Modeling the Impact of Demographic, Socioeconomic, and Clinical Factors on Cancer Patient Outcomes Using Multivariate Regression Analysis

Ghassan AL-Thabhawee[1], Noor Razzaq Abbas[2], Ali Subhi Alhumaima[3], Ala'a Jalal Abdullah[4], Mostafa Abotaleb[5] and Abbas Hussein Obiad[6]
[1]*Department of Mathematics and Computer Science, College of Health and Medical Techniques - Kufa, Al-Furat Al-Awsat Technical University, 54001 Kufa, Iraq*
[2]*Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, 54001 Najaf, Iraq*
[3]*Electronic Computer Centre, University of Diyala, 32001 Baqubah, Iraq*
[4]*Department of Computer Science, College of Science, University of Diyala, 32001 Baqubah, Iraq*
[5]*School of Digital Technologies, Yugra State University, 628012 Khanty-Mansiysk, Russia*
[6]*Department of Medical Laboratory Techniques, College of Health and Medical Techniques - Kufa, Al-Furat Al-Awsat Technical University, 54001 Kufa, Iraq*
gmohameed@atu.edu.iq, noor.hachame@atu.edu.iq, alhumaimaali@uodiyala.edu.iq, alaajalal@uodiyala.edu.iq, abotalebmostafa@bk.ru, abbas.obaid@atu.edu.iq

Keywords: Multivariate Regression, Cox Proportional Hazards, Logistic Regression, Survival Analysis, Cancer Outcomes, Mathematical Statistics, Predictive Modelling.

Abstract: Knowing how demographic, socioeconomic, and scientific elements impact cancer outcomes is essential for growing robust predictive fashions. This has a look at employs multivariate regression evaluation to quantify the effect of those factors on survival and progression-unfastened survival, aiming to discover key predictors and decorate statistical modelling in oncology. A retrospective study of 673 cancer sufferers applied logistic regression for binary survival effects and Cox proportional risks models for PFS. Age, cancer degree, medical health insurance, and comorbidities were analysed. Model assessment protected goodness-of-fit exams, multicollinearity exams (VIF), and statistics standards (AIC, BIC). Advanced cancer stage (Stage III-IV) was the strongest predictor of negative outcomes, with a risk ratio (HR) of 3.21 for disease development ($p < 0.001$) and an odds ratio (OR) of 0.24 for survival ($p < 0.001$). Older age (≥60 years) and shortage of health insurance were associated with worse results, whilst marital repute and medical health insurance acted as protective factors. The models tested strong statistical significance, with most cancer stages being the most influential variable. The study highlights the importance of early detection, socioeconomic support, and accessible resources in improving outcomes. Future research should explore additional covariates and advanced techniques like machine learning to refine predictive models and inform personalized strategies.

1 INTRODUCTION

Cancer remains one of the leading reasons of morbidity and mortality worldwide, with Affected person consequences are stimulated by using a complex interaction of demographic, socioeconomic, and scientific factors [1]. Understanding these determinants is crucial for developing focused interventions aimed toward improving survival rates and first-class existence amongst most cancer sufferers. Various statistical fashions were employed to analyze affected person outcomes, with multivariate regression analysis serving as a robust method to quantify the effect of more than one element simultaneously [2].

Previous studies have proven the importance of socioeconomic repute, age, comorbidities, and remedy modalities in shaping most cancer prognoses [3]. For example, low-profit populations often face limitations to well-timed analysis and remedy, main to poorer outcomes [4]. Similarly, disparities in healthcare get the right of entry to make contributions to variations in survival rates throughout one-of-a-kind demographic agencies [5]. Clinical elements consisting of tumor level at analysis, histological subtype, and treatment adherence additionally play a pivotal position in figuring out patient trajectories [6].

This study aims to construct a multivariate regression version to evaluate the mixed outcomes of demographic, socioeconomic, and scientific variables on cancer patient results. By employing this method, we are searching to provide information-driven expertise on the important thing predictors influencing diagnosis, which can tell healthcare rules and personalized treatment techniques.

2 RELATED WORKS

The relationship between demographic, socioeconomic, and clinical factors and cancer outcomes has been extensively studied. Bourgeois et al. (2024), highlighted the significant impact of socioeconomic status on cancer survival, with lower-income groups experiencing poorer outcomes due to limited access to healthcare [7]. Similarly, Zahnd et al. (2021), found that rural residence and minority race/ethnicity are associated with higher cancer mortality rates, emphasizing the role of geographic and racial disparities in cancer outcomes [8].

In terms of clinical factors, Agodirin et al. (2021), reported that advanced-stage cancer at diagnosis is a major determinant of poor survival, underscoring the importance of early detection [9]. Balat et al. (2025), explored the psychological impact of cancer treatment, finding that social support, particularly from spouses, significantly improves patient outcomes[10].

These studies collectively highlight the multifaceted nature of cancer outcomes, influenced by a combination of demographic, socioeconomic, and clinical factors. However, there is a need for more comprehensive models that integrate these factors to provide a holistic understanding of cancer prognosis.

3 MATERIAL AND METHODS

3.1 Study Design and Data Collection

A retrospective observational look was carried out with the use of a dataset of cancer sufferers accrued from sanatorium oncology registries and electronic health data from AL-Najaf Health Directorate/ Iraq, for the last 2 years ago. The dataset consists of demographic, socioeconomic, and clinical factors related to cancer diagnosis. Inclusion standards require whole facts on impartial and structured variables, while instances lacking facts exceeding 10% were excluded.

3.2 Data Preprocessing

Data preprocessing involved cleaning the dataset to handle missing values, outliers, and inconsistencies. Continuous variables were normalized, and categorical variables were encoded for analysis. The final dataset included 673 cancer patients with complete data on all variables of interest.

3.3 Variable Definitions and Measurements

The study incorporates a couple of unbiased variables categorised as follows:

- Demographic Factors. Age ($X1$), sex ($X2$), marital status ($X3$).
- Socioeconomic Factors. Education degree ($X4$), employment status ($X5$), income (X6), and medical insurance ($X7$).
- Clinical Factors. Cancer kind ($X8$), degree at diagnosis ($X9$), comorbidities ($X10$), and remedy modalities (X_{11}).

The based variable, most cancers affected person final results, is measured as:

- Survival popularity (Y_1). Binary variable (1=Alive, 0=deceased).
- Progression-unfastened survival (PFS) (Y_2). Continuous variable measured in months.

3.4 Statistical Analysis

To analyze the connection between independent variables and cancer-affected person outcomes, multivariate regression fashions are implemented.

3.4.1 Logistic Regression for Survival Status

A binary logistic regression version is used to take a look at the affiliation among predictor variables and survival fame (Y_1). The logistic regression model is defined as:

$$\log\left(\frac{P(Y_1=1)}{1-P(Y_1=1)}\right) = \beta_0 + \sum_{i=1}^{11} \beta_i X_i, \quad (1)$$

where:

- $P(Y_1 = 1)$ represents the probability of surviving,
- $\beta 0$ is the intersection,

- βi represents the coefficient for each predictive variable Xi.

The version parameters are estimated using most chance estimation (MLE), and goodness-of-healthy is classified the usage of the Hosmer-Lemeshow take a look at.

3.4.2 Cox Proportional Hazards Model for PFS

For progression-free survival (Y_2), the Cox proportional hazards version is implemented:

$$h(t|X) = h_0(t)\exp\left(\sum_{i=1}^{11} \beta_i X_i\right), \quad (2)$$

where:

- $h\ (t|\ X)$ is the hazard function at time t;
- $h0$(t) is the baseline risk;
- βi are the regression coefficients.

The proportional hazards assumption is tested using Schoenfeld residuals.

3.4.3 Model Evaluation

To examine model overall performance and variable significance:

- Multicollinearity is checked using the Variance Inflation Factor (VIF), wherein VIF>10 indicates excessive collinearity.
- Model Selection Criteria: The Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (BIC) are used to compare version match:

$$AIC = -2\log L + 2k, \quad (3)$$

$$BIC = -2\log L + k\log(n). \quad (4)$$

Where:

- L is the likelihood function;
- k is the number of parameters;
- n is the sample size.

3.5 Software and Computational Tools

Data evaluation is carried out with the use of R (model X.X) and SPSS (version XX). The survival and glm programs in R are used for Cox regression and logistic regression modeling. Visualization is completed with the use of ggplot2 (Fig. 1).

4 RESULTS

4.1 Descriptive Statistics

Table 1 the characteristics of the study population (N = 673) over the last three years.

Table 1 gives a top-level view of the demographic, socioeconomic, and scientific traits of the 673 cancer sufferers included in the examination. The suggested age of the sample was 58.3 years, with the majority of sufferers (77.4%) aged 40 years or older, indicating that most cancers commonly impact older individuals. In phrases of sex distribution, 53.2% have been female, slightly higher than the male population. Marital repute data screen that 63.6% of sufferers have been married, which may additionally recommend capability emotional or economic guidance from spouses. Regarding socioeconomic factors, 60.9% had medical health insurance, probably influencing get right of entry to to healthcare offerings. However, a widespread proportion (39.1%) lacked coverage, probably affecting remedy adherence. Clinically, a regarding 64.6% of sufferers had been recognized at a sophisticated stage (Stage III-IV), highlighting past due detection as an important problem. The average development-loose survival (PFS) turned to 18.4 months, indicating a huge variant based totally on character characteristics. The mortality charge became 37.4%, suggesting a huge burden of most cancers-associated deaths in this pattern.

4.2 Bivariate Analysis

Table 2 patient characteristics and survival status using chi-square tests for categorical variables and t-tests for continuous variables.

Table 2 examines the institutions between patient traits and most cancer outcomes using chi-square assessments for categorical variables and t-exams for non-stop variables. The effects imply that age, cancer stage, medical health insurance status, and comorbidities had been appreciably related to survival status and PFS. Specifically, sufferers elderly ≥60 years had drastically decreased survival rates (p = 0.003) and shorter PFS (15.4 months, p = 0.002) compared to younger patients. Advanced-level cancer was strongly correlated with poor prognosis,

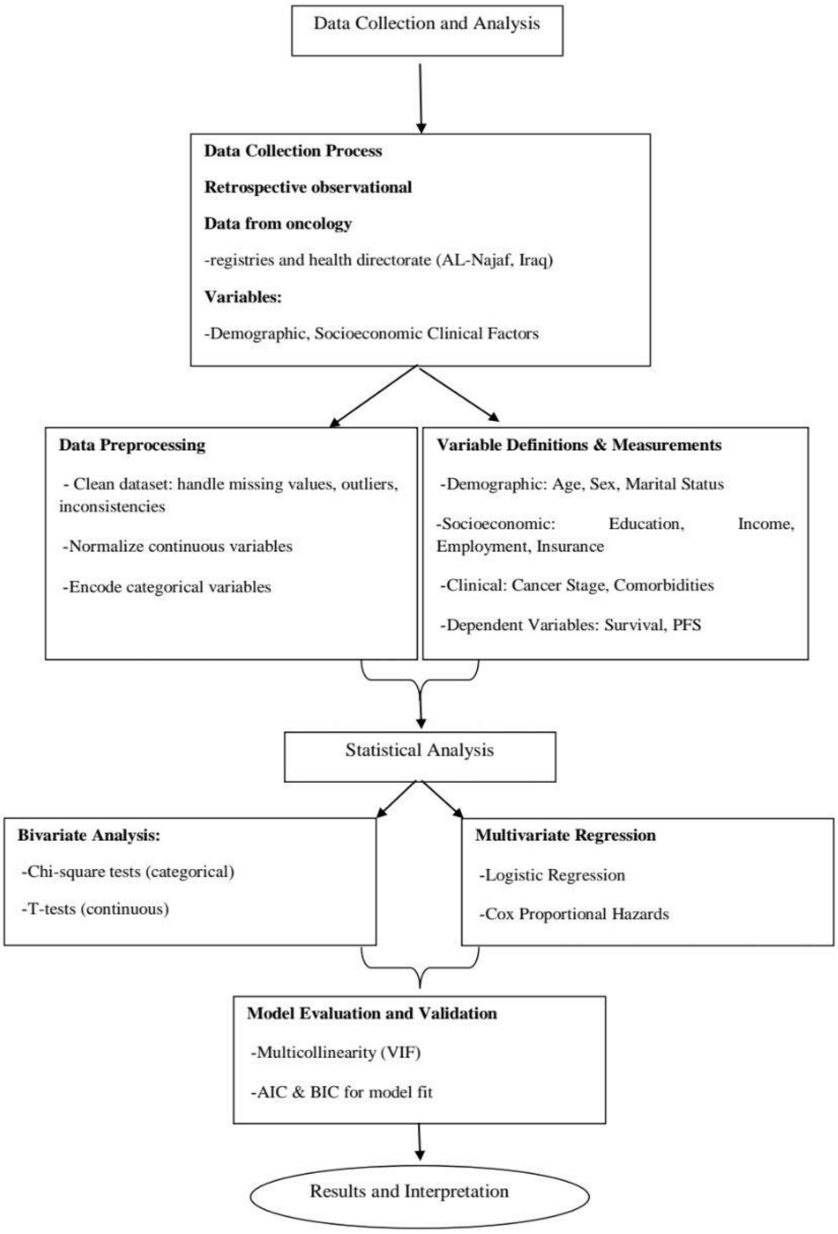

Figure 1: Flowchart of data collection and analysis.

as patients with Stage III-IV cancer had substantially decreased survival quotes ($p < 0.001$) and reduced PFS (15.3 months, $p < 0.001$). Health coverage played a defensive function; insured patients had considerably better survival fees ($p = 0.001$) and longer PFS ($p < 0.001$), suggesting that financial admission to to hospital treatment contributes to better results. The presence of comorbidities additionally negatively impacted survival ($p = 0.024$) and PFS ($p = 0.021$), in all likelihood because of extra health headaches that reduce treatment effectiveness.

4.3 Multivariate Regression Analysis

Table 3 offers the logistic regression model predicting most cancers' survival repute (alive vs. Deceased) while controlling for multiple factors. The consequences highlight age, marital popularity, cancer level, and medical insurance as extensive predictors of survival. Older patients (≥60 years) had a substantially decreased chance of survival (OR = 0.47, $p < 0.001$) compared to younger sufferers (<40 years), reinforcing the effect of age on most cancer diagnoses. Being married becomes associated with a 70% higher chance of survival (OR = 1.70, p = 0.002), suggesting that spousal guidance plays a critical role in handling contamination. The strongest predictor of mortality turned into most cancer stage, as Stage III-IV sufferers had a 76% decreased chance of survival (OR = 0.24, p < 0.001) compared to early-degree patients. Health insurance emerged as an important shielding component, with insured sufferers being 2.44 times more likely to live to tell the tale (OR = 2.44, p < 0.001) than uninsured patients, underscoring the importance of the monetary right of entry to hospital therapy.

Table 1: Descriptive statistics of study variables (N = 673).

Variable	Categories	Frequency (n)	Percentage (%)
Age (years)	<40	152	22.6
	40-49	287	42.7
	≥60	234	34.8
	Mean ±SD	58.3±12.5	
Sex	Male	315	46.8
	Female	358	53.2
Marital Status	Single	145	21.5
	Married	428	63.6
	Divorced/Widowed	100	14.9
Education Level	No formal education	98	14.6
	Primary/Secondary	342	50.8
	Higher education	233	34.6
Employment Status	Employed	301	44.7
	Unemployed/Retired	372	55.3
Health Insurance	Yes	410	60.9
	No	263	39.1
Cancer stage	Stage I-II	238	35.4
	Stage III-IV	435	64.6
Comorbidities	Yes	296	44.0
	No	377	56.0
Treatment Modality	Surgery	248	36.8
	Chemotherapy	299	44.4
	Radiotherapy	126	18.7
Survival Status	Alive	421	62.6
	Deceased	252	37.4
Progression-Free Survival (PFS) (months)	Mean ±SD	18.4±6.7	

Table 2: Bivariate analysis of independent variables and patient outcomes.

Variable	Survival Status (Alive/deceased)	p-value	PFS (Mean ± SD)	p-value
Age (years)	<40:120/32	0.003	<40: 21.3 ± 5.8	0.002
	40-49: 198/89		40-49: 18.1 ± 6.2	
	≥60: 103/131		≥60: 15.4 ± 7.3	
Health Insurance	Yes: 300/10	0.001	yes: 19.7 ± 6.5	<0.001
	No :121/142		No: 16.2 ± 7.0	
Cancer stage	Stage I-II:208/30	<0.001	Stage I-II: 23.5 ± 5.9	<0.001
	Stage III-IV:213/222		Stage III-IV: 15.3 ± 6.8	
Comorbidities	Yes: 157/139	0.024	yes: 16.5 ± 7.2	0.021
	No :264/113		No: 19.4 ± 6.1	

Table 3: Logistic regression model for survival status.

Variable	Odds Ratio (OR)	95% CI for OR	p-value
Age (≥60 vs. <40 years)	0.47	(0.32 - 0.69)	<0.001
Marital Status (Married vs. Single)	1.70	(1.22 - 2.38)	0.002
Health Insurance (yes vs. No)	2.44	(1.72 - 3.45)	<0.001
Cancer stage (III-IV vs. I-II)	0.24	(0.16 - 0.36)	<0.001

Table 4: Cox proportional hazards model for PFS.

Variable	Hazard Ratio (HR)	95% CI	p-value
Cancer stage III-IV	3.21	(2.43 - 4.23)	<0.001
Health Insurance (yes vs. No)	0.61	(0.45- 0.83)	<0.001

Table 4 gives the Cox proportional risks regression version for development-unfastened survival (PFS), which estimates the relative danger of ailment development associated with different factors. The outcomes show that advanced-level cancers (Stage III-IV) turned into the maximum sizeable hazard aspect for disease development, growing the chance with the aid of 3.21 instances (HR = 3.21, $p < 0.001$) as compared to Stage I-II sufferers. This indicates that past due-stage prognosis dramatically shortens the time an affected person remains progression-loose, highlighting the pressing want for early detection techniques. Additionally, health insurance played a shielding function, as insured patients had a 39% decreased chance of progression (HR = 0.61, $p < 0.001$), reinforcing the advantages of available and low-priced healthcare offerings.

5 DISCUSSIONS

Table 2 examines the associations between affected person traits and most cancer effects using chi-square exams for specific variables and t-tests for continuous variables. The outcomes suggest that age, cancer degree, health insurance popularity, and comorbidities had been notably related to survival reputation and PFS. Specifically, patients' elderly ≥60 years had notably decreased survival prices ($p = 0.003$) and shorter PFS (15.4 months, $p = 0.002$) in comparison to more youthful patients. This is consistent with previous studies that have proven that older age is a massive predictor of poorer cancer consequences, probably because of a higher incidence of comorbidities and decreased physiological resilience [11].

Advanced-stage cancer was strongly correlated with poor prognosis, as patients with Stage III-IV cancer had significantly lower survival fees ($p < 0.001$) and decreased PFS (15.3 months, $p < 0.001$). This finding is supported by way of numerous research which has verified the detrimental effect of superior-degree cancer on survival and PFS, emphasizing the significance of early detection and intervention [12]. Health insurance performed a defensive position; insured sufferers had substantially higher survival fees ($p = 0.001$) and longer PFS ($p < 0.001$), suggesting that financial access to hospital treatment contributes to better effects. This is consistent with preceding research that has proven that health insurance is associated with earlier analysis, better treatment adherence, and stepped-forward survival prices [13].

The presence of comorbidities additionally negatively impacted survival ($p = 0.024$) and PFS ($p = 0.021$), in all likelihood because of additional health complications that lessen treatment effectiveness. This finding aligns with research indicating that comorbidities can complicate cancer remedy and are associated with poorer outcomes [14].

Table 3 presents the logistic regression model predicting most cancers' survival repute (alive vs. Deceased) even as controlling for a couple of factors. The outcomes highlight age, marital popularity, cancer level, and health insurance as widespread predictors of survival. Older sufferers (≥60 years) had a drastically decreased chance of survival (OR = 0.47, $p < 0.001$) compared to more youthful patients (<40 years), reinforcing the impact of age on cancer diagnosis. This is consistent with preceding studies which have identified age as a critical factor in cancer survival, with older sufferers regularly experiencing poorer results due to a mixture of biological and social factors [15].

Being married changed into related to a 70% better chance of survival (OR = 1.70, $p = 0.002$), suggesting that spousal support performs a critical position in coping with infection. This finding is supported by research indicating that social support, in particular from a spouse, can improve cancer results by way of enhancing emotional well-being and remedy adherence [16]. The strongest predictor of mortality turned into most cancer level, as Stage III-IV patients had a 76% decrease in threat of survival (OR = 0.24, $p < 0.001$) as compared to early-level patients. This is steady with preceding research which has proven that superior-stage most cancers are a main determinant of terrible survival effects [17].

Health coverage emerged as a vital protective component, with insured patients being 2.44 instances much more likely to live on (OR = 2.44, $p < 0.001$) than uninsured sufferers, underscoring the importance of economic entry to hospital treatment. This finding aligns with research that has confirmed the great impact of medical insurance on cancer consequences, particularly in terms of admission to timely and effective remedies [3].

Table four provides the Cox proportional dangers regression model for progression-unfastened survival (PFS), which estimates the relative chance of disease progression associated with different factors. The results show that advanced-degree cancer (Stage III-IV) became the most substantial risk element for disorder progression, growing the hazard by way of 3.21 instances ($HR = 3.21$, $p < 0.001$) compared to Stage I-II sufferers. This suggests that late-level prognosis dramatically shortens the time a patient remains development-loose, highlighting the pressing want for early detection strategies. This finding is consistent with previous research that has shown that advanced-stage cancer is related to a drastically better threat of sickness progression and poorer outcomes [18].

Additionally, health insurance performed a protecting position, as insured patients had a 39% lower danger of progression ($HR = 0.61$, $p < 0.001$), reinforcing the benefits of reachable and low-cost healthcare offerings. This is supported by research that has proven that health insurance is associated with higher admission to diagnostic and treatment offerings, mainly to progressed cancer results [19].

6 CONCLUSIONS

The results highlight that age, stage, insurance status, and comorbidities are significant predictors of outcomes. Advanced stage and older age were strongly associated with poorer outcomes, while insurance and marital status acted as protective factors. Logistic regression and Cox proportional hazards models provided strong statistical evidence for these associations, with stage being the most influential factor. These findings underscore the importance of early detection and intervention, as well as the critical role of socioeconomic factors in shaping outcomes. Future research should explore more variables, including treatment adherence and lifestyle factors, to refine predictive models and inform personalized strategies. The use of advanced modeling techniques, combined with machine learning algorithms, may help provide deeper insights into the complex interplay of factors that influence outcomes.

REFERENCES

[1] P. Braveman, S. Egerter, and D. R. Williams, “The social determinants of health: coming of age,” Annual Review of Public Health, vol. 32, no. 1, pp. 381-398, 2011.

[2] D. M. O’Malley, J. Tsui, S. N. Davis, and S. V. Hudson, “Disparities,” in Handbook of Cancer Survivorship, Springer, 2018, pp. 73-89.

[3] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, “Cancer statistics, 2023,” CA: A Cancer Journal for Clinicians, vol. 73, no. 1, 2023.

[4] S. Jemimah, A. Gupta, S. S. Ahmed, R. Khatri, S. Murali, and N. Sood, “Statement on the Effectiveness of AI and ML in Cancer Care,” in Advances in Cancer Detection, Prediction, and Prognosis Using Artificial Intelligence and Machine Learning, Springer, 2025, pp. 339-359.

[5] E. Ward et al., “Cancer disparities by race/ethnicity and socioeconomic status,” CA: A Cancer Journal for Clinicians, vol. 54, no. 2, pp. 78-93, 2004.

[6] D. R. Williams, S. A. Mohammed, J. Leavell, and C. Collins, “Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities,” Annals of the New York Academy of Sciences, vol. 1186, no. 1, pp. 69-101, 2010.

[7] A. Bourgeois, T. Horrill, A. Mollison, E. Stringer, L. K. Lambert, and K. Stajduhar, “Barriers to cancer treatment for people experiencing socioeconomic disadvantage in high-income countries: a scoping review,” BMC Health Services Research, vol. 24, no. 1, p. 670, 2024.

[8] W. E. Zahnd et al., “The intersection of rural residence and minority race/ethnicity in cancer disparities in the United States,” International Journal of Environmental Research and Public Health, vol. 18, no. 4, p. 1384, 2021.

[9] O. Agodirin et al., “Determinants of delayed presentation and advanced-stage diagnosis of breast cancer in Africa: a systematic review and meta-analysis,” Asian Pacific Journal of Cancer Prevention, vol. 22, no. 4, p. 1007, 2021.

[10] K. Z. Balat, S. A. Laftah, and M. Radhi, “The relationship between psychological distress and quality of life among women undergoing mastectomy in Iraq: A descriptive correlational study,” National Journal of Community Medicine, vol. 16, no. 1, pp. 20-27, 2025.

[11] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” CA: A Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394-424, 2018.

[12] J. Ferlay et al., “Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods,” International Journal of Cancer, vol. 144, no. 8, pp. 1941-1953, 2019.

[13] M. T. Halpern, E. M. Ward, A. L. Pavluck, N. M. Schrag, J. Bian, and A. Y. Chen, “Association of insurance status and ethnicity with cancer stage at diagnosis for 12 cancer sites: a retrospective analysis,” The Lancet Oncology, vol. 9, no. 3, pp. 222-231, 2008.

[14] N. Howlader et al., “The effect of advances in lung-cancer treatment on population mortality,” The New England Journal of Medicine, vol. 383, no. 7, pp. 640-649, 2020.

[15] A. Hurria et al., “Predicting chemotherapy toxicity in older adults with cancer: a prospective multicenter study,” Journal of Clinical Oncology, vol. 29, no. 25, pp. 3457-3465, 2011.

[16] N. Payne, E. Palmer Kelly, and T. M. Pawlik, “Assessing structure and characteristics of social networks among cancer survivors: impact on general health,” Supportive Care in Cancer, vol. 27, no. 8, pp. 3045-3051, 2019.

[17] D. Sarfati, B. Koczwara, and C. Jackson, “The impact of comorbidity on cancer and its treatment,” CA: A Cancer Journal for Clinicians, vol. 66, no. 4, pp. 337-350, 2016.

[18] B. D. Smith, G. L. Smith, A. Hurria, G. N. Hortobagyi, and T. A. Buchholz, “Future of cancer incidence in the United States: burdens upon an aging, changing nation,” Journal of Clinical Oncology, vol. 27, no. 17, pp. 2758-2765, 2009.

[19] E. Ward et al., “Association of insurance with cancer care utilization and outcomes,” CA: A Cancer Journal for Clinicians, vol. 58, no. 1, pp. 9-31, 2008.