

# Estimating the 5-Year Breast Cancer Prognosis in a Cohort Residentially Exposed to Hazards at Hermosillo, Sonora, Mexico

DE Villa-Guillen<sup>1\*</sup> and JA Villa-Carrillo<sup>2</sup>

<sup>1</sup>Department of Chemistry and Biological Sciences, Universidad de Sonora, Mexico

<sup>2</sup>Department of Mathematics, Universidad de Sonora, Mexico

## ABSTRACT

Oncologists in Hermosillo, Sonora, Mexico, evaluate breast cancer prognosis based on tumor characteristics and breast cancer risk factors. Some risk factors are age, menopausal status, prior cancer, diabetes mellitus type 2, hypercholesterolemia, and hypertriglyceridemia. The estimation of 5-year overall survival, though, could be influenced by other modifiable risk factors not accounted for in clinical files such as environmental exposure to contaminants, residential exposure to hazardous air pollutants (HAPs), and residual water sites (RWS). We hypothesize that residential exposure to hazards can be added for improving the current criteria of using the variables; previous cancer condition, age, menopause, diabetes mellitus, high triglycerides and high cholesterol, for breast cancer prognosis in predicting the 5-year overall survival. The main result obtained by principal component analysis (PCA) indicated that the variables are positively correlated with the first dimension two by two. That means that size effect is observed where some patients have high values, while others have low values regardless of the variables. The previous statement means that, if a patient has a high value in one variable, then the value tends to be high for the others. The main findings obtained by correspondence analysis is that residentially exposed women with prior cancer(s), diagnosed at an older age ( $\geq 60$  years) had an overall unfavourable 5-year prognosis (survival rate of 69 % or less). This work considers the inclusion of hazard exposure for the estimation of breast cancer prognosis 5-years post-diagnosis in women in Hermosillo, Sonora, Mexico.

**KEYWORDS:** Breast cancer prognosis; Principal component analysis; Hazard exposure

**ABBREVIATIONS:** PCA: Principal Component Analysis; HAPs: Hazardous Air Pollutants; RWS: Residual Water Sites; GPP: Gas Power Plants

## INTRODUCTION

Breast cancer is considered the most common type of cancer diagnosed in women worldwide and the leading cause of female cancer-related deaths in developing countries Ferlay [1]. In Mexico, the National Institute of Statistics and Geography (INEGI) indicated breast cancer as the main cause of female morbidity in 2014 INEGI [2]. A population-based study showed breast cancer as the leading cause of cancer mortality in Mexican women from the period 2000-2013 Mohar-Betancourt [3]. Studies by the Health Secretary of Mexico (SS) also showed breast cancer as the main cause of cancer-

related deaths in 2002. Given the high mortality and morbidity rates in Mexico, the estimation of breast cancer prognosis or overall survival rate of 5 years, is crucial to provide an overview of the patient's scenario long-term. The assessment of breast cancer prognosis considers both tumor characteristics and cancer risk factors, seeking to determine the best therapeutic approach for the patient [4]. There are distinct criteria for estimating the overall 5-year survival for breast cancer. Most validated tools consider tumor characteristics (hormonal status, stage, size), potential invasion to lymph nodes, age, and screening methods for breast

### Quick Response Code:



**Address for correspondence:** Diana Evelyn Villa-Guillen, Ph.D., Department of Chemistry and Biological Sciences, Universidad de Sonora, Mexico

**Received:** November 09, 2020

**Published:** December 16, 2020

**How to cite this article:** DE Villa-Guillen, JA Villa-Carrillo. Estimating the 5-Year Breast Cancer Prognosis in a Cohort Residentially Exposed to Hazards at Hermosillo, Sonora, Mexico. 2020 - 2(6) OAJBS.ID.000240. DOI: [10.38125/OAJBS.000240](https://doi.org/10.38125/OAJBS.000240)

cancer diagnosis. Hormonal breast cancers have a higher mortality with increasing age van de Water et al. [5]. Tumor size at the time of diagnosis is also of ultimate importance for cancer prognosis. A clinically detectable tumor (weight > 1 gram) has 109 cancer cells, and when considering the tumor burden is around 1012 cells, breast cancer is fatal Kroman et al. [6]. Invasion to lymph nodes is an indicator for a poor outcome, and this includes both macro and micro-metastases. Patients with evidence of micro-metastases in axillary lymph nodes have a much more unfavorable prognosis in comparison to node-negative patients Andersson et al. [7]. Age, on the other hand, is considered as an independent risk factor for breast cancer prognosis. A study by Kroman et al. [6] suggest young patients (less than 40 years old) present a higher risk for aggressive, non-hormonal breast cancers with lymph node invasion, with a subsequent higher mortality rate in comparison to that of older groups Kroman et al. [6]. Screening methods for breast cancer diagnosis are considered as a variable for breast cancer prognosis. A study of 2,936 women compared the prognosis of those diagnosed by mammography with other screening methods. Those patients screened by mammography presented a better 15-year survival (86 %) in comparison to those diagnosed with other techniques (66 %) ( $P < 0.0001$ , HR = 2.91) Lehtimäki et al. [8]. Although tumor characteristics, age, and screening methods are well-established factors, prediction tools require a constant evaluation of the variables used for estimating the 5-year overall survival rate.

In this work, oncologists used their own criteria to estimate the overall 5-year prognosis, where their joined judgement represents an abstract evaluation of the variables. The aim of this research is to provide evidence that this criterion is useful to predict 5-year prognosis, where the residential exposure to hazards (hazardous air pollutants (HAPs), residual water sites (RWS), gas power plants (GPP)) can be included for its improvement. Outcomes of the present study seeks to establish hazard exposure as part of the medical criterion for estimating 5-year prognosis at Hermosillo city.

## METHODS

### Study Population

This work is based on a large observational retrospective study that collected clinical cases of female breast cancers from the period of 2013 to 2019 in Hermosillo, Sonora, Mexico [Villa-Guillen et al. 2020]. Briefly, a breast cancer database REDCap, Harris et al. [9] compiled data from women residing at least 10 years in the Hermosillo municipality. The participation criteria included female breast cancer cases, with current residence in Hermosillo, and with an ID-verifiable residential address (as registered by the National Electoral Institute (INE) of Mexico, or by their medical insurance company).

Exclusion criteria were male breast cancers, cancers other than breast, former Hermosillo residents, foreigners, or women without ID-verifiable residential address. Breast cancer cases were those reported by public (Hospital General del Estado de Sonora Dr. Ernesto Ramos Bours (HGE), Centro Estatal de Oncología Dr. Ernesto Rivera Claisse (CEO)) and private health institutions (CIMA, Hospital San José, Clínica del Noroeste) during the period of 2013-2019. Cases older than 2013 were not available in files. Clinical data was de-identified using the Safe Harbor method to be in compliance with national and international regulations for the protection and privacy of human subjects LFM [10]; DOF [11]; HHS [12].

### Oncologists' Criteria for Estimating the 5-Year Breast Cancer Prognosis

Physicians from public and private health institutions evaluated clinical files and estimated the 5-year prognosis post-diagnosis. The criteria were explained in a prior publication Villa-Guillen et al. [13]. Concisely, physicians estimated the 5-year overall survival rate based on clinical data as reported in files. The physicians classified this prognosis into three categories: favorable prognosis, if the expected 5-year survival rate was that of 90 % or higher; intermediary prognosis, if the survival rate was between 89 to 70 %; and unfavourable prognosis, if the patient had an estimated overall survival of 69 % or less.

### Variables for Estimating the 5-Year Breast Cancer Prognosis

Physicians used the following clinical data from files to estimate the overall 5-year survival:

**Tumor characteristics at the time of diagnosis:** Clinical data included breast cancer type by surgery (ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), lobular carcinoma in situ (LCIS), invasive lobular carcinoma (ILC), metastatic breast cancer (MBC), or rare cancers), tumor stage (I to IV) or grade (I to III), size ( $x \geq 2$  cm,  $2$  cm  $< x \leq 5$  cm,  $x > 5$  cm), and hormonal status of the tumor (hormonal positive (ER+ and PR-, ER+ and PR+, ER- and PR+), or triple negative breast cancer (TNBC)) at the time of diagnosis. Clinical data is described elsewhere Villa-Guillen et al. [13].

**Modifiable breast cancer risk factors:** Physicians evaluated the 5-year overall survival rate accounting for modifiable risk factors related to breast cancer. Those are, but are not limited to: breast density (BI-RADS, either by mammogram or ultrasound), parity (nulliparous, parous (only 1 child), multiparous), smoking status (never, current, former, passive smoker), alcohol consumption (abstainer, occasional drinker, alcoholism), weekly red meat consumption (1-2 times, 3-4 times, 5-7 times), exercise habits (sedentary life style, exercises frequently), use of hormonal contraception (never, current, or former user), use of hormonal replacement therapy (HRT) (not applicable, never, current or former user), and presence of elements of metabolic syndrome (as defined by the NCEP ATP III criteria) Huang et al. [14].

**Non-modifiable breast cancer risk factors:** Physicians considered the age at the time of diagnosis, age at menarche, menopausal status (premenopausal, perimenopausal, postmenopausal), personal history of cancer (prior breast cancer, cancer other than breast), family history of cancer (first-line relatives, second-line relatives), and prior chronic illnesses other than neoplasia (hypothyroidism, other chronic illness).

**Hazard exposure as a modifiable risk factor:** Other variables considered for this research as modifiable risk factors, but non-considered by physicians, were the residential exposure to hazards (HAPs, RWS, GPP). A prior study Villa-Guillen et al. [13] used proximity analysis using geographic information systems (ArcGIS version 10.7.1) to classify neighbourhoods as hazard-exposed or non-hazard exposed. Physicians had no access to the residential information of the patient, neither had knowledge of the patient's status in terms of hazard exposure. The geospatial analyst had no access to clinical data used by physicians for estimating the 5-year prognosis. Full data was disclosed only to the statistician.

**Hazard exposure as a modifiable risk factor:** Hazard exposure was evaluated in terms of the following variables per neighbourhood: gas power plants (GPPs), residual water sites (RWS), and the number of active industries (potential source of hazardous air pollutants or HAPs). Proximity analysis was used to determine the cut-offs for GPP-exposed and RWS-exposed neighbourhoods. Those within the established cut-off values were categorized as exposed, whereas neighbourhoods beyond the cut-off values were categorized as non-exposed. Geographical information software (ArcGIS version 10.7.1) was used to conduct proximity analyses for all neighbourhoods. The cut-off values from those analyses defined GPP-exposed neighbourhoods as those within  $X \leq 4$  km from a gas power plant (GPP). RWS-exposed neighbourhoods were those within  $X \leq 3$  km from a residual water site (RWS). Neighbourhoods with at least 7 active chemical industries were categorized as HAP-exposed, whereas those with equal or less than 6 industries were categorized as non-HAP-exposed.

**Residential exposure to Hazardous Air Pollutants (HAPs):** Women were considered as HAPs-exposed if their residential address was in a neighbourhood with at least 7 active industries producing chemical hazards. Women residing in neighbourhoods with 6 industries or less were considered as non-HAPs-exposed.

INEGI provided the coordinates (latitude, longitude) of active industries in Hermosillo city. Industries considered as potential producers of chemical hazards were those classified by INEGI Census 2010 as follows: production of industrial laminates, furniture, coolers, and AC, dental laboratories, production of pesticides and fungicides, manufacturers of orthopaedic materials, candle production, chemical laboratories, metal production, automobile assembly, manufacturers of agricultural products, technology machinery production and assembly, manufacturers of cleaning products, electronics production, and assembly, production of dyes and construction materials, iron foundries, printer suppliers, welding shops, production of toys and jewellery, manufacturers of chemical and clinical products, aquaculture products, polymers' production, mechatronics, electrics, among others.

**Residential exposure to Residual Water Sites (RWS):** INEGI Census 2010 provided the coordinates (latitude, longitude) of active and inactive RWS in Hermosillo city. Women considered as residentially exposed to RWS were those reporting to live in a neighbourhood within a radial distance of 3 km to an RWS. Women residing in a neighbourhood beyond 3 km were classified as non-RWS-exposed.

**Residential exposure to Gas Power Plants (GPPs):** INEGI provided the coordinates (latitude, longitude) of the two GPPs at Hermosillo city. Proximity analysis classified neighbourhoods as GPP-exposed if they were within a radial distance of 4 km. Neighbourhoods beyond 4 km were classified as non-GPP-exposed. Women were categorized as GPP-exposed if their residences were reported in a GPP-exposed neighbourhood. Those patients reporting as living in a non-GPP-neighbourhood were classified as non-GPP-exposed.

### Correspondence Analysis (CA) and Principal Component Analysis (PCA)

Correspondence Analysis (CA) provides visualizations of associations in a two-way contingency table in a small number of dimensions. Those tables are built using the variables prognosis and one or more of the remaining.

Principal Component Analysis (PCA) is a statistical method meant for reducing dimensionality of a dataset, while preserving as much of the variability, increasing interpretability and reducing information loss Jolliffe et al. [15]. In the present work, PCA analysed the variables used by physicians to estimate the overall 5-year breast cancer prognosis in a cohort of patients at Hermosillo, Sonora, Mexico. PCA analysis included the residential exposure to hazards (GPP-exposed, RWS-exposed, and/or HAP-exposed). The statistical analysis was carried out using R version 3.6.3.

### Ethics Statement

The present research was reviewed and IRB-approved by the Secretary of Health, the University of Sonora, and by the hospitals (HGE, CEO, CIMA, Hospital San José, Clínica del Noroeste) participating in this study. National regulations considered this research as without harm for the individual (General Law of Health for Clinical Research, Article 17); SSA [16].

## RESULTS

### The Overall 5-Year Breast Cancer Prognosis of the Study Population

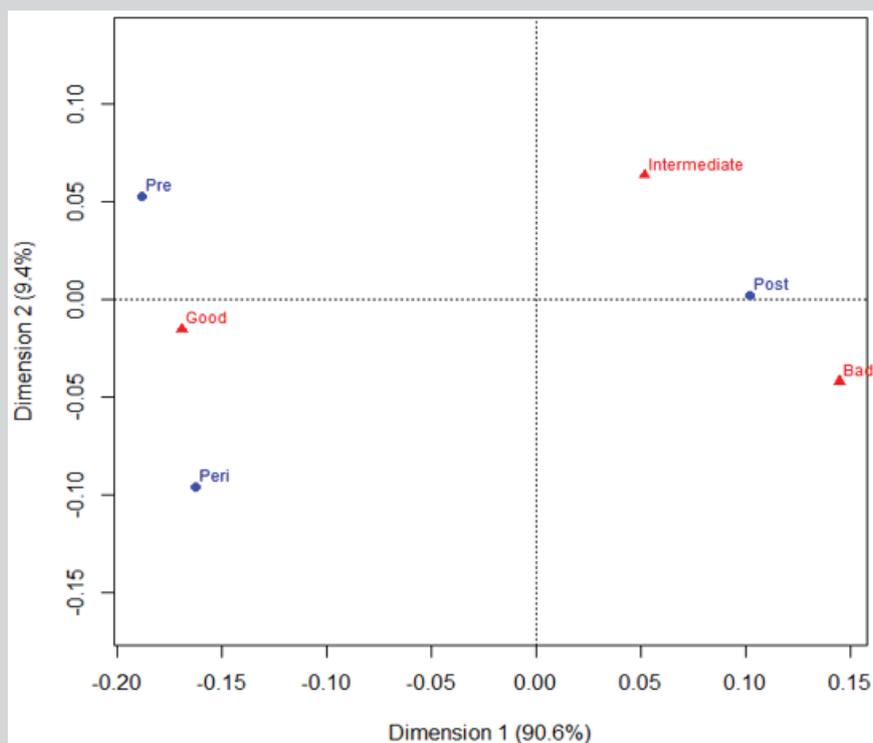
Physicians analysed the variables for estimating the overall 5-year breast cancer prognosis for  $n=297$  cases collected from physical files of the period of 2013-2018. All cases collected were of Hermosillo residents.

The clinical characteristics of the study cohort (tumor characteristics at the time of diagnosis, modifiable, and non-modifiable breast cancer risk factors) are described elsewhere Villa-Guillen et al. [13]. According to the physicians' criteria, the estimated 5-year prognosis classified 37.37 % of the patients ( $n=111$ ) with a favorable overall survival (life expectancy of 5 years post-diagnosis is that of 90 % to 100 %), 29.97 % ( $n=89$ ) with an intermediary prognosis (life expectancy of 70 % to 89 %) and 32.66 % ( $n=97$ ) with an unfavourable 5-year prognosis (life expectancy of 69 % or less). Geospatial analysis for evaluating the residential exposure to hazards is described in detail in a prior publication Villa-Guillen et al. [17]. Briefly, the study cohort showed an OR = 1.55 ( $p$ -value < 0.0001) for neighbourhoods GPP-exposed, an OR = 1.69 ( $p$ -value < 0.0001) for HAP-exposed neighbourhoods and an OR = 1.45 ( $p$ -value < 0.0001) for RWS-exposed locations. Cumulative residential exposure to all three hazards had an OR = 2.29 ( $p$ -value < 0.0001) in comparison to those women living in non-hazard-exposed neighbourhoods.

### Correspondence Analysis of the Variables Used for Estimating the 5-Year Breast Cancer Prognosis

The variables used by physicians to estimate the overall 5-year breast cancer prognosis were analysed by CA. Those described in this work are menopausal status, age at the time of diagnosis, residential exposure to hazards (zero, one, two, or three hazards), familiar history of cancer, prior cancer(s), diabetes mellitus type 2, hypertension, high cholesterol, and high triglycerides.

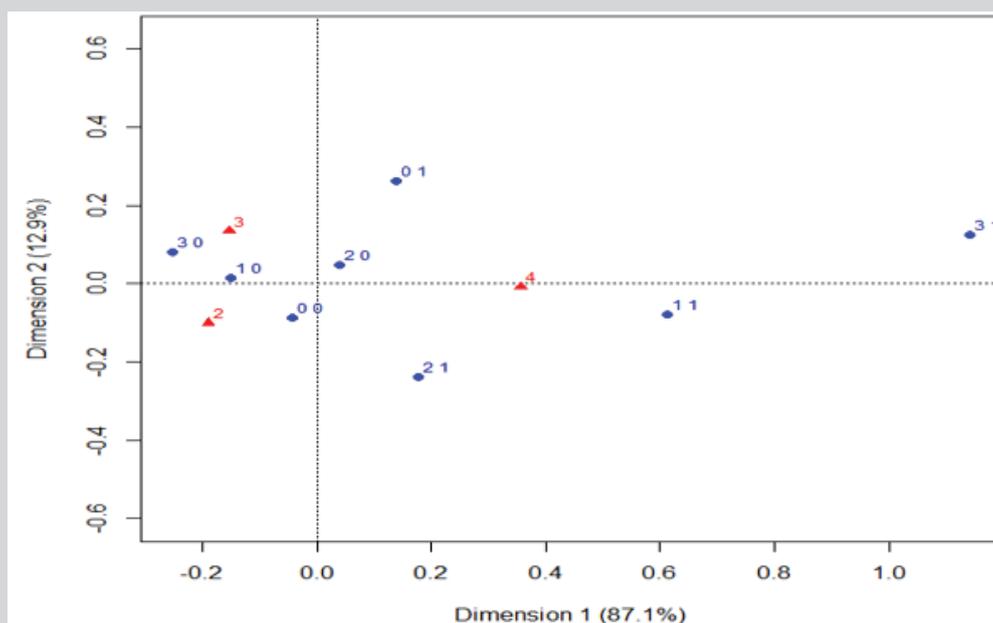
**Menopausal status and the overall 5-year breast cancer prognosis:** Menopausal status of the cohort indicates that those women classified as premenopausal or perimenopausal at the time of diagnosis shows association with an expected overall favorable 5-year prognosis (Figure 1). Postmenopausal women show association with either an intermediary or an unfavourable 5-year breast cancer prognosis. On the other hand, a favorable prognosis is related to pre-menopause or perimenopause.



**Figure 1:** Association between the variables of menopausal status at the time of diagnosis, and the estimated overall 5-year prognosis.

**Residential exposure to hazards, prior cancer(s), and the 5-year overall breast cancer prognosis:** CA analysed the variables of hazard exposure, prior cancers, and 5-year overall prognosis in the study cohort (Figure 2). It is observed that women with no prior cancer(s) are associated with a favorable prognosis. Women considered as residentially exposed to one or three hazards, but without prior cancer(s), are associated with an intermediary 5-year prognosis. Those women with prior cancer(s) are associated with

an overall unfavourable 5-year prognosis, whereas they were exposed to hazards (one hazard) or not. In other words, the graph shows association between favorable and intermediary prognosis, (triangles with numbers two and three) and points without previous cancer ((0, 0), (2, 0), (1, 0), and (3, 0)). Also, unfavourable prognosis (triangle with number four) is associated with several levels of exposition ((2, 1), (1, 1), and (3, 1)) and previous cancer.

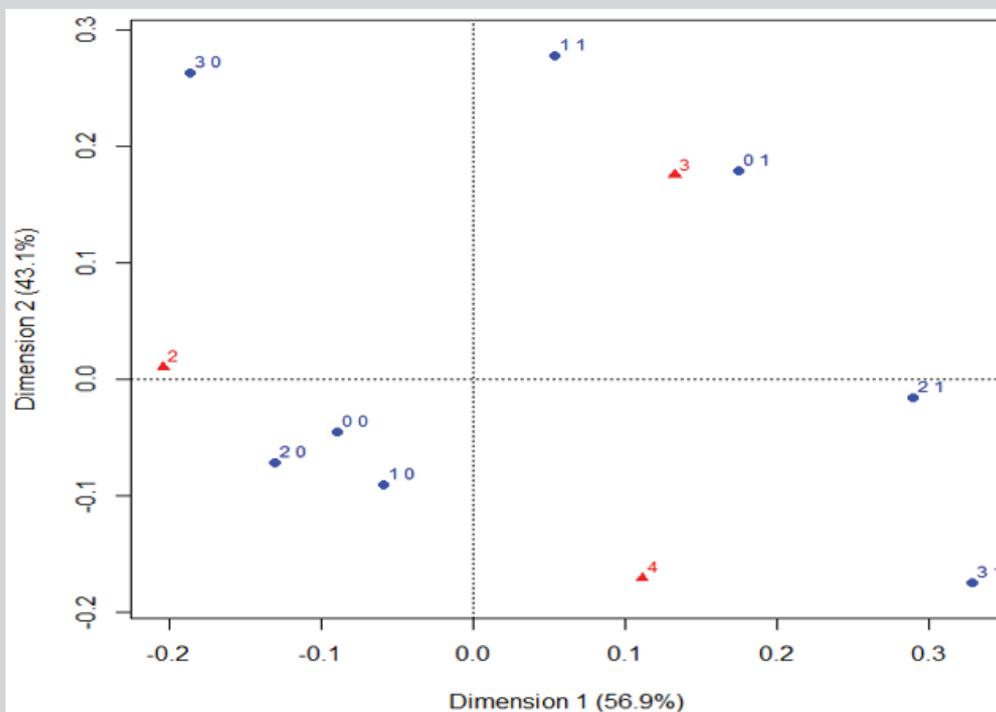


**Figure 2:** Correspondence analysis of the variables of hazard exposure, previous cancer, and the estimated overall 5-year prognosis.

**Residential Exposure to Hazards, Age at the Time of Diagnosis, and the 5-Year Overall Breast Cancer Prognosis:**

According to Figure 3, young women (less than 60 years old) are associated with a favorable 5-year prognosis, independently of their hazard exposure. Those women equal or older than 60 tend to have an intermediary 5-year prognosis and can be exposed to one hazard or non-exposed. Women with a poor 5-year prognosis are equal or older than 60 and can be residentially exposed to two

or three hazards. In other words, the graph shows association between favorable prognosis and younger age (triangle with number two), where there are points of women younger than 60 years ((0, 0), (1, 0), (2, 0), and (3, 0)). Also, both unfavourable and intermediary prognoses (triangle with numbers three and four) are associated with hazard exposure ((1, 1), (0, 1), (2, 1) and (3, 1)), and with an older age at the time of diagnosis (women equal or older than 60 years).

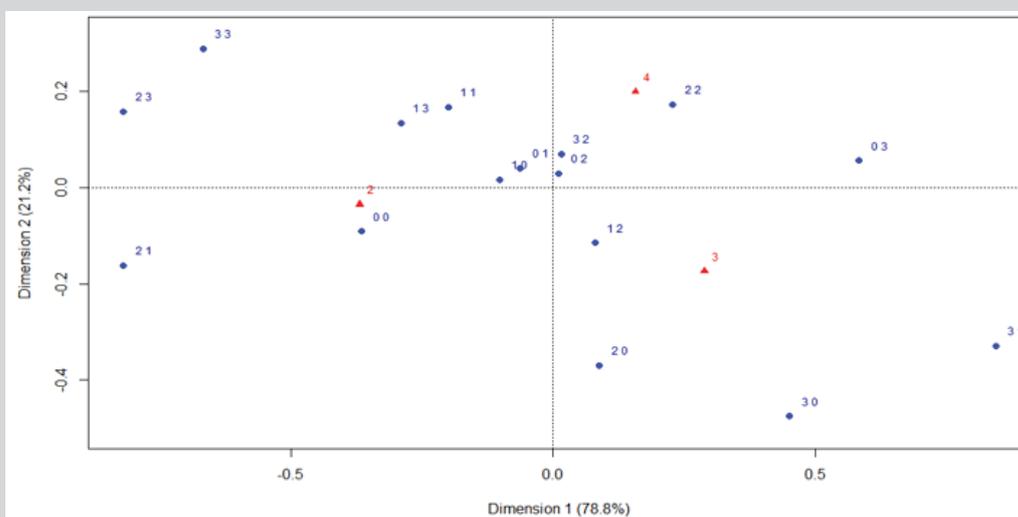


**Figure 3:** Correspondence analysis of the variables of hazard exposure, age at the time of diagnosis, and the estimated overall 5-year prognosis.

**Residential Exposure to Hazards, Menopausal Status, and the 5-Year Overall Breast Cancer Prognosis:**

Women with an overall favorable 5-year prognosis relate to non-exposure to hazards and premenopausal status (Figure 4). The graph shows an association between favorable prognosis for premenopausal women non-exposed to hazards (point (0,0)) or exposed to one

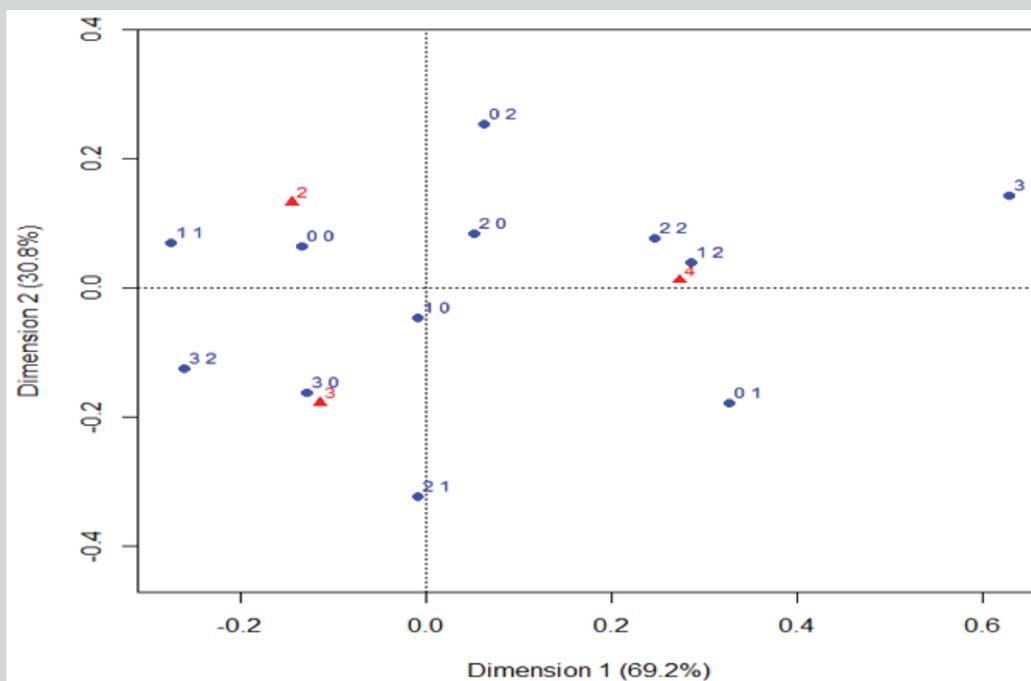
hazard (point (1,0)). Women with unknown menopausal status, but exposed to one hazard, also relate to a favorable prognosis (point (1,3)). Hazard exposure relates to both intermediary and an unfavourable 5-year prognosis, regardless of the menopausal status.



**Figure 4:** Correspondence analysis of the variables of hazard exposure, menopausal status, and the estimated overall 5-year prognosis.

**Menopausal status, diabetes mellitus type 2, and the overall 5-year breast cancer prognosis:** According to Figure 5, premenopausal women without diabetes mellitus (0,0) are expected to have an overall favorable 5-year prognosis. This prediction is the same for the peri-menopausal and non-diabetic women. Patients

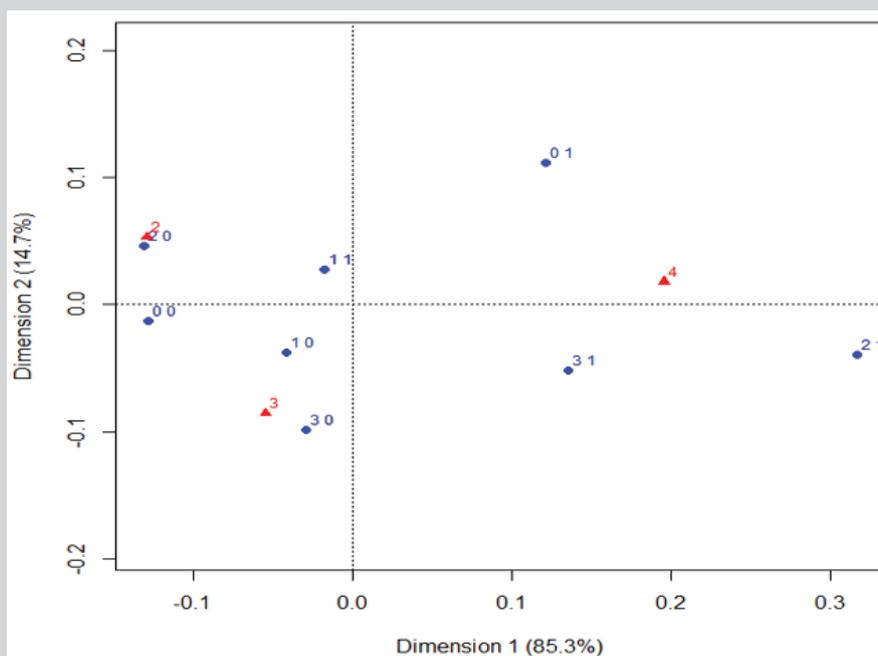
with unknown menopausal status, independently of the presence or absence of diabetes, are predicted to have an intermediary 5-year prognosis. Peri- and postmenopausal women with unknown diabetes status are expected to have an unfavourable overall 5-year survival.



**Figure 5:** Correspondence analysis of the variables of menopausal status, diabetes mellitus type 2, and the estimated overall 5-year prognosis.

**Residential exposure to hazards, hypertension, and the overall 5-year breast cancer prognosis:** CA (Figure 6) indicates women without hypertension, and residentially exposed to at most two hazards, are predicted to have an overall favorable 5-year breast cancer prognosis. Additionally, women residentially exposed

to at most one hazardous site (GPP-exposed, RWS-exposed, or HAP-exposed) are expected to have an intermediary 5-year prognosis. Women with hypertension are predicted to have an unfavourable 5-year overall survival.



**Figure 6:** Correspondence analysis of the variables of hazard exposure, hypertension, and the estimated overall 5-year prognosis.

### PCA Reduces Dimensionality for the Variables Used to Predict the Overall 5-Year Breast Cancer Prognosis

The main results are that the six variables considered in this study (age, menopausal status, residential exposure to hazards, high cholesterol, high triglycerides, diabetes mellitus type 2) are

highly correlated with the first dimension (Dim 1) (Figure 7) & (Table 1). This means that, if a variable has a high value, the values for the remaining variables tend to be high. If a variable has a low value, the others tend to have low values. Those subpopulations compose two groups. There is a third group, where there are distinct variables with both high and low values.

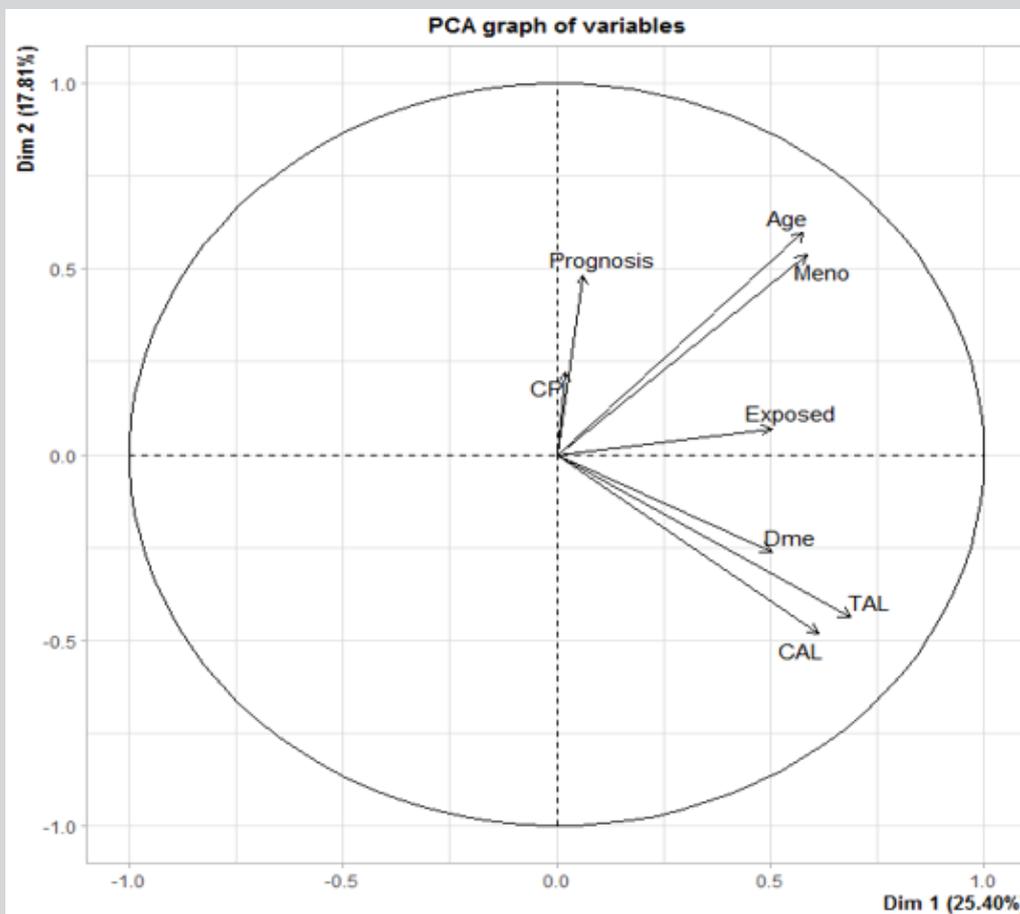


Figure 7: PCA graph of the variables used by physicians to estimate the overall 5-year breast cancer prognosis.

Table 1: Correlation and p-values with the first component.

	TAL	CAL	Meno	Age	Exposed	Dme
Correlation	0.68	0.61	0.58	0.57	0.5	0.5
P-value	1.32E-42	6.16E-32	1.66E-28	1.37E-27	1.67E-20	2.53E-20

All the active variables are within the same side of the first component, where the sign of the correlation coefficient is the same for all (Figure 7) & (Table 1). We are dealing with size effect, where some of the patients have high values, while others have low values regardless of the variable. In other words, the variables are positively correlated two by two. This correlation is positive, meaning that, when a variable has high values, there are high values for the other ones. For patients with an overall favorable prognosis, all variables present low values. Patients with a poor outcome, on the other hand, present variables with high values. There is a small quantity of patients (intermediary 5-year prognosis) where there are distinct variables with both high and low values.

The present study cohort has distinct correlations between variables. The variables of age and menopausal status at the time of diagnosis are highly correlated. The presence of diabetes mellitus type 2, high cholesterol ( $x > 200$  mg/dl), and high triglycerides ( $x > 150$  mg/dl) are also highly correlated. Residential exposure to

hazards (GPP-exposed, RWS-exposed, and/or HAP-exposed) by itself, has a small correlation to the overall 5-year breast cancer prognosis and to prior cancer(s).

Overall, correspondence analysis (CA) indicates that the variables tested for this cohort are positively correlated two by two. The later means that there is a size effect where some patients present high values, while others have low values regardless of the variables. The correlation observed for this cohort is positive, where if a patient has high values in one variable, then the value tends to be high for the remaining ones or vice versa.

Principal component analysis (PCA) indicates reduction to potentially four variables for this study cohort. Those are age, menopausal status, prior cancer(s), and the overall 5-year prognosis. The variables of age at the time of diagnosis, and menopausal status, are on the same direction. Similarly, the variables of prior cancer(s) and the overall 5-year prognosis follow

a similar directionality. Diabetes mellitus type 2, high triglycerides, and high cholesterol are in the same direction, which means that they are highly correlated. Prognosis and previous cancer(s) have the same direction, meaning that physicians are highly influenced for this information. Physicians are not influenced by the variable of hazard exposure. This result was expected, because physicians did not know about hazard exposure for each patient (Figure 7).

## DISCUSSION

This work produces two main findings. The first indicate that the variables age, menopause, levels of exposure, diabetes mellitus, high triglycerides and high cholesterol have a positive correlation. This means that, if a patient has a high value for one variable, then the value tends to be high for the others. For this study population, the correspondence analysis (CA) indicates that residentially exposed women with prior cancer(s), diagnosed at an older age ( $x \geq 60$  years old) have a poor overall survival 5-years post-diagnosis (life expectancy is less than 69 %). The variable of hazard exposure can add to the criterion used for physicians for estimating the survival rate at the time of diagnosis in Hermosillo residents.

Other studies evaluated clinical variables for estimating the overall survival. Prognostic tools use the tumor characteristics to determine aggressiveness and estimate the patient's lifespan. Among those characteristics are hormonal status, stage, size of the tumor, lymph node invasion, and age. Lower life expectancy is expected for triple-negative breast cancers, and those cancers detected at advanced stages (III or IV) with lymph node invasion or metastasis. The role of age is controversial for estimating the overall survival rate 5-years postdiagnosis. Some studies indicate patients younger than 40 years old present more aggressive cancers and lower survival rates. Older breast cancer patients also present poor prognosis. Screening methods may come into play for an early detection, where mammography remains as the gold standard for breast cancer diagnosis.

Other modifiable risk factors may influence breast cancer prognosis, like residential exposure to hazards. Prior works of this research group indicate a relation between a unfavourable 5-year prognosis and hazard exposure for breast cancer patients residing at Hermosillo city Villa-Guillen et al. [13].

PCA confirms those findings, where the variable of residential exposure has the same direction than that for age and menopausal status. This means that postmenopausal patients are more likely to be hazard-exposed, adding a risk factor that may compromise their overall 5-year survival rate. In this cohort, age and menopausal status are not related to hypercholesterolemia or high triglycerides. Residential exposure to contaminants, by itself, is not related to breast cancer prognosis, as physicians were unaware of this information. Those results indicate that hazard exposure may act in conjunction with other risk factors related to an unfavourable outcome.

## STRENGTHS AND LIMITATIONS

This work has several strengths regarding its statistical approach. PCA enables to preserve most of the variability of the data, allowing to increase the interpretability of the variables used by physicians to estimate the overall 5-year breast cancer prognosis, without compromising information loss. Limitations of the present work are the small cohort ( $n = 297$  patients), and the homogeneous ethnicity of the study population, as it is comprised solely by Hispanics residing at Hermosillo city. Current efforts are in course to include multi-ethnic populations in future works.

## CONCLUSION

For Hermosillo residents, principal component analysis (PCA) indicates women residentially exposed to hazards (GPP-exposed, RWS-exposed, and/or HAP-exposed) who had cancer before their current diagnosis, with obesity ( $BMI \geq 30 \text{ kg/m}^2$ ) and diagnosed at an older age ( $x \geq 60$  years old) have an overall reduced life expectancy 5-years post-diagnosis (69 % or less). The criterion used by physicians to estimate breast cancer prognosis is useful, where hazard exposure can be added as an additional variable for its improvement. Large studies could confirm these findings.

## ACKNOWLEDGEMENT

The first author of this research was the recipient of the Consortium Arizona-Mexico Arid Environments (CAZMEX) Award 2018 for Postdoctoral Stays. CAZMEX is formed by the University of Arizona and the Mexican Council of Science and Technology (CONACyT). Special thanks to the academic institution Universidad de Sonora and their staff for supporting the present research work.

The authors want to thank all the physicians collaborating in this work: Dr. Enrique Avila-Monteverde of the Clinica del Noroeste; Dr. Jose Heliodoro Gonzalez-Zepeda of the Hospital San Jose; Dr. Luis Felipe Munguia-Ibarra of the Hospital General del Estado de Sonora Dr. Ernesto Ramos Bours (HGE); Dr. Karla G. Aguilar-Gutierrez, Dr. Cynthia Roja-Camarena, Dr. Luis C. Durazo-Cons, Dr. Carlos D. Luque-Morales, Dr. Wendy B. Aguilar-Peraza, and Dr. Jorge A. Cordon-Guillen of the Centro Estatal de Oncología Dr. Ernesto Rivera Claisse (CEO); and Dr. Baldemar Corral-Villegas of CIMA. Special thanks to Armando Yanez from INEGI, and to Shelene Tarazon from HGE, for their help in providing an important quantity of the geospatial and clinical data. Thank you to the following university professors who contributed with data of hazardous pollutants from their prior research works: Martin Jara-Marini, CIAD; Graciela Caire-Juvera, CIAD; Ana Isabel Valenzuela-Quintanar, CIAD; Ivan Anduro-Corona, CIAD; Leticia Garcia-Rico, CIAD. Thank you to Dr. Eduardo Ruiz-Bustos for his help with IRB submission and registry of the present study at the Universidad de Sonora. We are very thankful to the teaching heads M.S. Maru Eugenia from Clinica del Noroeste, Dr. Carlos Gonzalez-Becuar from HGE, and Dr. Frankfer Urias from CEO, for their support in clinical data acquisition. Special thanks to Tanyha Zepeda from the University of Arizona, for her assistance with REDCap platform.

## REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, et al. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 136(5): E359-386.
2. Instituto Nacional de Estadística y Geografía (2014) Estadísticas a propósito del Día mundial de la lucha contra el cáncer de mama 19 de octubre. INEGI, México.
3. Mohar-Betancourt A, Reynoso-Noverón N, Armas-Texta D, Gutiérrez-Delgado C, Torres-Domínguez JA (2017) Cancer trends in Mexico: Essential data for the creation and follow-up of public policies. *J Glob Oncol* 3(6): 740-748.
4. Health secretary of Mexico (2002) Programa de Acción: Cáncer de mama. SS, México.
5. Van de Water W, Markopoulos C, Van de Velde CJH, Seynaeve C, Hasenburg A, et al. (2012) Association between age at diagnosis and disease-specific mortality among postmenopausal women with hormone receptor-positive breast cancer. *JAMA* 307 (6): 590-597.
6. Kroman N, Melbye N, Mouridsen HT (2002) Prognostic influence of age at diagnosis in premenopausal breast cancer patients. *SJS* 91: 305-308.

7. Andersson Y, Frisell J, Sylvan M, de Boniface J, Bergkvist L (2010) Breast cancer survival in relation to the metastatic tumor burden in axillary lymph nodes. *J Clin Oncol* 28(17): 2868-2873.
8. Lehtimäki T, Lundin M, Linder N, Sihto H, Holli K, et al. (2011) Long-term prognosis of breast cancer detected by mammography screening or other methods. *Breast Cancer Res* 13(6): R134.
9. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, et al. (2009) Research electronic data capture (REDCap) -- a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 42(2): 377-381.
10. Leyes Federales de México. (2010) Ley Federal de Protección de Datos Personales en Posesión de los Particulares.
11. Diario Oficial de la Federación (2012) NORMA Oficial Mexicana NOM-024-SSA3-2012, Sistemas de información de registro electrónico para la salud.
12. (2015) US Department of Health and Human Services, Office of the Secretary, Office for Civil Rights, & Ocr. Methods for De-identification of PHI.
13. Villa-Guillen DE, Avila-Monteverde E, Gonzalez-Zepeda JH, Munguia-Ibarra LF, Corral-Villegas B, et al. (2020) Breast cancer prognosis and exposure to hazardous contaminants: an observational retrospective study at Hermosillo, Mexico. *Open Acc J Bio Sci* 2(4): 425-443.
14. Huang PL (2009) A comprehensive definition for metabolic syndrome. *Dis Model Mech* 2(5-6): 231-237.
15. Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. *Phil Trans R Soc A* 374: 20150202.
16. (2003) Secretaría de Salud. Reglamento de la Ley General de Salud en Materia de Prestación de Servicios de Atención Médica.
17. Villa-Guillen DE, Avila-Monteverde E, Gonzalez-Zepeda JH, Munguia-Ibarra LF, Corral-Villegas B et al. (2019) Breast cancer risk and residential exposure to environmental hazards in Hermosillo, Sonora, Mexico. In: Proceedings of the 2019 San Antonio Breast Cancer Symposium; San Antonio, TX. Philadelphia (PA): AACR; *Cancer Res* 80(4 Suppl): P5-08-21.