



EPIc Series in Computing

Volume 58, 2019, Pages 236–246

Proceedings of 34th International Conference on Computers and Their Applications



Ethnicity-Related Survival Analysis of Patients with Triple-Negative Breast Cancer

M. Mehdi Owrang Ojaboni¹, Yasmine M. Kanaan², and Robert L. Dewitty JR³

¹ Department of Computer Science, American University, Washington, DC 20016

² Department of Microbiology, Cancer Center, Howard University, Washington, DC 20059

³ Providence Hospital, Washington, DC 20017

owrang@american.edu, ymkanaan@howard.edu, rldewittyjr@aol.com

Abstract

Breast cancer prognostication is a vital element for providing effective treatment for breast cancer patients. Different types of breast cancer can be identified based on the existence or lack of certain receptors (i.e., estrogen, progesterone, her2 receptors). Triple-negative breast cancer (TNBC) is characterized by a lack of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) expression. Existing studies suggest that TNBC patients tend to have worse prognosis compared to non-TNBC counterparts. The incidence of breast cancer and prognosis in women differ according to ethnicity. Given the poor prognosis of TNBC, cancer-related outcomes must be estimated accurately. Several factors responsible for the poor clinical outcomes observed in TNBC, including age, race/ethnicity, grade, tumor size, lymph node status among others, have been studied extensively. Available research data are not conclusive enough to make a convincing argument for or against a biological or clinical difference in TNBC patients based on these factors. This study was designed to investigate the effects of the ethnicity on breast cancer survivability among TNBC patients utilizing population-based Surveillance, Epidemiology, and End Results (SEER) data to confirm whether ethnicity factor has prognostic significance.

1 Introduction

Breast cancer is the most common female cancer in the US, the second most common cause of cancer death in women [1]. In 2018, an estimated 266,120 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S., along with 63,960 new cases of non-invasive (in situ) breast cancer [6]. About 40,920 women in the U.S. are expected to die in 2018 from breast cancer.

The lifetime probability of developing breast cancer is one in six overall (one in eight for invasive disease) [1, 26]. Breast cancer incidence is increasing while mortality is declining in many high-income countries [30].

Medical prognostication is an evaluative component of medicine that encompasses the science of estimating the complication and recurrence of disease and predictive survival of patients [21]. Medical prognosis plays an increasing role in health care outcome. Many factors, including tumor grade, tumor size, and lymph node status may influence or correlate with prognosis for breast cancer patients.

Different types of breast cancer can be identified based on the existence or lack of certain receptors (i.e., estrogen, progesterone, her2 receptors) [22]. Triple-negative breast cancer (TNBC) is characterized by a lack of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) expression [3, 5, 7, 10, 11, 12, 17, 23]. Previous and current studies suggest that TNBC patients tend to have worse prognosis compared to non-TNBC counterparts [2]. In addition, those studies suspect that the incidence of breast cancer, and prognosis in TNBC women, differ according to ethnicity.

Existing works on TNBC are mainly focused on the comparisons of the TNBC and non-TNBC with respect to the clinical, pathological, histological, prognostic features, and outcome associated with these two breast cancer subtypes. Available research data are not conclusive enough to make a convincing argument for or against a biological or clinical difference in TNBC patients based on these factors.

The aim of our study is to look at the TNBC patients to see if ethnicity is a main determinant of long-term survival in women suffered from TNBC tumor type. In this work, we present an analysis of the prognosis for the TNBC patients. We have done our survival analysis of the TNBC patients on National Cancer Institute's SEER dataset [25] based on ethnicity, ethnicity and age, ethnicity and tumor grade, and ethnicity and five year survival. We also applied the NPI (Nottingham Prognostic Index) [20] scheme on our breast cancer dataset to compare patients' survival rate based on their ethnicity and the NPI identified prognosis (Excellent, Good, Moderate, Poor). In addition, we used Kaplan-Meier survival curve [4] to look at TNBC survival based on ethnicity.

2 Triple Negative Breast Cancer (TNBC) Related Works

TNBC accounts for 10%–20% of all breast cancer cases. Although breast cancer is one of the most frequent malignancies, histopathological features, and long-term outcome of breast cancer especially TNBC type are already uncertain [8, 12, 23]. Compared to other breast carcinoma subtypes, TNBC correlates with more aggressive characteristics such as larger tumor size, higher histological grade, more positive lymph nodes, advanced stage, younger age and consequently more often in premenopausal women and is diagnosed more frequently in African-American/Hispanic women [7, 9, 13, 16, 24, 28]. Taken together, these adverse factors may be a major reason that TNBC patients suffer from poorer overall survival (OS), breast-cancer-specific survival (BCSS) and relapse-free survival (RFS) reported for this disease [2]. There have been several retrospective studies comparing the TNBC and non-TNBC patients [2, 29]. The TNBC subtype generally carries a worse prognosis than its non-TNBC counterpart does.

Review of different studies showed that race/ethnicity (African American ethnicity, Hispanic, White) as one of the main risk factors of TNBC. TNBC has a higher predilection for certain ethnicities, which is why its incidence has ranged from 11.2% in studies with a predominantly white patient population to as high as 39% in studies with a larger proportion of black patients [3, 29]. African American premenopausal women are known to have the highest incidence of TNBC, reporting up to 30% [28], contributing to increased mortality. In a large retrospective analysis on clinicopathological features of TNBC in China and another study from Taiwan presented discordant results with Western studies, concluding that race is an important factor for TNBC [18, 31]. In a study reported by [29], thirty-nine patients had TNBC and 303 had non-TNBC. TNBC was more common in black than in white patients (58.97% vs 35.90%; OR, 2.755; $P=.004$).

Clarke and colleagues [9] showed that incidence of TNBC is higher in African-American women than other racial or ethnic origin groups at all ages ($p<0.05$). African-American women were twice as likely to be diagnosed with ER-positive, PR-positive, and HER2-negative breast cancer; however, the ratio was substantially lower than reported for white (ratio 6.9:1) or Asian women (ratio 6.1:1). An analysis of the entire US SEER data of women diagnosed with breast cancer in 2010, added evidence to support this result, showing that African-American (odds ratio (OR)= 1.4, 95% CI 1.2–1.6) and Hispanic women (1.3, 1.2–1.5) were more likely to be diagnosed with TNBC than were white women [7].

In a study population conducted by Lara-Medina [16], after re-evaluating immunohistochemistry results from each sample, observed that the prevalence of TNBC was 23.1% in Hispanic patients. This prevalence was higher than that reported in white patients (range, 10%-13%) but was closer to the prevalence reported in African-American patients (range, 23%-30%) [19].

Overall, epidemiological data support the conclusion that although TNBC is not restricted to a specific age or ethnic group, this cancer is of a higher frequency and is a contributor to the survival disadvantage of women of African ancestry with breast cancer. Factors that might account for variations in the incidence and prevalence of TNBC in women of African ancestry include differences in methods of case ascertainment, population age structure, genetic and lifestyle risk factor distribution, and access to mammography screening [7, 27].

3 TNBC Survival Experiments and Analysis

Medical prognosis is a field in medicine that encompasses the science of estimating the complication and recurrence of disease and to predict survival of patient [15]. Survival analysis is a field in medical prognosis that deals with the application of various methods to estimate the survival of a particular patient suffering from a disease.

Traditionally, a number of clinicopathological characteristics, such as patient age, tumor size, histological grade, hormone receptor status, her2 status, lymphovascular invasion, and lymph node involvement have been used to determine the prognosis of breast cancer patients. However, there are reports that suggest, with the exception of lymph node involvement, those traditional markers are of little value in predicting the prognosis of TNBC patients [8].

In our experiments, we have used the Nottingham Prognostic Index (NPI) [20] on SEER breast cancer dataset to determine patient's prognosis. Its value is calculated using three pathological criteria: the size of the tumor; the number of involved positive lymph nodes; and the grade of the tumor.

The score for the index is calculated using the formula:

$$NPI = [0.2 \times S] + N + G, \text{ Where:}$$

S is the size of the tumor in centimeters,

N is the number of positive lymph nodes involved, where:

N=1, if nodes involved is 0,

N=2, if nodes involved is 1-3

N=3, if nodes involved is >3

G is the grade of tumor, where:

Grade I =1, Grade II =2, Grade III =3

The interpretation is as follows

<u>Score</u>	<u>5-Year Survival</u>	<u>Prognosis</u>
≥ 2.0 to ≤ 2.4	93%	Excellent
> 2.4 to ≤ 3.4	85%	Good
> 3.4 to ≤ 5.4	70%	Moderate
> 5.4	50%	Poor

With NPI, it is possible to group patients with respect to prognosis as those with “Excellent prognosis”, “Good prognosis”, etc. as defined above. That is why we have decided to do an exploratory analysis of breast cancer data sets using NPI and to group the patients based on their prognosis and ethnicity categories in better understanding the correlation between prognosis and ethnicity.

Despite the widespread acknowledgement of the poor clinical outcome of TNBC, the prognostic value of specific biological feature of these tumors continue to raise substantial degree of uncertainty and controversy.

The aim of this study is to analyze the significance of the ethnicity factor on the overall prognosis and survive rate of the TNBC patients. In the following subsections, we discuss the survival experiments that are performed on SEER breast cancer dataset and analyzed using IBM SPSS Survival tool, Cox Proportional Hazard Ratio and Kaplan-Meier survival curve [4, 14]. Survival rates of the TNBC patients based on NPI scheme are examined as well.

3.1 Survival Rates of the TNBC Patients based on Ethnicity

Based on Table 1, we find that the odds of survival of the TNBC white patients are 1.26 times as high as the ones for the black TNBC patients (Odds of Ratio OR=1.2568, P = 0.0066.). Figure 1 shows the Kaplan-Meier (KM) survival months for TNBC patients by Ethnicity. It is graphed based on the data contained in Table 1. As the P-value is less than 0.05, it can be concluded that there is a significant evidence of a difference in survival time for black and white TNBC patients. The difference, however, is too small to be noticeable from the KM graph in Figure 1.

3.2 Survival Rates of the TNBC patients based on Ethnicity and Age

Looking at Table 2, we can find that the odds of survival for the older black TNBC patients are 1.72 times as high as the younger patients (OR=1.72, P = 0.0109). In addition, the odds of survival for the older white TNBC patients are 2.2 times as high as the younger white patients (OR=2.23, P <

0.0001). The sample sizes for the Chinese and Japanese TNBC were too small to be considered in this analysis.

Ethnicity	Total N	N of Events	N of Censored
Black	2428	2223	205
Chinese	150	145	5
Japanese	124	121	3
Other	648	614	34
White	9011	8395	616
Overall	12361	11498	863

Table 1. KM Case Processing Summary data for TNBC patients based on Ethnicity group (N of Events = survived, N of Censored=died)

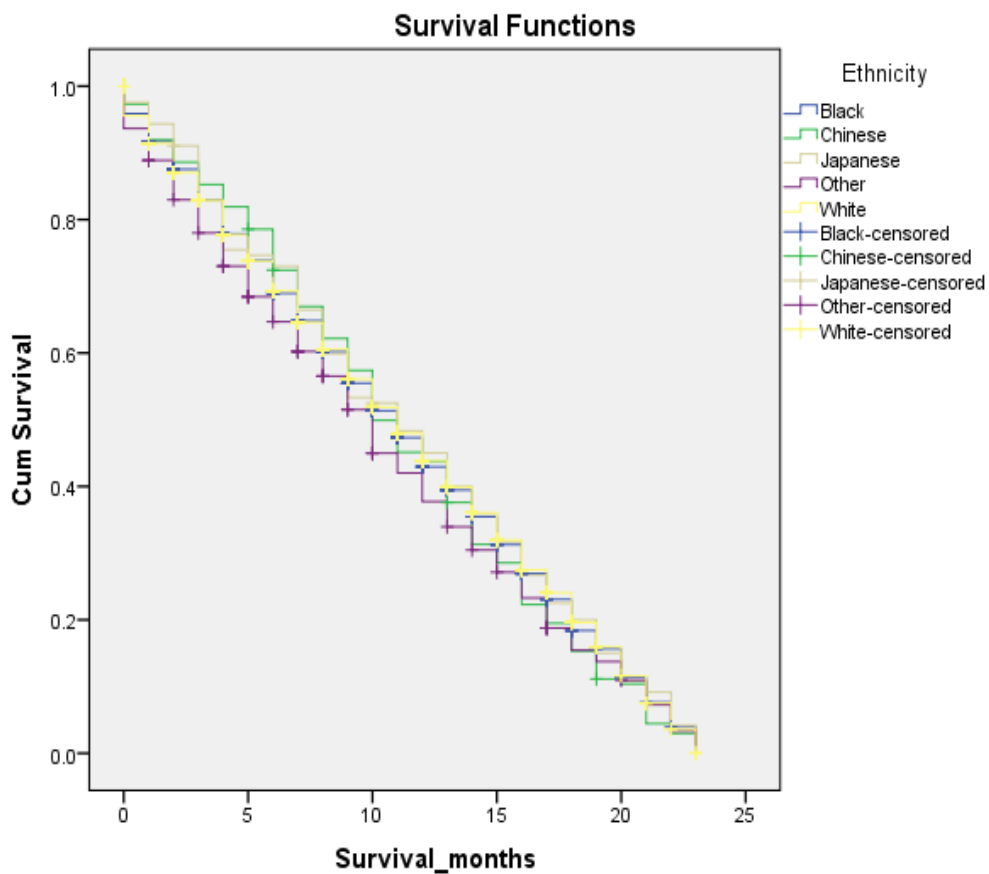


Figure 1. KM survival months for TNBC patients by Ethnicity.

Ethnicity	Age	Total N	N of Events	N of Censored
Black	<= 45	487	460	27
	> 45	1941	1763	178
	Overall	2428	2223	205
Chinese	<= 45	28	27	1
	> 45	122	118	4
	Overall	150	145	5
Japanese	<= 45	16	15	1
	> 45	108	106	2
	Overall	124	121	3
Other	<= 45	165	159	6
	> 45	483	455	28
	Overall	648	614	34
White	<= 45	1512	1459	53
	> 45	7499	6936	563
	Overall	9011	8395	616
Overall	Overall	12361	11498	863

Table 2. KM Case Processing Summary data for TNBC patients based on Ethnicity and Age groups. (N of Events=survived, N of Censored=died)

3.3 Survival Rates of the TNBC patients based on Ethnicity and Grade

Considering the ethnicity and grade (Table 3), we found no statistical significance in the difference between black and white TNBC patients with respect to grade 2 (OR=1.23, P=0.31). However, for Grade 3, black patient had slightly higher odds for survival than the white patient (OR=1.27, P=0.01). Grade 1 and 4 were not considered for analysis due to the low incidences for black and white ethnicities.

3.4 Survival Rates of the TNBC based on Ethnicity and Five Year Survival

From Table 4, we can find out that the odds of survival for 5 or less years for the white TNBC patients are 1.3 times as high as the one for the black TNBC patients (OR=1.2566, P = 0.0066). The samples for the ethnicity group Chinese and Japanese were too low to consider in this analysis.

GradeN	Ethnicity	Total N	N of Events	N of Censored
Grade2	Black	426	391	35
	Chinese	22	22	0
	Japanese	27	27	0
	Other	125	122	3
	White	1546	1441	105
	Overall	2146	2003	143
Grade3	Black	1852	1692	160
	Chinese	121	116	5
	Japanese	84	81	3
	Other	470	443	27
	White	6800	6329	471
	Overall	9327	8661	666

Table 3. Partial KM Case Processing Summary data for TNBC patients based on Ethnicity and Grade groups (N of Events=survived, N of Censored=died)

Five Year Survival	Ethnicity	Total N	N of Events	N of Censored
LE5Year	Black	2428	2223	205
	Chinese	150	145	5
	Japanese	124	121	3
	Other	648	614	34
	White	9010	8394	616
	Overall	12360	11497	863
Overall	Overall	12361	11498	863

Table 4. KM Case Processing Summary data for TNBC patients based on Five Year Survival and Ethnicity groups (N of Events=survived, N of Censored=died)

3.5 Survival Rates of the TNBC patients based on Ethnicity and NPI Prognosis

In Table 5, the ethnicity groups were compared according to their NPI prognostic groups. For example, in the category of “Excellent Prognosis”, the odds of survival of the white patient was about 1.50 times as high as the one for the black patient (statistically significant, OR=1.51, P=0.0065). Likewise, in the category of “Good Prognosis”, the odds of survival of the white patient was 1.34 times as high as the one for the black patient (statistically significant, OR=1.34, P=0.0008). In the

category of “Moderate Prognosis”, the odds of survival of the white patient was slightly higher than the one for the black patient (statistically significant, OR=1.18, P=0.0003). However, in the category of “Poor Prognosis”, the odds of survival of the black patient was about 1.4 times as high as the one for the white patient (statistically significant, OR=1.4, P=<0.0001).

Comparing the Japanese and white patients, in the category of “Good Prognosis”, the odds of survival of the Japanese patient was 1.9 times as high as the one for the white patient. (Statistically significant, OR=1.9, P=0.009). In the “Moderate and Poor Prognosis” categories, the odds of survival for the White patient was slightly higher than that of the Japanese patient; however, the differences were not statistically significant ((OR=1.16, P=0.40), (OR=1.4, P=0.18), Moderate and Poor Prognosis, respectively).

Comparing the Japanese and black patients, we obtained very similar results as noted above in the case of Japanese and white patients. That is, Japanese patient had a higher odd of survival, for “Good Prognosis” category, than the one for black patient (OR=2.55, P=0.0003), but slightly lower odds of survival for “Moderate and Poor Prognosis” categories, than the one for black patient (the differences, however, were not statistically significant).

For the “Excellent Prognosis” category, the TNBC incidences for the Japanese were too small to be considered in the analysis.

Ethnicity	N P I					Total
	Excellent	Good	Moderate	Poor	Other	
Japanese	4	20	64	19	17	124
White	289	827	4990	1822	1083	9011
Black	52	170	1245	571	390	2428

Table 5. TNBC based on Ethnicity/NPI Prognosis differences according to Ethnicity groups

4 Conclusion and Future Direction

Breast cancer prognostication is a vital element for providing effective treatment for breast cancer patients. Existing studies suggest that TNBC patients tend to have worse prognosis compared to non-TNBC counterparts.

Our experimental results on TNBC patients and their odds of survival can be summarized as follows (in most cases, the samples for the Chinese and Japanese TNBC patients were too small to be considered in the analysis):

- For the ethnicity factor, the white patients had odds of survivability slightly better than the black patients.
- Regarding the ethnicity and NPI prognosis, white patients had higher odds than the black patients on the Excellent, and Good prognoses, whereas the black patients had better odds with the Poor prognosis.
- For the ethnicity and age factors, both black and white older patients have better odds of survival than the corresponding younger patients.

This study was designed to investigate the effects of the ethnicity prognostic factor on breast cancer survivability among TNBC patients utilizing population-based of NIH SEER data to confirm whether ethnicity has prognostic significance. The results do not conclusively show that one ethnicity group (White, Black, Chinese, and Japanese) is doing better than the other. However, the results definitely show that ethnicity plays important role on the overall survivability of the TNBC patients.

The disparity in the prognosis of TNBC based on ethnicity can't be fully explained as there are other factors, as suggested by current medical literature [29], such as age, food habit, work/occupational environment, genetics and family history, obesity, and an access to health care could influence the survivability rate. These factors were not available in the SEER dataset that we used for our experiments. Subsequently, additional experiments should be conducted based on these factors to determine their impact on the prognosis in connection with ethnicity. In addition, existing breast cancer prognostic tools (i.e., NPI) should be further evaluated to accommodate other prognostic factors (i.e., age and ethnicity, etc.) in determining the survivability rate of breast cancer patients more accurately. Factors can be ranked based on their significance on the overall survivability (using sensitivity analysis, feature/factor selection, principle component analysis etc.) and weighted according to their ranking. Based on the results of ethnicity and other factors on the significance on survivability, it is possible to consider generating a simple rule-based prognostic decision system. A typical rule in such system would be:

If TNBC patient is African-American
and age is between 50-60
and has tumor grade 2
and NPI score of xx
Then the probability of a five year survivability is yy%.

Finally, such extended prognostic tool/system need to be evaluated for its usefulness in a clinical practice environment.

References

- [1] American Cancer Society. Breast Cancer Facts & Figures. Retrieved from <http://www.cancer.org/>, 2016.
- [2] G. Agarwal, G. Nanda, P. Lal, A. Mishra, A. Agarwal, V. Agrawal, and N. Krishnani. Outcomes of Triple-Negative Breast Cancers (TNBC) Compared with Non-TNBC: Does the Survival Vary for All Stages? *World Journal of Surgery*, Volume 40, Issue 6, pp 1362–1372, June 2016.
- [3] C. Anders and LA Carey. Understanding and Treating Triple Negative Breast Cancer. *Oncology*. 22 (11): 1233–9; discussion 1239–40, 1243, 2008.
- [4] Hui Bian. Survival Analysis Using SPSS., [core.ecu.edu/cfe/StatisticalReserach/Survival Analysis Using SPSS.pdf](http://core.ecu.edu/cfe/StatisticalReserach/Survival%20Analysis%20Using%20SPSS.pdf).
- [5] P. Boyle. Triple-negative breast cancer: epidemiological considerations and recommendations. *Annals of Oncology*. Aug; Vol. 23, issue Suppl 6: vi7-12, 2012.
- [6] Breastcancer.org. U.S. Breast Cancer Statistics. http://www.breastcancer.org/symptoms/understand_bc/statistics
- [7] AM Brewster, M. Chavez-MacGregor, and P. Brown. Epidemiology, biology, and treatment of triple-negative breast cancer in women of African ancestry. *Lancet Oncol*;15(13): 625-34, Dec 2014.
- [8] Oliver Brouckaert, Hans Wildiers, Giuseppe Floris, and Patrick Neven. Update on triple-negative

- breast cancer: prognosis and management strategies. *International Journal of Women's Health* Vol:4 511–520, 2012.
- [9] CA Clarke, TH Keegan, J Yang, *et al.* Age-specific incidence of breast cancer subtypes: understanding the black-white crossover. *J Natl Cancer Inst*, Vol. 104, 1094-1101, 2012.
- [10] SJ Dawson, E. Provenzano, and C. Caldas. Triple negative breast cancers: Clinical and prognostic implications. *Eur J Cancer* Vol. 45: 27-40, 2009.
- [11] William D. Foulkes, Ian E. Smith, and Jorge S. Reis-Filho. Triple-negative breast cancer. *New England journal of medicine* Vol. 363, No.20: 1938-1948, 2016.
- [12] C. A Hudis and L. Gianni. Triple-Negative Breast Cancer: An Unmet Medical Need. *The Oncologist*. 16: –11, 2011.
- [13] D. Huo, F. Ikpat, A. Khrantsov, JM Dangou, R. Nanda, J. DignamB. Zhang, T. Grushko, C. Zhang, O. Oluwasola, D. Malaka, S. Malami, A. Odetunde, AO Adeoye, F. Iyare, A. Falusi, CM Perou, and OI Olopade. Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer. *J Clin Oncol* 27:4515–4521, 2009.
- [14] Kaplan_Meier, Kaplan-Meier survival curve, <https://www.medcalc.org/manual/kaplan-meier.php>
- [15] S. Kharya. Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease, *International Journal of Computer Science, Engineering and Information Technology (IJCSIT)*, Vol. 2, No. 2, 55-56, April 2012.
- [16] F. Lara-Medina, V. Pérez-Sánchez, D. Saavedra-Pérez, M. Blake-Cerda, C. Arce, D. Motola-Kuba, C. Villarreal-Garza, AM González-Angulo, E. Bargalló, AJL guilar, A. Mohar, and O. Arrieta. Triple-negative breast cancer in Hispanic patients: high prevalence, poor prognosis, and association with menopausal status, body mass index, and parity., *Cancer*, Aug 15; 117(16):3658–69,2011.
- [17] Brian D Lehmann, and Jennifer A. Pietenpol. Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes. *The Journal of pathology* 23:.2, 142-150, 2014.
- [18] C. Lin, SY Chien, LS Chen, SJ Kuo, TW Chang, and DR Chen. Triple negative breast carcinoma is a prognostic factor in Taiwanese women [serial online]. *BMC Cancer*, 9:192, 2009.
- [19] MJ Lund, EN Butler, BY Hair, *et al.* Age/race differences in HER2 testing and in incidence rates for breast cancer triple subtypes: a population-based study and first report. *Cancer*, 116: 2549-2559, 2010.
- [20] NPI, https://en.wikipedia.org/wiki/Nottingham_Prognostic_Index
- [21] L. Ohno-Machado. *Modeling medical prognosis: survival analysis techniques*. *J Biomed Inform*; 34:428–39, 2001.
- [22] AA Onitilo, JM Engel, RT Greenlee, and BN Mukesh. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clin Med Res.*, 7:4-13, 2009.
- [23] Tanja Ovcaricek, Snjezana Grazio Frkovic, Erika Matos, Barbara Mozina, and Simona Borstnar, Triple negative breast cancer – prognostic factors and survival; *Radiol. Oncol.* Mar; 45(1): 46–52, 2011.
- [24] JS Reis-Filho, and AN Tutt. Triple negative tumours: a critical review. *Histopathology*. Jan; 52(1):108–18, 2008.
- [25] SEER, <https://seer.cancer.gov/data/options.html>
- [26] R. Siegel, D. Naishadham, and A. Jamal. *Cancer Statistics, CA Cancer J Clin*, Vol. 62, No. 10, 2012.
- [27] A. Stark, CG Kleer, I Martin, *et al.* African ancestry and higher prevalence of triple-negative breast cancer: findings from an international study, *Cancer*, 116, pp. 4926-4932, 2010.
- [28] LA Stead, TL Lash, JE Sobieraj, et al. Triple-negative breast cancers are increased in black women regardless of age or body mass index [serial online]. *Breast Cancer Res.*, 11: R18, 2009.

- [29] Khurram Tariq, Arezzo Farhangi, and Fauzia Ran. TNBC vs Non-TNBC: A Retrospective Review of Differences in Mean Age, Family History, Smoking History, and Stage at Diagnosis. *Clinical Advances in Hematology & Oncology* Volume 12, Issue 6 June, 377-381, 2014.
- [30] Lindsey A. Torre, Rebecca L. Siegel, Elizabeth M. War, and Ahmedin Jemal, Global Cancer Incidence and Mortality Rates and Trends—An Update, Online, December 14, 2015.
- [31] WJ Yin, JS Lu, GH Di, YP Lin, LH Zhou, GY Liu, J Wu, KW Shen, QX Han, ZZ Shen, and ZM Shao. Clinicopathological features of the triple-negative tumors in Chinese breast cancer patients. *Breast Cancer Res Treat* 115:325–333, 2009.