Proceedings of ICEGEE 2026 Symposium: AI-Based Medicine and Biological Data Analysis
DOI: 10.54254/2753-8818/2026.AU31888

Survival Analysis of Breast Cancer Patients: A Statistical Study Based on the SEER Database

Songyanqiu Guo

College of Science, Purdue University, West Lafayette, USA
rachel03.guo@gmail.com

Abstract. In recent times, breast cancer has become a global healthcare challenge, whereby women of all ages have been diagnosed with the dangerous disease. Further, breast cancer is understood and seen as complex due to its relatively high mortality rates. This study's objective was to analyze and reveal the long-term patterns in breast cancer occurrences and the prevailing mortality rates, as well as survival rates across the US. A quantitative research design was applied while also using data from SEER website and using both descriptive and regression methods. IBM SPSS version 26 software was specifically used to complete the statistical analysis. The results indicated that there was a gradual increase in breast cancer occurrences recently. Besides, this was followed by a major decline in mortality rates with a slower improvement in the 5-year survival rates. These findings further indicate that early diagnosis and treatment greatly improve population-level survival outcomes among breast cancer patients. The study implies the need for biostatistics research and analysis in understanding survival rates, hence aiding in informed public healthcare decisions.

Keywords: Breast Cancer, Survival Analysis, SEER Database, Biostatistics

1. Introduction

In recent decades, cancer has been a global public healthcare problem that affects many individuals, cutting across many healthcare systems. Particularly, breast cancer is the most prevalent malignancy reported among women and still remains the leading cause of increased mortality rates worldwide [1]. According to the recent data from the cancer surveillance databases, breast cancer significantly contributes to new cases each and every year, which reflects on the demographic shifts and improvements in screening practices [2]. The majority of the breast cancer patients are from vulnerable populations, especially in low- and middle-income countries (LMC's) where social and demographic factors pose a significant challenge since cancer is costly when it comes to diagnosis and treatment [3]. Despite the body of growing research and advancements in early detection and treatment, this disease continues to impose a significantly higher clinical and economic burden on affected individuals as well as public healthcare systems. Additionally, the heterogeneity of this disease when it comes to biological behaviors, clinical practices and prognosis further imposes a major challenge as far as management and survival outcomes are concerned [4]. In this perspective, there is a dire need for researchers, healthcare professionals, the government and associated stakeholders to collaborate and come up with relevant measures and practices to minimize cases of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.AU31888

breast cancer. Biostatistics is a key part of research work and greatly contributes to important epidemiological data analysis when it comes to systematic methods particularly on a large-scale dimension. Furthermore, statistical methods are important because they facilitate the identification of temporal trends in incidence, mortality, and survival of breast cancer, which is needed in minimizing and controlling the incidence of this disease [5]. In such a case, statistical analysis is very important in supporting evidence-based clinical and also public health decision-making, which is mostly seen as an important practice. Additionally, population-based websites such as the Centre for Disease Control (CDC), the World Health Organization (WHO), surveillance, epidemiology and End Results (SEER) programs provide researchers and medical scientists with valuable opportunities that are critical in examining long term survival patterns among breast cancer patients [6]. This is also a key significant idea for both controlling and managing cancer strategies, especially when dealing with vulnerable populations. The primary goal of this study was to explore and examine global trends in breast cancer incidence, mortality and survival rates among the most vulnerable individuals. Specifically, the paper utilizes data from the SEER database, which has critical information regarding breast cancer incidences, mortality and survival rates with age-specific criteria among other socio-economic demographics. At the same time, the paper sought to apply statistical trend analysis and methods to provide a comprehensive survival outcomes among breast cancer patients and future implications, which are critical for clinical and public healthcare outcomes. This innovation solely relies on the associated analysis of incidence, relative survival rates within the chosen rationale, while applying an advanced statistical analysis.

2. The epidemiology and clinical characteristics of breast cancer

2.1. The global and U.S. trends in breast cancer incidence rates

The United States is among the leading countries according to recent statistics, where there are prevalent cases of breast cancer as a result of genetics, lifestyle habits, and other factors. Breast cancer incidence rates have significantly demonstrated an upward trend, especially in the US [7]. At the same time, data from the SEER website indicates that there exists an increase in breast cancer incidence in recent decades, but it is accompanied by periods of stabilization due to recent advancements in early detection and treatment methods [4]. However, mortality rates have been reported to show a gradual decline, which is typically associated with early detection and therapeutic technologies. The differences in incidence and mortality rates are also evident across all vulnerable populations, which have been impacted by age, race, socioeconomic status, and access to improved healthcare services.

2.2. Clinical staging and treatment for breast cancer

Clinical staging of breast cancer is a key factor for breast cancer prognosis [8]. Additionally, the early-stage breast cancer diagnosis is mainly related with accelerated long-term survival outcomes, while advanced-stage disease continues to accelerate poor prognosis [9]. The most common standard treatment methods include surgery, radiotherapy, and chemotherapy [10]. However, the preferred choice of treatment is determined by tumor stage, genetics, and other underlying biological factors among patients. Furthermore, the recent improvements when it comes to treatment methods have significantly paved the way to improved survival rates among breast cancer patients across all healthcare disparities in the United States.

DOI: 10.54254/2753-8818/2026.AU31888

3. Survival analysis methods

3.1. Overview of research methods

The present study employed a quantitative research design based on secondary data analysis from SEER database. Both descriptive and multivariate regression methods were utilized to examine the extent to which longitudinal trends in breast cancer incidences, mortality, and 5-year relative survival rates have changed over time in the United States. Specifically, the descriptive statistics were used to summarize central tendencies and measures of variability according to recent data. On the other hand, inferential statistics where regression-based trend analysis were used to assess the changes in breast cancer rates over time. The statistical analysis involved the IBM statistical software version 26 to derive the required output based on the research topic.

3.2. Univariate and multivariate analysis

The univariate analysis is commonly utilized in analyzing the temporal patterns in incidence rates, mortality rates, and relative survival independently based on a disease that poses a major challenge in the healthcare systems. Furthermore, the multivariate regression models are co-currently utilized to analyze the relationship between calendar year and survival outcomes by linking any major or prevalent temporal effects. Despite incomplete individual-level covariates such as age or treatment methods from the SEER database, the survival trends are unique and provide important ideas and patterns into breast cancer disease outcomes.

4. Breast cancer analysis based on the SEER database

4.1. Data sources and variable description

The data utilized in this study were extracted from the Surveillance, Epidemiology, and End Results (SEER) database, which is a popular surveillance system administered by the National Cancer Institute in the United States. The dataset further consisted of aggregates and total annual statistics for breast cancer, which have been typically collected and recorded over multiple time periods. At the same time, the variables included calendar year, the breast cancer incidence rates, breast cancer mortality rates, age variable and five-year relative survival estimate as retrieved from the SEER database. All these indicators were key and significant in measuring and defining long term trends among affected individuals as opposed to individual patient records, which might cause discrepancies due to bias and self-reporting, hence causing erroneous statistical results as observed by Feleke et al. [11] The dataset was cleaned in Microsoft Excel version 2016 to remove outliers and missing values before initial statistical analysis.

4.2. Descriptive statistical results

Descriptive statistics were run in SPSS statistical software to come up with trends in breast cancer incidence, mortality, and survival over time. The mean breast cancer incidence rates for SEER 8 and SEER 12 were approximately (M=126.8, SD = 6.8), and (M=129.8, SD=4.9) cases. This showed the existence of a relatively stable but gradually increasing incidence in breast cancer. The mean mortality rate was M=26.8, SD=5.2, which is a downward trend. The mean five-year relative survival rate was given to be (M=83.6%, SD = 6.2), hence showing an improvement in the rates compared to previous years. These results are displayed in Table 1 below:

DOI: 10.54254/2753-8818/2026.AU31888

Table 1. Descriptive statistics of breast cancer indicators

	N	Minimum	Maximum	Mean	Std. Deviation
New_Cases_SEER8	48	101.68	143.76	126.7567	12.07528
New Cases_ SEER12	31	121.83	138.90	129.7784	4.94437
Deathrate	49	18.55	33.25	26.8169	5.23702
5_Year_Relative_ Survival (%)	43	75.24%	93.20%	86.2549%	6.22152%

4.3. Trend analysis of incidence rates

Linear regression model was carried out to analyze trends in breast cancer incidence rates. The results showed a statistically significant association between calendar year, death and incidence rates with $\beta = 3.578$, $p<0.022$. This finding depicted that breast cancer incidence increased gradually over the years. Typically, this factor was impacted by improved screening and diagnostic practices in the United States. Figure 1 below shows the trends in breast cancer incidence over time.

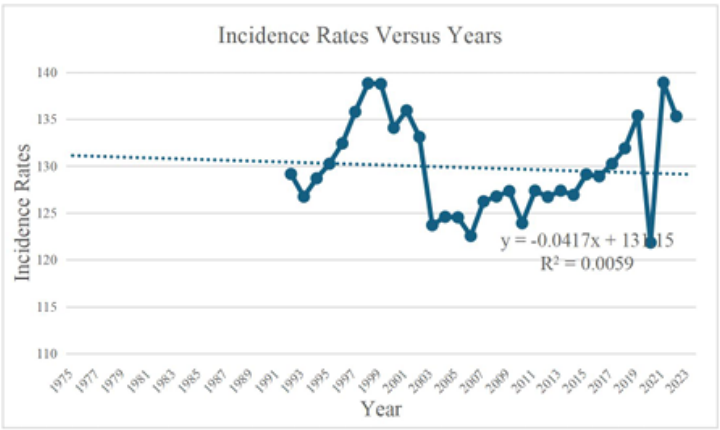

Figure 1. Trends in breast cbancer incidence rates [12]

4.4. Trend analysis of mortality rates

Regression analysis revealed a statistically significant positive association between the calendar year and breast cancer mortality rates with $\beta = -3.49$, $p < 0.000$. This finding was consistent, showing that there exists a decline in breast cancer mortality over time, which further shows how improvements in treatment methods and disease management have helped many patients [13]. These results are shown in Table 2 below:

Table 2. Linear regression results for breast cancer mortality trends

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	724.099	32.810		22.069	.000	658.093	790.104
	Year	-.349	.016	-.952	-21.252	.000	-.382	-.316

a. Dependent Variable: Deathrate

DOI: 10.54254/2753-8818/2026.AU31888

4.5. Trend analysis of five-year relative survival

The 5-year relative survival rates were analyzed using the linear regression models. The results indicated a strong positive association between calendar year and survival outcomes ($\beta = 0.470$, $p < 0.000$). This finding revealed that there exist consistent improvements in population-level survival among breast cancer patients for the sampled decades. The results are shown in Figure 2 below:

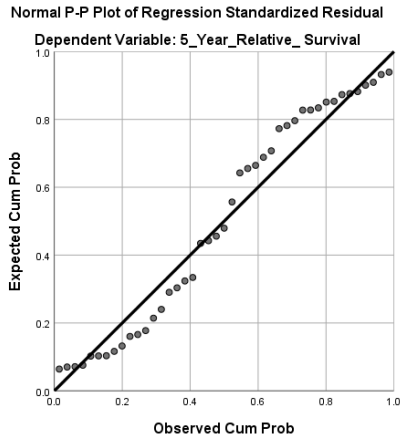

Figure 2. Five-year relative survival trends for breast cancer

5. Discussion

5.1. Factors affecting breast cancer survival

According to the findings of this study, the cancer stage (year prior to diagnosis) was found to be the strongest predictor or determinant of survival rates as documented by previous studies [6]. However, it is worth noting that age-related factors and differences in biological survival rates are key determinants in predicting comorbidities and tolerance of breast cancer. At the same time, the descriptive and inferential results highlighted the effectiveness of early diagnosis and treatment of breast cancer to increase survival rates among patients in the US. For instance, age-specific, race and period of diagnosis measured by years were strong predictors of breast cancer according to the analysis, which was consistent with Huang et al. [5] study, which found a statistically significant association between these variables. Furthermore, the study revealed that the majority of breast cancer survivors are likely to benefit when they are diagnosed on time before the disease escalates or progresses to late stages, which are complex to treat using existing methods.

5.2. Implications for clinical practice and public health

From the findings of the present study, there was a critical emphasis on the need to seek early diagnosis or detection where equitable access to effective treatment methods is important. Survival analysis further provided a chance for healthcare professionals, the government and other stakeholders with important information to support individualized care among breast cancer patients [14]. In this case, they can easily collaborate and come up with necessary intervention strategies to control and minimize the severity of breast cancer, which shortens patients' lifespan and their quality of life.

DOI: 10.54254/2753-8818/2026.AU31888

5.3. Value of biostatistics in medical decision-making

The current paper is important and will play an important role in the purpose of biostatistical methods in large-scale health data analysis. Debela et al [10] find that the incorporation of both descriptive and advanced inferential statistical methods is key in predicting breast cancer incidences, survival rates and mortality rates, which are important in healthcare decision making.

6. Conclusion

The current study analyzed all the long-term breast cancer survival outcomes using data from the Surveillance, Epidemiology, and End Results (SEER) database. The data extracted was analyzed using IBM SPSS software version 26. Both the descriptive and inferential statistical methods were used in determining the relationship between demographics of cancer patients, changes in incidence rates over time, mortality and a 5-year survival rate. The results indicated a statistically significant gradual increase in breast cancer incidence. Additionally, the results showed significant improvements in the 5-year relative survival rates across the selected decades. These findings were consistent with recent literature, which holds that early detection, diagnosis and treatment play an important role in improving survival outcomes among breast cancer patients. The analysis further supported the need for the implementation of advanced cancer treatment methods as well as collaboration between stakeholders to control the instances of severe breast cancer disease. Although the present study provides useful insights into breast cancer survival trends over time, it was limited in a number of ways. First, the data used was aggregated data, hence restricting the ability to assess specific prognostic factors for the chosen rationale. Secondly, the sample size utilized was small due to limited access to large datasets, hence reducing the statistical power. Future research should consider utilizing patient-level clinical and demographic factors as a way to guarantee a more detailed survival model stratification for all groups. In addition, including socio-economic factors such as employment, age, education attainment and healthcare access variables would further help to inform stakeholders on disparities in breast cancer and survival outcomes, especially in marginalized populations. Finally, the advancement in breast cancer research and the application of advanced biostatistical methods would further help healthcare systems curb the incidence of breast cancer, hence providing evidence-based public health interventions.

References

[1] Toffaha, K., Emre Simsekler, M. C., Shehhi, A. A., Sleptchenko, A., & AlAwadhi, A. (2025). Comprehensive survival analysis of breast cancer patients: a bayesian network approach. BMC Medical Informatics and Decision Making, 25(1). https: //doi.org/10.1186/s12911-025-03197-z

[2] Bazzi, T., Muneer Al-husseini, Saravolatz, L., & Zyad Kafri. (2023). Trends in Breast Cancer Incidence and Mortality in the United States From 2004-2018: A Surveillance, Epidemiology, and End Results (SEER)-Based Study. Cureus, 15(4). https: //doi.org/10.7759/cureus.37982

[3] Caswell-Jin, J. L., Sun, L. P., Munoz, D., Lu, Y., Li, Y., Huang, H., Hampton, J. M., Song, J., Jayasekera, J., Schechter, C., Alagoz, O., Stout, N. K., Trentham-Dietz, A., Lee, S. J., Huang, X., Mandelblatt, J. S., Berry, D. A., Kurian, A. W., & Plevritis, S. K. (2024). Analysis of Breast Cancer Mortality in the US—1975 to 2019. JAMA, 331(3), 233–241. https: //doi.org/10.1001/jama.2023.25881

[4] Dolatkhah, R., Somi, M. H., Jafarabadi, M. A., Hosseinalifam, M., Sepahi, S., Belalzadeh, M., Nezamdoust, M., & Dastgiri, S. (2020). Breast Cancer Survival and Incidence: 10 Years Cancer Registry Data in the Northwest, Iran. International Journal of Breast Cancer, 2020, e1963814. https: //doi.org/10.1155/2020/1963814

[5] Huang, J., Chan, P. S., Lok, V., Chen, X., Ding, H., Jin, Y., Yuan, J., Lao, X., Zheng, Z.-J., & Wong, M. C. (2021). Global incidence and mortality of breast cancer: a trend analysis. Aging, 13(4). https: //doi.org/10.18632/aging.202502

DOI: 10.54254/2753-8818/2026.AU31888

[6] Davis, E.A. and Joe, B.N. (2025) 'Rising Rates of Distant-Stage Breast Cancer: Trends among U.S. women, ' Radiology Imaging Cancer, 7(3), p. e259012. https: //doi.org/10.1148/rycan.259012.
[7] National Cancer Institute. (2024). Female Breast Cancer - Cancer Stat Facts. SEER. https: //seer.cancer.gov/statfacts/html/breast.html
[8] McCormack, V., McKenzie, F., Foerster, M., Zietsman, A., Galukande, M., Adisa, C., Anele, A., Parham, G., Pinder, L. F., Cubasch, H., Joffe, M., Beaney, T., Quaresma, M., Togawa, K., Abedi-Ardekani, B., Anderson, B. O., Schüz, J., & dos-Santos-Silva, I. (2020). Breast cancer survival and survival gap apportionment in sub-Saharan Africa (ABC-DO): a prospective cohort study. The Lancet Global Health, 8(9), e1203–e1212. https: //doi.org/10.1016/S2214-109X(20)30261-8
[9] Kaur, R., Bhardwaj, A. and Gupta, S. (2023) 'Cancer treatment therapies: traditional to modern approaches to combat cancers, ' Molecular Biology Reports, 50(11), pp. 9663–9676. https: //doi.org/10.1007/s11033-023-08809-3.
[10] Debela, D. T., Muzazu, S. G., Heraro, K. D., Ndalama, M. T., Mesele, B. W., Haile, D. C., Kitui, S. K., & Manyazewal, T. (2021). New Approaches and Procedures for Cancer treatment: Current Perspectives. SAGE Open Medicine, 9(9), 205031212110343. https: //doi.org/10.1177/20503121211034366
[11] Feleke, B., Tesfaw, L.M. and Mitku, A.A. (2022) 'Survival analysis of women breast cancer patients in Northwest Amhara, Ethiopia, ' Frontiers in Oncology, 12, p. 1041245. https: //doi.org/10.3389/fonc.2022.1041245.
[12] National Cancer Institute. (2025, May 15). Types of Cancer Treatment. National Cancer Institute. https: //www.cancer.gov/about-cancer/treatment/types
[13] Tang, D.-D., Ye, Z.-J., Liu, W.-W., Wu, J., Tan, J.-Y., Zhang, Y., Xu, Q., & Xiang, Y.-B. (2024). Survival feature and trend of female breast cancer: A comprehensive review of survival analysis from cancer registration data. The Breast, 79, 103862. https: //doi.org/10.1016/j.breast.2024.103862
[14] Noone, A.-M., Lund, J. L., Mariotto, A., Cronin, K., McNeel, T., Deapen, D., & Warren, J. L. (2016). Comparison of SEER Treatment Data with Medicare Claims. Medical Care, 54(9), e55–e64. https: //doi.org/10.1097/mlr.000000000000007