

Study of a breast cancer epidemiology among the patients: Case of Tiaret region, Algeria.

Bacha Soufiane¹, ADIM Hakim¹, Mostefa Belarbi¹, Chahrazed ADDA¹, BENAZZOUZ Embarka Ikram², MIMOUNI Amina, CHOUANA Hadjer ²,

¹Faculty of Mathematics and informatics, University of Tiaret, LIM Laboratory , Tiaret Algeria

²Faculty of Material Sciences, Department of Physics, Synthesis and Catalysis Laboratory, University of Tiaret,

gl.soufianebacha@gmail.com

hakim@univ-tiaret.dz

mbelarbi@univ-tiaret.dz

chahrazedadda@gmail.com

embarkaikram.benazzouz@univ-tiaret.dz

amina.mimouni1@univ-tiaret.dz

hajer.chouana@univ-tiaret.dz

Abstract—Cancer poses the highest clinical, social and economic among all human diseases. Breast cancer is one of the most dangerous and frequently occurring cancers among women. Breast cancer starts when cells in the breasts grow out of control and form a tumor. According to the American Cancer Society, factors that affect one's chances of developing breast cancer include age, lifestyle (e.g., alcohol consumption), and family history. Factors affecting the mortality of breast cancer patients include education level and race [18]. Mar et al. [19] assessed breast cancer rates based on age and other risk factors, such as family history and genetic factors. [20] also assess mortality trends in breast cancer by age. The purpose of this study is to assess the effect of marital status on breast cancer specific mortality in women and evaluate the role of race in this association. Furthermore, The goal of this study was to determine the incidence rate and some risk factors of breast cancer in women with various sub-types and stages, to identify subgroup(s) that might potentially benefit from breast cancer. Due to this, we propose here an analytical method and a tool for periodic and long-term measurement of cancer behaviors and age, race, and marital status parameters. These data will then be analyzed using computer tools (SPSS) to extract knowledge and eventually develop predictive cancer models. The Tiaret region will be a field for data collection and validation of the models or tools developed.

Index Terms— Breast cancer, statistical analytical method, Arcgis, SPSS, cancer registry, incidence, epidemiology.

I. INTRODUCTION

Cancer is a major public health problem worldwide [1]. The abnormal growth of human cells is widely known as cancer that attacks healthy cells [2]. Cancer cells have the ability to spread to other parts of the body through the blood and lymphatic systems. It is the leading cause of death among middle-aged and older women [3]. However, systematic data collection in cancer registry over the last 40 years has facilitated the assessment and control of the disease [4].

Cancer registries have been widely used in epidemiological research. They are an essential component of a fully developed cancer control program. In addition to providing information on current and future needs for services, they are used to monitor programs of prevention, early detection, and cure [5]. Breast cancer (BC) has become the most common cancer in women worldwide in both developed and developing [6] [7], and female breast cancer is among the cancers of Arab countries (Algeria), with the maximum incidence among women of the world. The concepts of the etiology of breast cancer are changing rapidly and the variability in both incidence and survival among breast cancer sub-types so, physicians need to be aware of these changes [8] and the risk factors that are associated with the development of breast [9]. Tumor characteristics such as tumor grade and age as well as social factors including race and marital status have been found to be associated with both breast cancer incidence and survival. The main objective of this paper is to evaluate some risk factors for breast cancer mentioned, in this case, age and marital status, and geographical factor, as well as their impact on the spread of breast cancer, which will enable us to propose well thoughtout solutions, in line with the medical diagnosis context. For this, we propose an analytical method and a tool for periodic and long-term measurement of some breast cancer factors. These data will then be analyzed using computer tools (SPSS) to extract knowledge and eventually develop medical models. The Tiaret(Algeria) region will be a field for data collection and validation of the models or tools developed.

II. SOURCE OF DATA

Cancer is one of the new priority public health needs today. To fight effectively against this disease, It is therefore essential to have precise data by setting up registers. The availability of cancer data is a key element in setting up a program to fight this disease. These data provide reliable information on the cancer profile, including incidence, trend, and survival rates. In recent years, WHO and several other research groups have recognized the importance of

population cancer registries as an essential tool for health research. An articulated thanks to Order No. 22 of February 18, 2014, allowed to set up of a National Network of Cancer Registries, with national coordination of the East, Center, and West Networks.

Description of cancer registry networks

The national registry network was created in 2015 as part of the 2015-2019 cancer plan [22], in strategic axis number 6 relating to the development of the cancer information and communication system. The implementation of this national network is reinforced by order N 22 of February 18, 2014. This order institutionalizes population registers, with the consolidation of existing registers and the establishment of new existing registers in all cities of the country. this institutionalization is articulated in the form of regional coordination of East, Center, and West [21] to allow a wide coverage of cancer registration in Algeria. The cities of the country are divided into three regional networks East and South-East, Center, and South Center, and West and South-West, coordinated by the National Network of Cancer Registries [21].

- **East and Southeast Network** coordinated by the Setif Cancer Registry and includes 20 cities: Setif, Annaba, Bejaia, Taref, Batna, Khenchla, Skikda, Souk-Ahras, Biskra, Constantine, Om El Bouaghi, El Oued, Guelma, BBA, Tebessa, Jijel, Msila, Mila, Illizi, Ouargla.
- **Central and South-central network** The Center network is coordinated by the INSP and includes 13 cities: Algiers, Blida, Medea, Tipaza, Ghardaïa, Tizi-Ouzou, Djelfa, Tamanrasset, Ain Defla, Boumerdes, Bouira, Laghouat, Chlef. and south-center
- **West and South-West network** coordinated by the Oran Cancer Registry includes 15 cities: Oran, Mostaganem, Tlemcen, Adrar, Relizane, Tiaret, Mascara, Tissemsilt, Sidi-Bel-Abb'es, Tindouf, Naama, Saida, Bechar, Ain Temouchent, El Bayed.

Cancer data collection begins by identifying people with cancer who have been diagnosed or received cancer care in hospitals, outpatient clinics, radiology departments, doctors' offices, laboratories, surgical centers, or from other providers (such as pharmacists) who diagnose or treat cancer patients. Table I shows many different types of data tacked from the cancer registry of Tiaret, including patient features.

TABLE I
TABLE FOR BREAST CANCER REGISTRY OF TIARET

Variables recorded by cancer registries
Personal identification (names (in full))
Sex
Date /place of birth
Age
Job
Address (usual residence)
Topography (site) of primary cancer
Date of last contact
Status at last contact (at least dead or alive)

This was a retrospective study based on the secondary data collected through patient recorders from the website with strengthened projection and similarity in the content with

Tiaret cancer register to avoid human rights privacy by the collaboration with Hospital Public Establishment of Tiaret, oncology and epidemiology services as base references in our study. The population size is 4005 patients. Table II shows breast cancer data-set attributes:

TABLE II

TABLE FOR BREAST CANCER DATA-SET ATTRIBUTES

Patient ID	Age	Race	Marital status	Grade	Survival Months	Status
Grade						

III. EXPLANATORY VARIABLES

All explanatory variables included in this study are those that were obtained from the data set during the time of the study by the collaboration with the Hospital Public Establishment. The description of variables in the data set is given as **Patient ID**: identification codes used to identify the patient (replace patient name). **Age**: This variable provides the patient's age at diagnosis. **Grade**: The grade of a tumor describes how abnormal the tumor cell and tissue look under a microscope. It indicates how quickly a tumor can grow and spread. The tumor is well-differentiated as if the tumor cells and the organization of the tumor's tissue are close to those of normal cells. These tumors tend to grow and spread at a slower rate. The undifferentiated or poorly differentiated tumors have abnormal-looking cells and may lack normal tissue structures. In our data set, we have used the value "1" for Grade I (well-differentiated or low grade), the value "2" for Grade II (moderately differentiated or intermediate grade), the value "3" for Grade III (Poorly differentiated or high grade). **Marital status**: In the data set, we have used the value "2" for divorced, "1" for married, "0" for single, "3" for separated and "4" for widowed. **Race**: In the data set, we have used the value "0" for white, "1" for black and "2" for others. **Status**: In the data set, we have used the value "0" for Dead, "1" for Alive.

IV. SPECIFIC RISK FACTORS

A. Age

Age is among the most important risk factors for female breast cancer patients in the Tiaret region, Algeria. The incidence rate of female breast cancer patients has an agedependent curve, where the risk increases with increasing age.

B. Environmental factors

The main environmental factors involved in breast cancer risk are ionizing radiation, which is a known carcinogen, and night/shift work with disturbances of circadian rhythms, probably classified as carcinogenic [12]. Daily exposure to solar radiation the population of the Tiaret region especially during the summer months of every year works in an irradiated environment for the woman. All these radiation causes, whereas among that we have breast cancer.

C. Geographical origin of women

The geographical origin of women also appears to be a factor to be taken into consideration. Li ET all., Showed

that the majority of Asian studies of Chinese populations had different results from Caucasian studies in the United States and Scandinavia in the year 2015 [13][14]. Indeed, the study of breast cancer in the female population of this region during the year 2016 showed that patients are geographically very far from the mountains and the Mediterranean Sea. There are two important areas in the Tiaret region in Algeria.

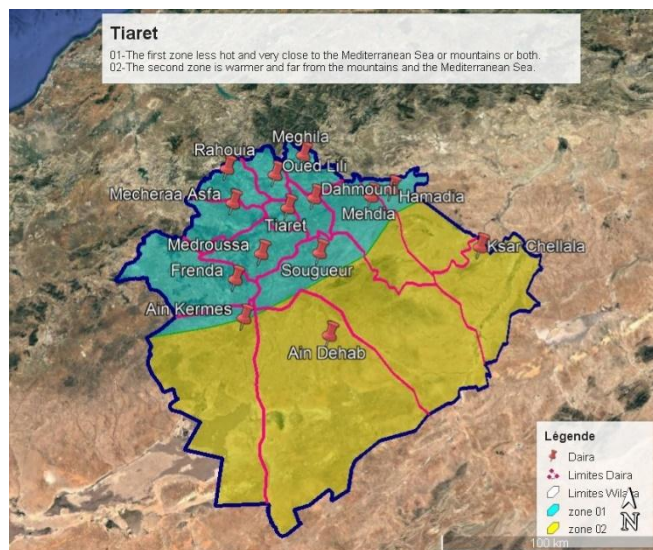


Fig. 1. Google earth image represents the two main regions in Tiaret.

In our study and to determine the relationship between the case study and the geographical data we used a geographic information system (GIS) which is a type of database containing geographical data that is, descriptions of phenomena for which location is relevant, combined with software tools for managing, analyzing, and visualizing those data, in our case this system is ArcGIS.

The previous figures represent the first zone less hot and very close to the Mediterranean Sea or mountains or both, the relative risk of female breast cancer and includes Dahmouni, Frenda, Rahouia...etc. The second zone is warmer and far from the mountains and the Mediterranean Sea and includes Ain Dehab and Ksar Chellala. So, geographical origin plays a very



Fig. 2. ArcGIS image represents the two main regions in Tiaret.

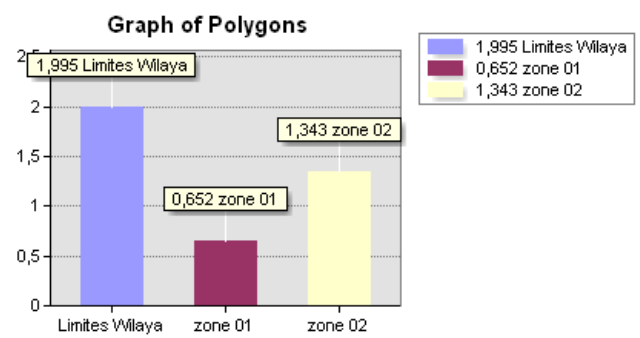


Fig. 3. Aria comparison in the two main regions in Tiaret.

important role in reducing breast cancer. This study showed higher breast cancer in warmer regions.

ArcGIS provides the capability to relate previously unrelated information, through the use of location as the "key index variable". Locations and extents that are found in the Earth's space-time are able to be recorded through the date and time of occurrence, along with x, y, and z coordinates; representing, longitude (x), latitude (y), and elevation (z). All Earth-based, spatial, temporal, location, and extent references should be relatable to one another, and ultimately, to a "real" physical location or extent. This key characteristic of ArcGIS has begun to open new avenues of scientific inquiry and studies.

This work offers a geographical perspective so that the analyses that are being carried out with ArcGIS and spatial statistical tools can give us a spatiotemporal analysis of disease mapping, the latest health and social geography, environmental variables, data mining, and web-based mapping have been the most efficient approaches for cancer studies, therefor ArcGIS and SPSS could be valuable tools in medical decision making.

TABLE III
ARCGIS DATABASE.

	Daira	Number of towns	towns	Area (km2)	Daira limits (km)	Population (hab)
01	Ain Deheb	3	A'inDeheb •Chehaima • Naima	5528,47	1287.565	46082
02	Ain Kermes	5	A'in Kermes • Madna • Medrissa • Djebilet Rosfa	3073,5	1116.479	49307
03	Dahmouni	2	• Sidi Dahmouni • Ain Bouchekif	315,68	576.946	35402
04	Frenda	3	Frenda • Ain El Hadid •Takhemaret	1640,97	853.258	95088
05	Hamadia	3	Hamadia • Bougara • Recha'iga	1508,2	470.758	89509
06	Ksar Chellala	3	Ksar Chellala • Serghine • Zmalet El Emir Abdelkader	1433,18	272.405	46505
07	Mahdia	4	Mahdia • Ain Zarit • Nadorah • Seba'ine	1309,5	187.266	57884
08	Mechraa Safa	3	Mechraa Safa • Djillali Ben Amar • Tagdemt	600,54	441.909	26398
09	Medroussa	3	Medroussa • Sidi Bakhti • Mellakou	635,93	345.092	31537
10	Meghila	3	Meghila • Sebt • Sidi Hosni	411,61	1077.773	12329
11	Oued Lilli	3	Oued Lilli • Sidi Ali Mellal • Tidda	470,34	428.891	23332
12	Rahouia	2	Rahouia • Guertoufa	270,61	942.037	24657
13	Sougueur	4	Sougueur • Faidja • Si Abdelghani • Tousnina	3363,02	1078.426	107530
14	Tiaret	1	Tiaret	111,45	181.409	201263
Zone 01	10		-	7239	5097.774	597399
Zone 02	4		-	13434	6242.538	249424
Total	14		-	20673	9260.214	846823

V. ANALYTICAL APPROACH

We used exploratory research methods and analytical tools to process clinical data extracted, with particular emphasis on cancer in the Tiaret region. the main objective of this study is to process the clinical data of the breast cancer of the Tiaret region in order to provide a reliable decision for the complex questions of the spread of breast cancer and determine some risk factors.

A. Hypothesis testing statement

For exploring the relationship between patient survival and marital status factor, we conducted the Kaplan–Meier to test the hypothesis(assumption): marital status influences on breast cancer survival rate. On another hand does marital status consider to be a risk factor for breast cancer ?

B. Statistical analysis

Kaplan–Meier survival analysis and the Log-Rank test were used to compare unadjusted survival rates by marital status. Secondly, Pearson's chi-squared test was used to test the association between the race and status of breast cancer. The race with the status of breast cancer (dead/alive) interaction was tested to determine if the breast cancer status was associated with the Patient's race. Furthermore, to accept or reject an existing relationship, we formulate two hypotheses:

- **Null Hypothesis (H0):** There is no significant relationship between the marital status and breast cancer survival rate.
- **Null Hypothesis (H0):** There is no significant relationship between the geographical location factor and breast cancer .
- **Null Hypothesis (H0):** There is no significant relationship between the patient's race(black /white) and breast cancer .
- **Alternative Hypothesis (H1):** There is a relationship between the marital status and breast cancer survival rate.
- **Alternative Hypothesis (H1) :** There is a relationship between geographical location factor and breast cancer .
- **Alternative Hypothesis (H1) :** There is a relationship between the patient's race and breast cancer .
- All analyses were performed using IBM SPSS [15].

VI. RESULTS AND DISCUSSION

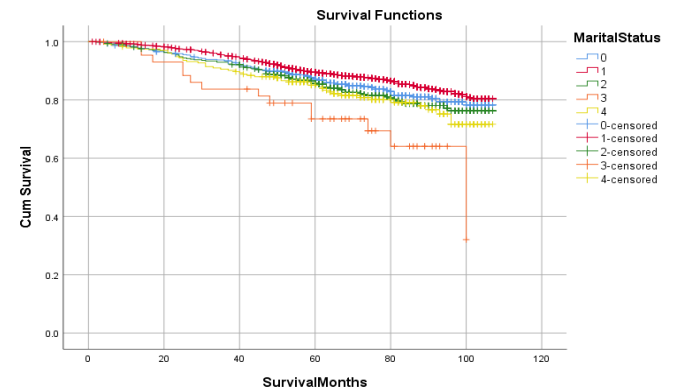


Fig. 4. Kaplan-Meier curve.

Fig. 4. Shows the Kaplan-Meier breast cancer-specific survival graphs according to marital status for white and black women.

Fig. 4 Shows the results of the Kaplan–Meier survival curves for marital analysis and the Log-Rank test was used to compute the risk of mortality for women who were single, separated, divorced, and widowed when compared with women who were married In this study, we found that in all cases, patients who were classified as widowed, separated,

divorced, single were likely to die from breast cancer than their married counterparts. These findings are similar to those reported by Aizer et al. [16], who found that divorcees were about 21% more likely to die of breast cancer. A further analysis by Martinez et al. [17] similarly revealed that unmarried and widowed patients were 28% and 35% more at risk of death, respectively, compared to married ones. Their analysis was stratified according to cancer stage and is in accordance with our findings. Our results concur with studies that found that married women with breast cancer fare better than unmarried women.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	84.050 ^a	39	.000
Likelihood Ratio	79.641	39	.000
Linear-by-Linear Association	12.315	1	.000
N of Valid Cases	4005		

a. 9 cells (11.3%) have expected count less than 5. The minimum expected count is .76.

Fig. 5. Shows the table of significance and the relationship between race and the status of breast cancer.

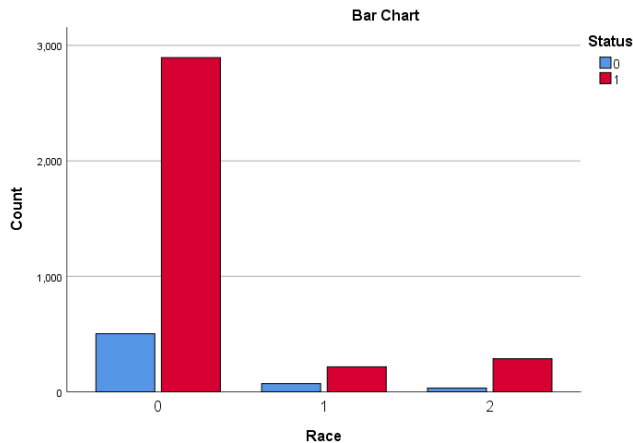


Fig. 6. Shows the Pearson's chi-squared test statistic for race and status variables.

Fig. 5 Presents the table results for testing Pearson's chisquared test statistic for race and status variables. We found a statistically significant difference across the race after applying Pearson's chi-squared test statistic. The p-value of the chisquared test statistic was $<.05$, which means there is a correlation between the race and status of breast cancer.

Fig. 6 Shows the Pearson's chi-squared graphs according to breast cancer status (dead:0/alive:1) for white (0), black (1), and other (2) women. The chart indicates that black women had a higher risk of mortality than white women(white women had a higher alive). These findings are similar to those reported in [24] [25], which found that black women have the highest breast cancer death rate than white.

VII. CONCLUSION AND PERSPECTIVE

Cancer has become a major public health concern in Algeria. Prevention, early diagnosis, and different care and treatment play an important role in reducing this chronic

disease in this region, and why not over the world, there are many factors that affect the spread of breast cancer such as marital status, age, and race. We recommend an intensive health education program for men and women that targets the consequences of divorce, especially considering the increasing divorce rate in Tiaret –Algeria. Furthermore, we plan to apply more tests on our database and determine an effective method using artificial intelligence and deep learning for cancer study.

VIII. REFERENCES

- [1] J. S´aez, J. Luengo, J. Stefanowski, F. H.-I. Sciences, and undefined 2015, "SMOTE-IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering," Elsevier, Accessed: Jul. 04, 2022.
- [2] V. Chaurasia and S. Pal, "Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer," SN Comput. Sci., vol. 1, no. 5, Sep. 2020, doi: 10.1007/S42979-020-00296-8.
- [3] A. Qasem, S. Abdullah, . . . S. S.-P. J. S., and undefined 2017, "An accurate rejection model for false positive reduction of mass localization in the mammogram," pertanika.upm.edu.my, Accessed: Jul. 04, 2022.
- [4] Parkin, D. M., L`a`ar`a, E., Muir, C. S. (1988). Estimates of the worldwide frequency of sixteen major cancers in 1980. International journal of cancer, 41(2), 184-197.
- [5] Parkin DM. The evolution of the population-based cancer registry. Nat Rev Cancer 2006;6:603-12.
- [6] Nadine Zielonke and al. Evidence for reducing cancer-specific mortality due to screening for breast cancer in Europe: A systematic review. European Journal of Cancer 127 (2020) 191e206.
- [7] Jing-Yi and al.The Undervalued Effects of Polychlorinated Biphenyl Exposure on Breast Cancer. Clinical Breast Cancer, Vol. 20, No. 1, (2019) 12–18.
- [8] M.J.M. BROEDERS and A.L.M. VERBEEK. Breast Cancer Epidemiology and Risk Factors. Springer-Verlag Berlin • Heidelberg 2000.
- [9] Soley Bayraktar and Banu K. Arun. Epidemiology, Risk Factors, and Prevention. Springer International Publishing Switzerland 2016.
- [10] <https://www.ibm.com/fr-fr/analytics/spss-statistics-software>
- [11] <https://www.arcgis.com/index.html>.
- [12] Centre L´eon BERARD. Cancer du sein. Cancer environnement comprendre informer pr´evenir, juin (2017).
- [13] Vieira, V., Webster, T., Weinberg, J., Aschengrau, A., & Ozonoff, D. (2005). Spatial analysis of lung, colorectal, and breast cancer on Cape Cod: an application of generalized additive models to case-control data. Environmental Health, 4(1), 1-18.
- [14] J.C. Benabu, F. Stoll, M. Gonzalez, C. Mathelin. Night work, shift work: Breast cancer risk factor?. Gyn´ecologie Obst´etrique and Fertilit´e 43 (2015) 791–799.
- [15] IBM SPSS Statistics for Windows. vol 21.0. IBM Corp. Armonk NY. 2012.

[16] Aizer AA, Chen MH, McCarthy EP, Mendu ML, Koo S, Wilhite TJ, et al. Marital status and survival in patients with cancer. *Journal of clinical oncology*. 2013;31(31):3869. pmid:24062405

[17] Mart'inez ME, Unkart JT, Tao L, Kroenke

[18] Bartonville. Reasons Behind Racial Differences in Breast Cancer Mortality. Women in Academic Report, Bru Con Publishing Company; 2018.

[19] Mar SJ C, Wilson C. Breast cancer screening in British Columbia: A guide to discussion with patients. *BRITISH COLUMBIA MEDICAL JOURNAL*. 2018;60:20–26.

[20] Hollecze B, A V, S C, B H. Trends in breast cancer survival in Germany from 1976 to 2008 period analysis by age and stage. *The International Journal of Cancer Epidemiology, Detection, and Prevention*. 2011;35:339–406.

[21] Arrê'té n 98 du 27 septembre 2015 fixant le Réseau national des registres de cancer. Ministère de la Santé et de la réforme Hospitalière

[22] Plan National Cancer 2015-2019, nouvelle vision stratégique centrée sur la maladie, Octobre 2014. Available at.

[23] Boukabcha2021, Basic study of a breast cancer epidemiology among the female patients using a regional data source: Case of Chlef region, Algeria,2021.

[24] DeSantis, C. E., Ma, J., Gaudet, M. M., Newman, L. A., Miller, K. D., Goding Sauer, A., ... Siegel, R. L. (2019). Breast cancer statistics, 2019.CA: a cancer journal for clinicians, 69(6), 438-451.

[25] <https://www.medicalnewstoday.com/articles/breast-cancer-death-rate-inblack-women-does-dna-repair-play-a-role> Study-implications