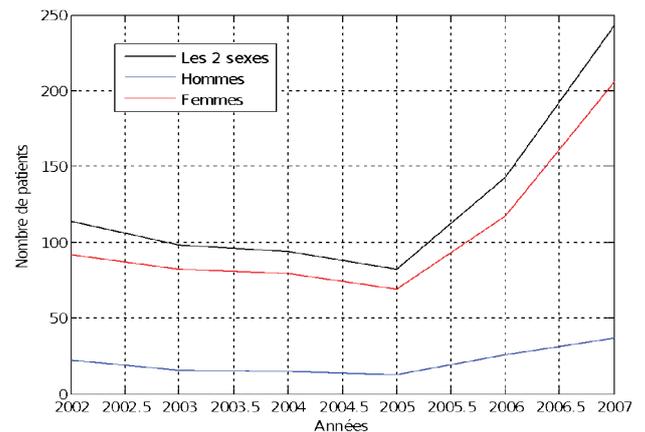


Figure 1 Evolution of the number of patients treated from 2002 to 2007 in Morocco [7].

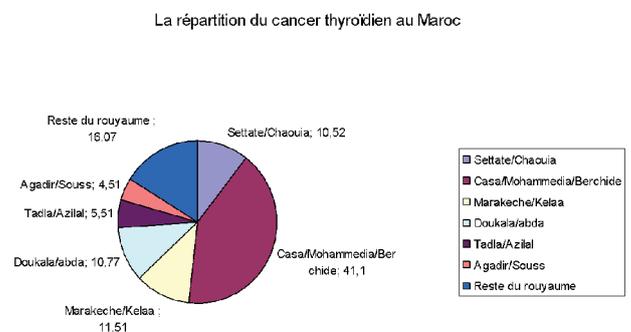


Figure 2 Geographic distribution of thyroid cancer in Morocco.

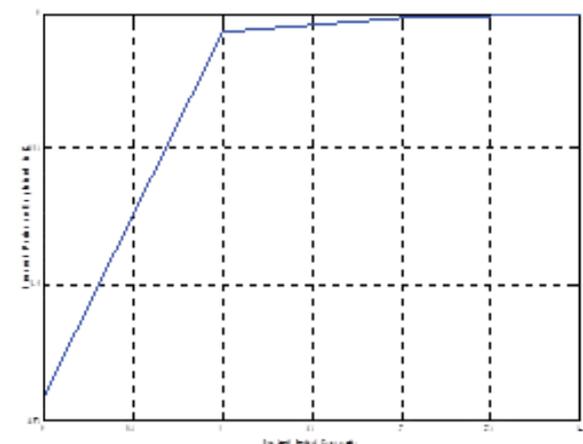


Figure 3 Evolution of the variance explained in X according to the number of principal components used by the PCA.

region is Greater Casablanca with a percentage of about 40%. Also, it should be noted that 65.71% of people from non-coastal cities develop vesicular carcinoma and 60.98% of these people from coastal cities suffer from papillary cancer. The other data collected were the subject of a descriptive statistical study, the results of which are summarized in **Table 1** [8].

Analyse par la PCA

The collected data were presented in the form of a matrix whose rows represent the 399 observations and the columns represent the variables: sex of the subject, his age, the type of carcinoma developed, and his region of origin. The latter is based on the administrative division of the kingdom that cuts Morocco into 17 different regions [12]. The choice of the dimension of the projection subspace is calculated automatically. **Figure 3** illustrates the graph of the variance explained in X as a function of the number of principal components used by the PCA that we programmed under the MATLAB™ environment [13]. The results obtained show that the minimum number of components to be used is three.

We calculated a performance index (PI) to evaluate the method used. This index is given by Equation (eq.1) below. A performance index of zero (PI = 0) indicates the best performance of the method. In practice, the smallest PI value obtained reflects the good performance.

$$PI = 1 - \frac{Y}{\hat{Y}} \quad (eq.1)$$

Where Y is the set of original observations and \hat{Y} all projected data

In our study, the best performance of the PCA corresponds to a performance index of PI = 0.1168, and the optimal orthonormal basis of the data projection subspace consists of three vectors: {"Region", "Sex of the patient", "Type of carcinoma of the thyroid"}.

The analysis of the representation of the projected in this database, allowed us to conclude that the parameter "region" is not relevant in our case, since it made it possible to classify the data according to only two classes of regions (**Figure 4**). This is also reflected by the high value of the method's performance index. For this reason, we have opted for a reorganization of the observation matrix by separating the variable "region" into two variables according to the proximity or not to the sea and according to the urban or rural character of the agglomeration.

Applying the PCA to the new observations matrix, we were able to improve the performance index to reach PI = 0.0574. In addition, we obtained three possible configurations for the orthonormal database, namely:

- Base 1: {"Urban / Rural Character", "Proximity to the Sea", "Type of Thyroid Carcinoma"};
- Base 2: {"Subject Sex", "Proximity to the Sea", "Type of Thyroid Carcinoma"};
- Basis 3: {"Subject Sex", "Urban / Rural Character", "Type of Thyroid Carcinoma"};

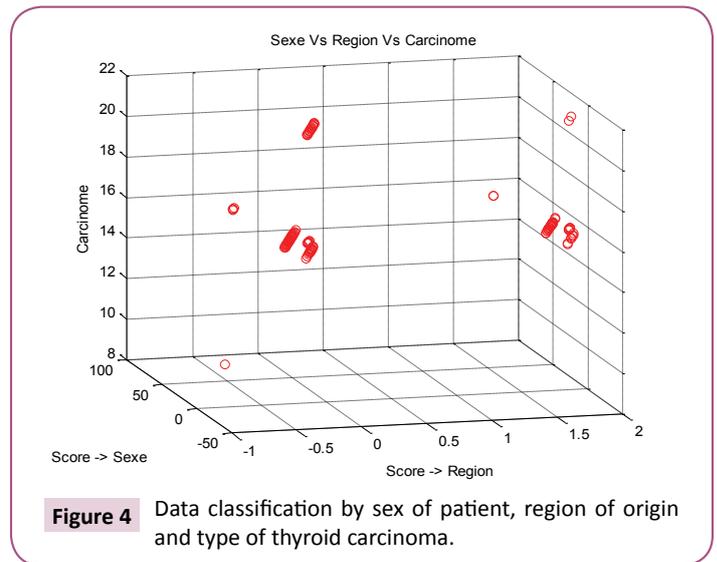


Figure 4 Data classification by sex of patient, region of origin and type of thyroid carcinoma.

Table 1 Summary Table of the Distribution of Thyroid Cancer in Morocco by Age, Sex and Histological Classification of Carcinoma Developed.

Age (Years)	CancerPapillary		CancerVesicular		CancerUndifferentiated	
	Female	Male	Female	Male	Female	Male
[0, 15]	0,25%	0,25%	0%	0%	0%	0%
[16, 19]	1%	0,25%	0%	0%	0%	0%
[20, 45]	44,86%	5,76%	2,76%	0%	0%	0%
Plus que 45 ans	33,58%	5,1%	5,26%	0,75%	0,25%	0%
% tout âge	79,69%	11,36%	8,02%	0,75%	0,25%	0%

In each of these bases, the PCA allowed us to represent the entire projected cloud as shown in **Figure 5** below.

Discussion

It appears from the analysis of the representations obtained by the PCA, that the regional administrative division made it possible to obtain a classification of the data in only two classes, whereas a division according to the proximity or not of the sea and according to the aspect urban and rural areas has led to better results. This can be explained by the large area of administrative areas that are inhabited by heterogeneous populations from the point of view of crops and diets. The second geographic approach is based on the results of a recent study on Moroccan household consumption patterns, which states that urban dwellers consume twice as much seafood (fish, crustaceans, mollusks) than those living in rural areas and this consumption is accentuated in coastal areas.

This may explain the results obtained by the application of PCA that show that papillary carcinoma remains the most dominant form among the three histological types studied, with a high incidence in urban coastal areas (**Figure 5a**). Vesicular carcinoma is also present in these areas with a slightly lower impact. In addition, (**Figures 5b and 5c**) show that women are more likely to develop all three types of thyroid cancer than men, especially in

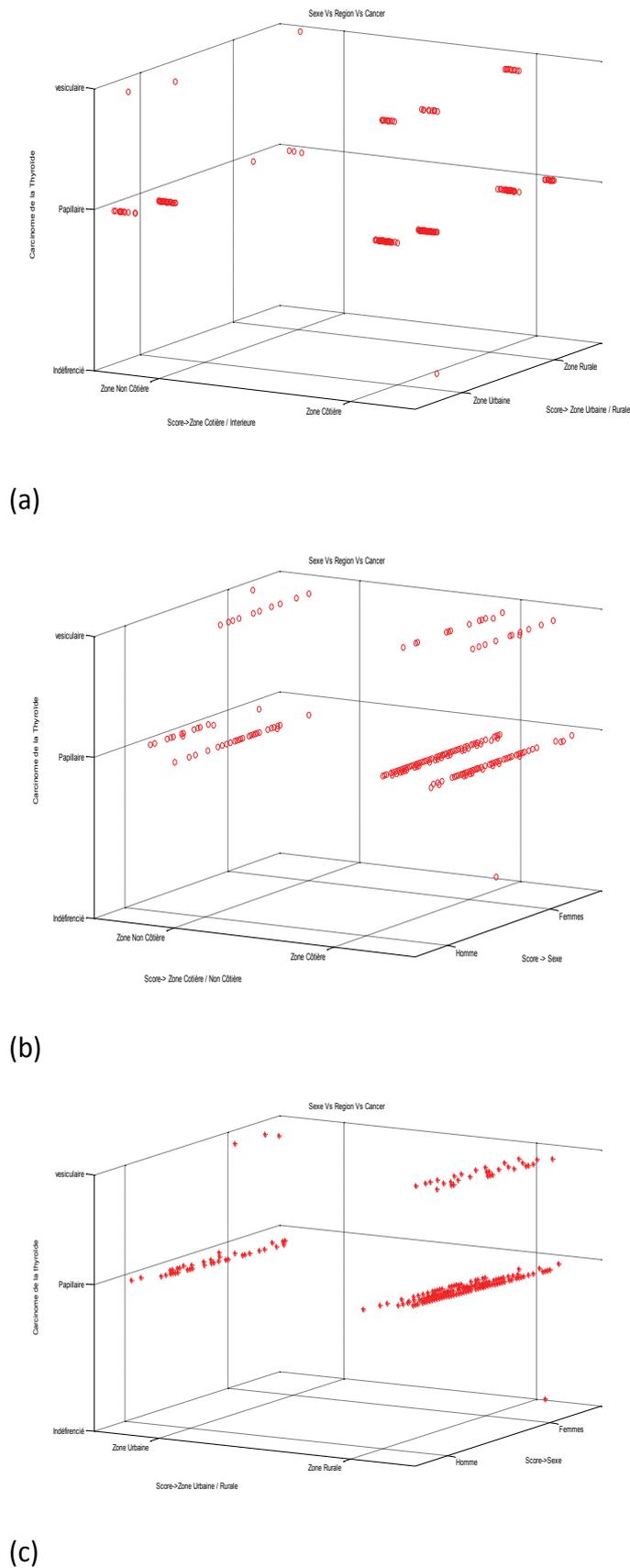


Figure 5 Classification of data by the PCA. (a) In the base 1. (b) In the base 2. (c) In the base 3.

coastal urban areas. These graphical results made it possible to confirm the descriptive statistics of our sample. Thus, unlike other cancers, thyroid cancer can be developed in cases of a young age. 54.63% of people affected by this disease are between 20 and 45 years old. Also, this study showed that women with thyroid cancer accounted for 87.97% compared to men (12.03%). Of these, 54.13% are between the ages of 20 and 45, followed by women over the age of 45 (44.44%), and teenage girls make up only 1.14%. While among men, we found that 48.63% of cases are older than 45 years, 47.88% are of average age (between 20 and 45 years) and 3.49% are very young age (under 19 years).

In terms of cancer histology, it was revealed that in both sexes papillary carcinoma is the most dominant form (90.98%) of which 93.75% in men and 79.69% in women. The vesicular form is less frequent (8.52%). It occurs much more in women (91.18%), while men developing this carcinoma are older than 57 years and represent 8.82% of the sample studied. Undifferentiated carcinoma was not recorded in any male subject during these

3 years of study. In contrast, only one woman, aged 45, was identified during this study period.

Conclusion

In this study, we combined principal component analysis (PCA) and descriptive statistics to analyze data on thyroid cancer developed by Moroccan patients. The results of the PCA made it possible to classify the data by projection on bases composed of three characteristic vectors. These graphic representations showed that women are the most affected by the three types of cancer studied and that the high incidence is recorded in urban and coastal areas. This allowed us to assume that there is no direct relationship between the consumption of iodine-rich foods and the risk of developing thyroid cancer. Also, she showed that the majority of patients in this sample develop a well-differentiated thyroid cancer without good prognosis metastasis and well-coded treatment.

References

- 1 Suzuki K, Mitsutake N, Saenko V, Yamashita S (2015) Radiation signatures in childhood thyroid cancers after the Chernobyl accident: possible roles of radiation in carcinogenesis. *Cancer Sci* 106: 127-133.
- 2 Wakaskar RR (2017) Passive and Active Targeting in Tumor Microenvironment. *Int J Drug Dev & Res* 9: 37-41.
- 3 Bouchbika Z, Haddad H, Benchakroun N, Kotbi S, Megrini A, et al. (2014) Cancer incidence in Morocco: report from Casablanca registry 2005-2007. *Pan Afr Med J* 16: 31.
- 4 Abdelhakim A, Barlier A, Kebbou M, Benabdeljalil N, Timinouni M, et al. (2009) RET genetic screening in patients with medullary thyroid cancer: the Moroccan experience *J Cancer Res Ther* 5: 198-202.
- 5 Rachid A (2011) Carcinome Papillaire de la Thyroïde (A Propos de 40 cas). Thèse de doctorat en Médecine, Université Sidi Mohammed Ben Abdellah, Faculté de Médecine et de Pharmacie, Fès, Février 2011.
- 6 <http://www.springer.com/us/book/9780387954424>
- 7 P Sanguansat (2012) Principal Component Analysis Multidisciplinary Applications. Ed. InTech.
- 8 Bricha MR (2008) Contribution à l'étude épidémiologique du cancer de la thyroïde au Maroc. Mémoire de Master. Université Ibn Tofail, faculté des Sciences de Kenitra.
- 9 Hedinger C (1988) Histological typing of thyroid tumours. (WHO Classification).
- 10 Teijeiro JC, Sobrinho-Simoes M (2003) Carcinoma papilar de la glándula tiroides Problemas en el diagnóstico y controversias. *Revista española de Patología* 36: 373-382.
- 11 Leenhardt L, Ménégau F, Franc B, Hoang C, Salem (2005) Cancers de la thyroïde. *EMC-Endocrinologie*, 2 : 1-38.The Mathworks Inc.
- 12 <https://www.hcp.ma/file/103035/>
- 13 <https://cabinetbassamat.com/codes-et-lois/id/46587>