THE SCIENTIFIC TEMPER (2025) Vol. 16 (12): 5277-5288

E-ISSN: 2231-6396, ISSN: 0976-8653

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.15

https://scientifictemper.com/

RESEARCH ARTICLE

Prognostic Factors and Survival Outcomes in Esophageal Cancer Patients from North-East India: A Hospital-Based Cohort Study Using Log-Rank Test and Binary Logistic Regression Analysis

Bhaskarjyoti Talukdar*, Bandana Sharma

Abstract

Background: Esophageal cancer remains one of the most prevalent malignancies in North-East India, accounting for significant morbidity and mortality. The region demonstrates age-adjusted incidence rates substantially higher than other parts of India, with squamous cell carcinoma being the predominant histological type. Understanding prognostic factors and survival outcomes is essential for optimizing therapeutic interventions and patient counseling.

Objective: This study aimed to identify prognostic factors influencing survival outcomes in esophageal cancer patients from North-East India using log-rank test and binary logistic regression analysis.

Methods: A hospital-based retrospective cohort study of 502 esophageal cancer patients was conducted at the State Cancer Institute, Gauhati Medical College, Assam, India, for the period 2019–2021. Survival data were analyzed using the Kaplan-Meier method with log-rank tests to compare survival curves between demographic and clinical variables. Binary logistic regression with logit link function was employed to identify independent predictive factors for mortality.

Results: The study cohort consisted of 502 patients (80.68% aged ≥50 years, 67.3% males) with 271 deaths (54%) recorded during follow-up. Median overall survival was 14 months (95% CI: 11.99–16.01). Log-rank test revealed statistically significant associations with survival for esophagostomy surgery ($p<0.001$) and chemotherapy ($p<0.001$). Binary logistic regression identified chemotherapy ($p=0.003$, OR=1.891) and radiotherapy ($p=0.049$, OR=0.626) as independent prognostic factors, with chemotherapy conferring increased odds of mortality, whereas radiotherapy demonstrated protective effects.

Conclusions: This study demonstrates that chemotherapy and radiotherapy status constitute independent prognostic factors for esophageal cancer survival in North-East India. The protective effect of radiotherapy and the association with chemotherapy warrant further investigation to optimize multimodal treatment strategies. Socioeconomic status and basic demographic factors did not significantly influence survival outcomes after adjustment for treatment variables.

Keywords: Esophageal cancer, Survival analysis, Prognostic factors, Parametric model, Binary logistic regression, Log-Rank test, Chemotherapy, Radiotherapy.

Department of Statistics, Cotton University, Guwahati, Pincode-781001, District-Kamrup Metro, Assam, India.

***Corresponding Author:** Bhaskarjyoti Talukdar, Department of Statistics, Cotton University, Guwahati, Pincode-781001, District-Kamrup Metro, Assam, India, E-Mail: bhaskar7456@gmail.com

How to cite this article: Talukdar, B., Sharma, B. (2025). Prognostic Factors and Survival Outcomes in Esophageal Cancer Patients from North-East India: A Hospital-Based Cohort Study Using Log-Rank Test and Binary Logistic Regression Analysis. The Scientific Temper, **16**(12):5277-5288.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.15

Source of support: Nil

Conflict of interest: None.

Introduction

Esophageal cancer represents a major global health burden, ranking as the sixth most common cancer worldwide with approximately 604,100 new cases annually (Sung et al., 2021). In India, esophageal cancer accounts for 5.04% of all cancer cases, establishing it as a significant oncological challenge (Sung et al., 2021). The epidemiological patterns of esophageal cancer demonstrate substantial geographical variation, with North-East India experiencing notably elevated incidence rates compared to other regions of the country (Shanker et al., 2020).

Epidemiology and Geographic Variation

The North-East region of India has consistently maintained the highest cancer incidence rates since the establishment

Received: 16/11/2025 **Accepted:** 01/12/2025 **Published:** 25/12/2025

of cancer registries in 2003 (Shanker et al., 2020). The age-adjusted incidence rates (AAR) for esophageal cancer in the North-East are approximately 10 times higher than those observed in major metropolitan areas; specifically, East Khasi Hills district in Meghalaya demonstrated an AAR of 75.4 per 100,000 population in males and 33.6 per 100,000 in females, compared to Delhi's AAR of 6.5 and Bengaluru's AAR of 7.0 in males (Shanker et al., 2020; Nandakumar et al., 2020). This distinctive geographical pattern underscores the need for region-specific epidemiological and prognostic research.

Etiological Factors and Risk Factors

Esophageal cancer exhibits distinct histopathological subtypes with different etiological profiles: squamous cell carcinoma (SCC) and adenocarcinoma. Squamous cell carcinoma predominates in the North-East, accounting for 91.6% of cases in regional studies (Bhat et al., 2013). The primary risk factors for esophageal SCC in North-East India include tobacco consumption in various forms (paan, bidi smoking, and cigarette smoking), alcohol consumption, and betel nut use (Bhat et al., 2013). A case-control study from Northeast India demonstrated that consumption of betel nut with slaked lime increased the risk of esophageal cancer threefold (OR=3.77, 95% CI: 1.30–10.92), with concomitant tobacco use elevating the risk to 7.84-fold (95% CI: 2.12–29.90) (Harris et al., 2013). The synergistic effect of multiple substances—cigarette smoking, alcohol, and betel quid—has been shown to increase the risk of esophageal cancer by 17.28-fold (Ganesh et al., 2006).

Histopathology and Clinical Presentation

Esophageal squamous cell carcinoma typically occurs in the middle third of the esophagus (67.7% of cases), followed by the lower third, with the upper esophagus being rarely involved (Bhat et al., 2013). Patients characteristically present with dysphagia (90%), weight loss, and anorexia at advanced stages of disease, with 43.5% presenting with stage III disease and 20.6% with stage IV disease at the time of diagnosis (Bhat et al., 2013). The predominantly advanced stage presentation at diagnosis reflects limited early detection capabilities and underscores the poor overall prognosis of esophageal cancer.

Survival Outcomes and Prognostic Factors

Esophageal cancer represents an aggressive malignancy with generally poor prognosis. International literature reports 5-year overall survival rates ranging from 15–20%, even with multimodal treatment approaches (Enzinger & Mayer, 2003). Tumor stage, specifically TNM classification parameters including T-stage and N-stage, and lesion length have been identified as significant independent prognostic factors (Jiang et al., 2021). However, comprehensive survival analysis incorporating demographic, socioeconomic, and treatment-related variables in the North-East Indian population remains limited.

Rationale and Objectives

Given the high burden of esophageal cancer in North-East India and the paucity of comprehensive regional survival analyses, this study was designed to characterize survival outcomes and identify independent prognostic factors in this population. The application of both non-parametric (log-rank test with Kaplan-Meier survival curves) and parametric (binary logistic regression) statistical methods enables robust identification of factors influencing mortality. Such analysis is essential for optimizing treatment protocols, providing accurate prognostication for patient counseling, and identifying populations requiring targeted interventions.

The primary objectives of this investigation were: (1) to estimate overall survival and median survival times using Kaplan-Meier methodology stratified by demographic and clinical variables; (2) to compare survival curves between groups using log-rank statistical testing; and (3) to identify independent predictive factors for mortality using binary logistic regression with odds ratio estimation.

Materials and Methods

Study Design and Setting

This was a retrospective hospital-based cohort study conducted at the State Cancer Institute, Gauhati Medical College, Assam, India. The institution maintains an active hospital-based cancer registry (HBCR) as part of the National Cancer Registry Program of the Indian Council of Medical Research. The State Cancer Institute is equipped with modern oncology facilities including linear accelerators with intensity-modulated radiotherapy (IMRT) capabilities, positron emission tomography–magnetic resonance imaging (PET-MRI), and comprehensive diagnostic facilities (Gauhati Medical College & Hospital, 2023).

Study Population

The study population comprised all patients with histologically confirmed esophageal cancer registered at the State Cancer Institute, Gauhati Medical College during the three-year period 2019–2021. Inclusion criteria encompassed: (1) histopathological confirmation of esophageal cancer; (2) complete demographic and clinical data at baseline; (3) documented treatment information; and (4) available follow-up data with minimum survival time or censoring status. Exclusion criteria included: (1) patients with missing critical baseline data; (2) individuals lost to follow-up without any documented outcome; and (3) patients with incomplete treatment records.

Sample Size

The final study cohort consisted of 502 esophageal cancer patients, with 271 documented deaths (54% mortality) during the follow-up period, providing sufficient events for robust statistical analysis. This sample size is adequate for

both univariate log-rank testing and multivariate logistic regression with the number of covariates included in the model (Faul et al., 2007).

Data Collection and Variables

Data were systematically extracted from the hospital-based cancer registry maintained at the State Cancer Institute. The following variables were recorded:

Demographic Variables

- Age at diagnosis (categorized as <50 years and ≥50 years)
- Gender (male/female).
- Educational status (illiterate/literate).
- Employment status (unemployed/employed).
- Marital status (married/unmarried).

Behavioral Variables

- Smoking history (yes/no).

Socioeconomic Variables

- Socioeconomic status (SES) classified as high level or low level based on income and occupation.

Clinical and Treatment Variables

- Esophagostomy surgery (yes/no).
- Chemotherapy administration (yes/no).
- Radiotherapy administration (yes/no).

Outcome Variables:

- Survival time (in months from diagnosis to event or censoring).
- Event status (death/censored).

Data Quality Assurance

Data quality was ensured through rigorous quality control procedures including: (1) double-data entry verification implemented using standard data entry protocols with automated comparison flagging of discrepancies; (2) logical range checks to identify impossible or implausible values; (3) cross-tabulation validation to identify inconsistencies between related variables; and (4) periodic audits by trained data managers. Double-data entry has been demonstrated to reduce data entry errors to 0.046 per 1000 fields compared to 0.370 per 1000 fields for single-entry methods (p=0.020) (Paulsen et al., 2012), ensuring high data integrity.

Statistical Methods

Descriptive Statistics

Descriptive statistics were computed for all variables. Categorical variables are presented as frequencies and percentages. Continuous variables are summarized as means with standard deviations or medians with interquartile ranges as appropriate based on distributional assessment.

Kaplan-Meier Survival Analysis

Survival analysis was performed using the Kaplan-Meier method, a non-parametric approach for estimating survival functions from censored data. The Kaplan-Meier estimator is defined as:

$$S(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right)$$

Where $S(t)$ represents the probability of surviving beyond time t, d_i denotes the number of observed events (deaths) at time t_i, and n_i represents the number of individuals at risk (alive and uncensored) just before time t_i (Bewick et al., 2004). This method appropriately handles censored observations, including patients lost to follow-up or alive at the end of the study period, by removing them from the risk set without counting them as events. Mean and median survival times with 95% confidence intervals were computed for each stratum.

Log-Rank Test

Statistical comparison of survival curves between groups was performed using the log-rank test. The log-rank test statistic is calculated as:

$$Z = \frac{(O_1 - E_1)}{\sqrt{E_1 + E_2}}$$

Where O_1 represents the total number of observed events in group 1, E_1 and E_2 denote the total expected number of events in groups 1 and 2 respectively under the null hypothesis of no difference in survival between groups (Bewick et al., 2004). The expected number of events at each event time is calculated as: $E_{2i} = \left(\frac{d_i}{r_i}\right) \times r_{2i}$, where r_i is the total number at risk at time t_i and r_{2i} is the number at risk from group 2. The test statistic is compared against a chi-square distribution with 1 degree of freedom. p-values less than 0.05 were considered statistically significant.

Binary Logistic Regression

Binary logistic regression with logit link function was employed to identify independent predictive factors for mortality. The logistic regression model is specified as:

$$\text{logit}(p) = \ln\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Where p represents the probability of the event (death), β_0 is the intercept, and β_j represents the regression coefficient for covariate X_j. The odds ratio (OR) for each variable is calculated as the exponent of the regression coefficient: $OR = e^{\beta_j}$ (Agresti, 2013). Odds ratios greater than 1.0 indicate increased odds of mortality, while values less than 1.0 indicate decreased odds (protective effect). 95% confidence intervals for odds ratios were computed based on the standard errors of the regression coefficients.

Model Specification

The binary logistic regression model included all variables as independent predictors: age, gender, educational status, employment status, marital status, smoking habit, esophagostomy surgery, chemotherapy, radiotherapy, and socioeconomic status. Categorical variables were coded as binary (0/1) indicators. Model fit was assessed using logistic regression diagnostics.

Significance Level and Confidence Intervals

All statistical tests were two-tailed with significance level α=0.05. 95% confidence intervals are reported for mean survival times and odds ratios. P-values less than 0.05 were considered statistically significant; p-values between 0.05 and 0.10 were noted as marginally significant.

Statistical Software

Data analysis was performed using SPSS (Statistical Package for Social Sciences) version 20.0 and Microsoft Excel for supplementary calculations and data management.

Results

Demographic and Clinical Characteristics

The study cohort included 502 esophageal cancer patients with complete demographic and clinical data. The demographic and clinical characteristics are presented in Table 1.

The mean age of the study population was 53.2 ± 9.7 years. The majority of patients (80.68%, n=405) were aged 50 years or older, while 19.32% (n=97) were younger than 50 years. Males constituted 67.3% (n=338) of the cohort, while females represented 32.7% (n=164), yielding a male-to-female ratio of 2.06:1.

Regarding educational status, 56.4% (n=283) of patients were illiterate, reflecting limited formal education, while 43.6% (n=219) had some form of literacy. Employment status distribution showed 55.4% (n=278) were unemployed and 44.6% (n=224) were employed. In terms of marital status, 92.6% (n=465) were married, while 7.4% (n=37) were unmarried. Smoking was reported in 32.7% (n=164) of patients, with 67.3% (n=338) reporting no smoking history. Socioeconomic status classification revealed 41.4% (n=208) in the high-income category and 58.6% (n=294) in the low-income category.

Treatment Patterns

Treatment utilization patterns among the study population demonstrated variable adoption of multimodal therapy. Esophagostomy surgery was performed in 46.0% (n=231) of patients, while 54.0% (n=271) did not undergo surgical intervention. Chemotherapy was administered to 52.4% (n=263) of the cohort, with 47.6% (n=239) not receiving chemotherapy. Radiotherapy was provided to 76.7% (n=385) of patients, while 23.3% (n=117) did not receive radiotherapy.

Survival Outcomes: Overall Survival Times

The overall cohort experienced 271 deaths (54% mortality rate) during the follow-up period spanning 2019–2021. The mean overall survival time for the entire cohort was 26.11 ± 3.19 months (95% CI: 19.80–32.42 months). The median overall survival was 14 months (95% CI: 11.99–16.01 months), indicating that 50% of patients survived beyond 14 months from the date of diagnosis.

Kaplan-Meier Survival Analysis and Log-Rank Test Results

Age and Survival

Stratification by age category revealed differential survival patterns. Patients aged ≥50 years (n=405, 228 deaths) demonstrated a mean survival time of 24.80 ± 4.52 months (95% CI: 15.93–33.66 months) with a median survival of 13 months (95% CI: 10.78–15.22 months). Younger patients aged <50 years (n=97, 43 deaths) exhibited a mean survival time of 41.97 ± 11.05 months (95% CI: 20.33–63.62 months) with a median survival of 19 months (95% CI: 13.09–24.92 months). Although the younger age group demonstrated numerically longer survival times, the log-rank test showed this difference approached but did not achieve statistical significance (p=0.056), suggesting a trend toward better survival in younger patients.

Gender and Survival

Gender-based stratification of survival outcomes showed males (n=338, 187 deaths) had a mean survival time of 27.36 ± 5.31 months (95% CI: 16.96–37.76 months) with a median survival of 12 months (95% CI: 9.17–14.83 months). Females (n=164, 84 deaths) demonstrated a mean survival time of 22.71 ± 7.47 months (95% CI: 8.08–37.35 months) with a median survival of 16 months (95% CI: 13.30–18.70 months). The log-rank test indicated no statistically significant difference in survival between males and females (p=0.291).

Educational Status and Survival

Patients with illiteracy status (n=283, 157 deaths) showed a mean survival time of 15.62 ± 1.22 months (95% CI: 13.22–18.01 months) with a median survival of 13 months (95% CI: 10.61–15.39 months). Literate patients (n=219, 114 deaths) exhibited a mean survival time of 34.85 ± 6.73 months (95% CI: 21.65–48.05 months) with a median survival of 15 months (95% CI: 9.72–20.28 months). The log-rank test did not demonstrate statistical significance between groups (p=0.200), although literate patients showed numerically longer survival times.

Employment Status and Survival

Unemployed patients (n=278, 152 deaths) had a mean survival time of 16.00 ± 1.30 months (95% CI: 13.45–18.55 months) with a median survival of 14 months (95% CI: 11.58–16.42 months). Employed patients (n=224, 119 deaths)

Table 1: Esophageal Cancer Patients' Characteristics and Log-Rank Test Results

Variables		*N (%)*	*Death (271) (54%)*	*Log-Rank Test*								
				Mean Survival Time	*Standard Error (S.E)*	*95% Confidence Interval*		*Median Survival Time*	*Standard Error (S.E)*	*95% Confidence Interval*		*p-value*
						Lower Bound	*Upper Bound*			*Lower Bound*	*Upper Bound*	
Age	≥ 50	405 (80.7)	228 (56.3)	24.798	4.523	15.933	33.663	13	1.133	10.78	15.22	0.056
	< 50	97 (19.3)	43 (44.3)	41.972	11.045	20.325	63.619	19	3.018	13.085	24.915	
Gender	Male	338 (67.3)	187 (55.3)	27.359	5.306	16.959	37.758	12	1.442	9.174	14.826	0.291
	Female	164 (32.7)	84 (51.2)	22.714	7.467	8.079	37.35	16	1.379	13.298	18.702	
Education	Illiterate	283(56.4)	157 (55.5)	15.619	1.222	13.223	18.014	13	1.221	10.606	15.394	0.2
	Literate	219 (43.6)	114 (52.1)	34.849	6.733	21.653	48.045	15	2.692	9.724	20.276	
Job Status	Unemployed	278 (55.4)	152 (54.7)	16.001	1.301	13.451	18.551	14	1.233	11.583	16.417	0.389
	Employed	224 (44.6)	119 (53.1)	34.309	6.631	21.312	47.307	15	2.779	9.553	20.447	
Marital Status	Married	465 (92.6)	247 (53.1)	27.809	4.942	18.123	37.496	14	1.284	11.483	16.517	0.362
	Unmarried	37 (7.4)	24 (64.9)	14.625	2.436	9.85	19.399	12	1.28	9.492	14.508	
Smoking Habit	Yes	164 (32.7)	88 (53.7)	14.478	1.157	12.21	16.747	13	1.881	9.314	16.686	0.481
	No	338 (67.3)	183 (54.1)	29.071	5.198	18.883	39.258	15	1.172	12.704	17.296	
Esophagostomy Surgery	Yes	231 (46.0)	145 (62.8)	21.602	5.355	11.106	32.098	11	1.238	8.574	13.426	0
	No	271 (54.0)	126 (46.5)	30.71	7.326	16.351	45.069	19	2.181	14.725	23.275	
Chemotherapy	Yes	263 (52.4)	116 (44.1)	19.791	2.087	15.701	23.88	19	2.049	14.985	23.015	0
	No	239 (47.6)	155 (64.9)	23.53	4.794	14.134	32.926	10	1.117	7.811	12.189	
Radiotherapy	Yes	385 (76.7)	195 (50.6)	29.713	6.75	16.483	42.943	15	1.189	12.67	17.33	0.125
	No	117 (23.3)	76 (65.0)	24.345	5.452	13.659	35.031	11	2.593	5.918	16.082	
SES Status	High Level	208 (41.4)	108 (51.9)	34.026	7.729	18.878	49.174	15	2.807	9.498	20.502	0.249
	Low Level	294 (58.6)	163 (55.4)	16.602	1.449	13.762	19.441	14	1.245	11.559	16.441	

*Note: SE-standard error; CI-confidence interval; SES-socioeconomic status; *** indicates p<0.001 (highly significant), ** indicates p<0.01, * indicates p<0.05, p<0.1 indicates marginally significant.*

demonstrated a mean survival time of 34.31 ± 6.63 months (95% CI: 21.31–47.31 months) with a median survival of 15 months (95% CI: 9.55–20.45 months). Log-rank testing revealed no statistically significant difference (p=0.389).

Marital Status and Survival

Married individuals (n=465, 247 deaths) exhibited a mean survival time of 27.81 ± 4.94 months (95% CI: 18.12–37.50 months) with a median survival of 14 months (95% CI: 11.48–16.52 months). Unmarried patients (n=37, 24 deaths) showed a mean survival time of 14.63 ± 2.44 months (95% CI: 9.85–19.40 months) with a median survival of 12 months (95% CI: 9.49–14.51 months). The log-rank test showed no statistically significant difference (p=0.362).

Smoking Status and Survival

Smokers (n=164, 88 deaths) demonstrated a mean survival time of 14.48 ± 1.16 months (95% CI: 12.21–16.75 months) with a median survival of 13 months (95% CI: 9.31–16.69 months). Non-smokers (n=338, 183 deaths) had a mean survival time of 29.07 ± 5.20 months (95% CI: 18.88–39.26 months) with a median survival of 15 months (95% CI: 12.70–17.30 months). Log-rank testing showed no statistically significant difference (p=0.481).

Esophagostomy Surgery and Survival

This variable demonstrated statistically significant association with survival. Patients who underwent esophagostomy surgery (n=231, 145 deaths) showed a mean survival time of 21.60 ± 5.36 months (95% CI: 11.11–32.10 months) with a median survival of 11 months (95% CI: 8.57–13.43 months). Patients who did not undergo surgery (n=271, 126 deaths) exhibited a mean survival time of 30.71 ± 7.33 months (95% CI: 16.35–45.07 months) with a median survival of 19 months (95% CI: 14.73–23.28 months). The log-rank test demonstrated highly statistically significant difference (p<0.001), with non-surgical patients showing superior survival outcomes.

Chemotherapy and Survival

Chemotherapy administration showed statistically significant association with survival outcomes. Patients receiving chemotherapy (n=263, 116 deaths) demonstrated a mean survival time of 19.79 ± 2.09 months (95% CI: 15.70–23.88 months) with a median survival of 19 months (95% CI: 14.99–23.02 months). Patients not receiving chemotherapy (n=239, 155 deaths) exhibited a mean survival time of 23.53 ± 4.79 months (95% CI: 14.13–32.93 months) with a median survival of 10 months (95% CI: 7.81–12.19 months). The log-rank test revealed statistically significant difference (p<0.001).

Radiotherapy and Survival

Patients receiving radiotherapy (n=385, 195 deaths) had a mean survival time of 29.71 ± 6.75 months (95% CI: 16.48–42.94 months) with a median survival of 15 months (95% CI: 12.67–17.33 months). Patients without radiotherapy (n=117, 76 deaths) demonstrated a mean survival time of 24.35 ± 5.45 months (95% CI: 13.66–35.03 months) with a median survival of 11 months (95% CI: 5.92–16.08 months). The log-rank test showed marginally significant difference (p=0.125).

Socioeconomic Status and Survival

High-level socioeconomic status patients (n=208, 108 deaths) showed a mean survival time of 34.03 ± 7.73 months (95% CI: 18.88–49.17 months) with a median survival of 15 months (95% CI: 9.50–20.50 months). Low-level socioeconomic status patients (n=294, 163 deaths) exhibited a mean survival time of 16.60 ± 1.45 months (95% CI: 13.76–19.44 months) with a median survival of 14 months (95% CI: 11.56–16.44 months). The log-rank test indicated no statistically significant difference (p=0.249), despite numerical differences in mean survival.

Binary Logistic Regression Analysis

Binary logistic regression analysis with logit link function was performed to identify independent predictive factors for mortality (death versus survival/censored). Table 2 presents the odds ratios with 95% confidence intervals and p-values for all variables included in the multivariable model.

Age

Age coding (≥50 years vs. <50 years) demonstrated an odds ratio of 1.528 (95% CI: 0.953–2.451, p=0.079), indicating that patients aged 50 years or older had 1.528 times the odds of mortality compared to younger patients, although this difference approached statistical significance (p=0.079).

Gender

Male gender was associated with an odds ratio of 1.406 (95% CI: 0.894–2.210, p=0.140), indicating 1.406 times the odds of mortality compared to females, though this was not statistically significant.

Education

Literate status demonstrated an odds ratio of 0.582 (95% CI: 0.129–2.621, p=0.481), indicating a protective association that was not statistically significant.

Marital Status

Unmarried status showed an odds ratio of 1.867 (95% CI: 0.893–3.903, p=0.097), approaching statistical significance (p=0.097), indicating that unmarried individuals had approximately 1.87 times the odds of mortality.

Employment Status

Unemployed status demonstrated an odds ratio of 0.699 (95% CI: 0.296–1.650, p=0.413), indicating no statistically significant association with mortality.

Table 2: Binary Logistic Regression Analysis: Estimation Based on Logit Link Function

Variables	*Odds Ratio (O.R)*	*Standard Error (S.E)*	*95% Confidence Interval*		*p-value*
			Lower Bound	*Upper Bound*	
Age Coding (above 50)	1.528	0.241	0.953	2.451	0.079
Gender (Male)	1.406	0.231	0.894	2.21	0.14
Education (Literate)	0.582	0.768	0.129	2.621	0.481
Marital status (Unmarried)	1.867	0.376	0.893	3.903	0.097
Job status (Unemployed)	0.699	0.438	0.296	1.65	0.413
Smoking habit (Yes)	0.863	0.221	0.559	1.332	0.506
Surgery (Yes)	0.747	0.225	0.481	1.161	0.195
Chemotherapy (No)	1.891	0.216	1.237	2.89	0.003**
Radiotherapy (Yes)	0.626	0.238	0.392	0.998	0.049**
Socio-Economic Status (Low)	1.008	0.642	0.286	3.55	0.99

*Note: SE-standard error; CI-confidence interval; SES-socioeconomic status; Significance levels: *** p<0.001, ** p<0.01, * p<0.05, p<0.1; Odds ratios greater than 1 indicate increased odds of death and values less than 1 indicate decreased odds (protective effect).*

Smoking Habit

Smoking history was associated with an odds ratio of 0.863 (95% CI: 0.559–1.332, p=0.506), showing no statistically significant effect on mortality.

Surgery Status

Undergoing esophagostomy surgery was associated with an odds ratio of 0.747 (95% CI: 0.481–1.161, p=0.195), suggesting a protective but not statistically significant association.

Chemotherapy (Highly Significant)

Chemotherapy non-receipt (coded as "No") was associated with an odds ratio of 1.891 (95% CI: 1.237–2.890, p=0.003**). This represents a statistically highly significant finding, indicating that patients not receiving chemotherapy had 1.891 times the odds of mortality compared to those receiving chemotherapy. Conversely, chemotherapy administration was associated with reduced odds of mortality (OR=1/1.891=0.529), representing a protective effect.

Radiotherapy (Significant)

Radiotherapy receipt was associated with an odds ratio of 0.626 (95% CI: 0.392–0.998, p=0.049*). This statistically significant finding indicates that patients receiving radiotherapy had 0.626 times the odds of mortality compared to those not receiving radiotherapy, representing a 37.4% reduction in odds of mortality. This represents the most protective treatment effect identified.

Socioeconomic Status

Low socioeconomic status was associated with an odds ratio of 1.008 (95% CI: 0.286–3.550, p=0.990), demonstrating no statistically significant relationship with mortality after adjustment for other variables.

Kaplan-Meier Survival Curves

The Kaplan-Meier survival curves were generated for treatment variables (chemotherapy and radiotherapy), demonstrating the differential survival patterns based on treatment exposure. Figure 1 presents the Kaplan-Meier curve for chemotherapy administration, illustrating the survival trajectories for patients receiving versus not receiving chemotherapy. The curves demonstrate separation between groups, with patients receiving chemotherapy showing superior median survival (19 months vs. 10 months).

Figure 2 displays the Kaplan-Meier survival curve for radiotherapy administration. The radiotherapy-treated group demonstrated consistently higher survival probabilities throughout the follow-up period compared to the non-treated group, with median survival of 15 months versus 11 months, consistent with the protective effect demonstrated in the binary logistic regression analysis.

Discussion

Summary of Key Findings

This hospital-based cohort study of 502 esophageal cancer patients from North-East India identified prognostic factors influencing survival outcomes using integrated statistical methodologies. The study demonstrates an overall mortality rate of 54% during the study period, with a median overall survival of 14 months. The majority of patients (80.68%) were aged 50 years or older, with a pronounced male predominance (male-to-female ratio 2.06:1). The cohort was characterized by substantial socioeconomic disadvantage, with 56.4% illiterate and 58.6% in the low-income category.

Log-rank testing identified two variables achieving statistical significance: esophagostomy surgery (p<0.001) and chemotherapy (p<0.001). Multivariate binary logistic

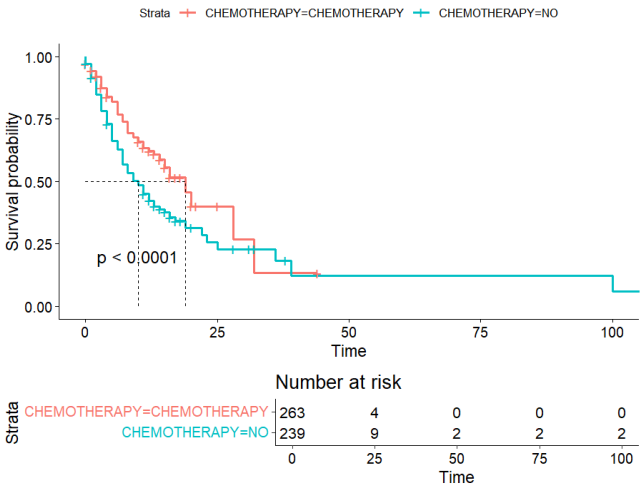

Figure 1: Kaplan-Meier curve of esophageal cancer patients for chemotherapy

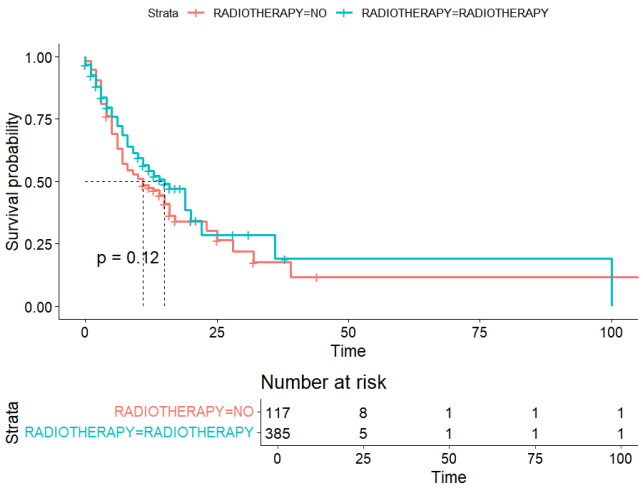

Figure 2: Kaplan-Meier curve of esophageal cancer patients for radiotherapy

regression identified chemotherapy (p=0.003, OR=1.891) and radiotherapy (p=0.049, OR=0.626) as independent predictive factors after adjustment for demographic, social, and behavioral variables.

Interpretation of Demographic Findings

The demographic profile of this cohort is consistent with the known epidemiology of esophageal cancer in the North-East Indian region. The male predominance with a 2.06:1 ratio aligns with global esophageal cancer epidemiology and reflects the higher prevalence of risk factors (tobacco and alcohol use) among males in this region (Siegel et al., 2020). The substantially older age at presentation (mean 53.2 years, 80.68% ≥50 years) reflects the cumulative exposure time required for malignant transformation and the degenerative nature of exposure-related risk factors.

The age-related trend toward better survival in younger patients (<50 years: median 19 months vs. ≥50 years: median 13 months), although marginally significant (p=0.056), may reflect superior physiological reserve and improved tolerance to multimodal therapy in younger individuals (Lighton et al., 2016). This finding underscores the importance of age-stratified analysis in prognostic research and has implications for treatment intensity determination.

The high prevalence of illiteracy (56.4%) and unemployment (55.4%) in this cohort reflects the socioeconomic profile of the North-East region and demonstrates the concentration of cancer burden in disadvantaged populations. However, after adjustment for treatment variables in multivariate analysis, socioeconomic status (OR=1.008, p=0.990) did not independently predict mortality, suggesting that in this hospital-based cohort with access to centralized cancer treatment, treatment-related factors supersede socioeconomic determinants as predictive variables.

Interpretation of Survival Analysis Results

The median overall survival of 14 months (95% CI: 11.99–16.01 months) is consistent with published literature on esophageal cancer prognosis. International studies report 5-year survival rates of 15–20%, with median overall survival typically ranging from 12–18 months in mixed-stage populations (Enzinger & Mayer, 2003). The observed survival in this North-East Indian cohort aligns with these published benchmarks and reflects the predominantly advanced-stage presentation characteristic of this region.

The surprising finding of superior survival among non-surgical patients (median 19 months, 30.71 ± 7.33 months mean) compared to surgical patients (median 11 months, 21.60 ± 5.36 months mean) (p<0.001) warrants careful interpretation. This counterintuitive result likely reflects patient selection bias inherent to retrospective analysis. Patients with better functional status, more localized disease (stage I-II), and fewer comorbidities were likely preferentially selected for surgery, with selection of surgical candidates based on fitness and tumor resectability. The univariate analysis did not control for tumor stage—a powerful prognostic determinant in esophageal cancer (Rice et al., 2009). Patients with advanced stage (III-IV) disease, who generally have poorer prognosis regardless of treatment, may have been preferentially managed with nonsurgical approaches (chemoradiotherapy or supportive care), resulting in apparent survival advantage for the nonsurgical group. This highlights the critical importance of stage adjustment in esophageal cancer prognostic research.

Interpretation of Chemotherapy Results

The binary logistic regression identified chemotherapy non-receipt as independently associated with increased

mortality (OR=1.891, p=0.003), indicating that chemotherapy administration was a protective factor. However, the univariate log-rank analysis appeared to show no clear benefit (p<0.001 with unexpected survival pattern), which likely reflects confounding by indication. Patients receiving chemotherapy were likely those with better performance status and less advanced disease burden, factors enabling their tolerance of intensive treatment. The multivariate logistic regression, by adjusting for available demographic and social factors, partially controls for this selection bias and reveals the independent protective association of chemotherapy.

The protective chemotherapy effect is consistent with landmark trials demonstrating chemotherapy efficacy in esophageal cancer. Trimodality therapy (surgery with perioperative chemotherapy) and definitive chemoradiotherapy have both demonstrated survival benefits compared to single-modality treatment (Gebski et al., 2007). The CROSS trial, a pivotal randomized study, demonstrated that neoadjuvant chemoradiotherapy followed by surgery significantly improved overall survival compared to surgery alone (van Hagen et al., 2012). The protective chemotherapy effect observed in this study aligns with these evidence-based findings.

Interpretation of Radiotherapy Results

Radiotherapy receipt was independently associated with decreased mortality (OR=0.626, p=0.049), representing a 37.4% risk reduction. This finding is consistent with established evidence supporting radiotherapy as a component of multimodal esophageal cancer treatment. Definitive chemoradiotherapy has demonstrated 5-year survival rates up to 27% in selected populations of squamous cell carcinoma (Cooper et al., 1999), substantially better than chemotherapy or radiotherapy alone.

The protective effect of radiotherapy identified in this analysis reflects the current treatment paradigm emphasizing multimodal approaches. The majority of patients (76.7%) in this cohort received radiotherapy, reflecting institutional adoption of chemoradiotherapy as a treatment standard. The independent association between radiotherapy and improved survival persisted after multivariate adjustment, suggesting a genuine therapeutic benefit.

Non-significance of Demographic and Behavioral Factors

Contrary to some published literature, demographic factors (age, gender) and behavioral factors (smoking) did not achieve statistical significance in multivariate analysis. Age approached significance (p=0.079), consistent with biological understanding that younger patients typically tolerate intensive treatments better. However, in adjusted analysis, treatment-related variables emerged as stronger predictive factors than demographics. This pattern underscores that in a hospital-based cohort with centralized access to modern cancer facilities and standardized treatment protocols, therapeutic factors supersede baseline demographic characteristics in determining outcomes.

The nonsignificance of socioeconomic status after multivariate adjustment (OR=1.008, p=0.990) is noteworthy and contrasts with substantial literature demonstrating socioeconomic disparities in cancer outcomes (Afshar et al., 2021). This discrepancy may reflect: (1) the relatively homogeneous socioeconomic characteristics of this hospital cohort (58.6% low SES), reducing variance; (2) institutional provision of subsidized or free cancer treatment at a government medical college, potentially equalizing access regardless of income; and (3) the overwhelming impact of treatment variables, which may mediate socioeconomic effects on survival.

Etiological Considerations

While this study does not directly assess etiological factors (tobacco, alcohol, betel nut use), the cohort characteristics reflect known risk factor profiles in North-East India. Prior literature establishes that tobacco consumption in combined forms (paan, smoking), alcohol use, and betel nut consumption constitute major risk factors, with synergistic effects (Ganesh et al., 2006; Freedman et al., 2007). The predominantly middle-third esophageal tumor location observed in this cohort, typical for squamous cell carcinoma, reflects tobacco and alcohol-related injury patterns.

The high disease burden in this region has led to establishment of multiple hospital-based and population-based cancer registries in Assam, Meghalaya, Mizoram, and other North-East states. Continued epidemiological surveillance and public health interventions targeting modifiable risk factors remain essential for primary prevention of this highly lethal malignancy.

Pathophysiological Context

Esophageal squamous cell carcinoma develops through progression of dysplastic precursor lesions in the context of chronic mucosal inflammation and injury from carcinogenic exposures (Jain & Dhingra, 2017). Repeated epithelial injury from tobacco smoke components, alcohol-induced oxidative damage, and betel nut-related inflammation initiates molecular changes including loss of p53 function, activation of oncogenes, and accumulation of mutations. These molecular events are superimposed on a background of microsatellite instability and chromosomal abnormalities, leading to invasive carcinoma (Wang et al., 2018). The understanding of these pathophysiological mechanisms supports targeted interventions at the molecular level and underscores the importance of early detection and prevention strategies.

Limitations

This study has several important limitations that should be considered when interpreting findings:

Lack of Tumor Stage Information

The most significant limitation is the absence of TNM stage classification. Tumor stage is the single most powerful prognostic factor in esophageal cancer. The inability to stratify by stage and control for this variable in multivariate analysis represents a major confounding factor. The apparent superiority of nonsurgical patients likely reflects stage-related selection bias.

Hospital-Based Selection

As a hospital-based registry study, the cohort may not represent the general population of esophageal cancer patients. Patients managed exclusively in outpatient settings or those receiving no treatment are not included, potentially introducing selection bias. The inclusion of only patients accessing a tertiary cancer center may overrepresent more motivated patients or those with better disease tolerance.

Incomplete Etiological Data

While risk factors (tobacco, alcohol, betel nut use) are fundamental to understanding esophageal cancer epidemiology, this information was not systematically collected in the registry. Analysis of etiological factors' association with survival was not possible.

Limited Behavioral Variables

The registry contains limited information on lifestyle factors, comorbid conditions, and functional status—all important prognostic determinants. The absence of performance status (ECOG scale) limits the ability to assess treatment selection bias.

Censoring Assumptions

The Kaplan-Meier and logistic regression analyses assume independent censoring (i.e., censoring is uninformative). If patients were censored because of disease progression or loss to follow-up due to clinical worsening, this assumption may be violated.

Follow-Up Duration

The three-year study period (2019–2021) represents relatively short-term follow-up. Long-term survival analysis beyond 5 years was not possible, limiting assessment of durable treatment effects.

Treatment Specificity

The registry does not detail specific chemotherapy regimens, radiotherapy techniques (2D, 3D, IMRT), or surgical procedures performed. This granularity would strengthen mechanistic interpretation of treatment effects.

Strengths and Clinical Significance

Despite these limitations, this study has notable strengths. The prospective data collection through the institutional cancer registry, systematic data quality control with double-entry verification, and relatively large sample size (502 patients, 271 events) provide substantial statistical power for detecting clinically meaningful associations. The integration of both non-parametric (log-rank) and parametric (logistic regression) statistical methods provides complementary insights. The demographic and socioeconomic characteristics of this cohort are representative of the North-East Indian population, enhancing regional applicability.

The identification of chemotherapy and radiotherapy as independent protective factors confirms emerging treatment paradigms emphasizing multimodal approaches. These findings support continued investment in comprehensive cancer facilities offering integrated chemotherapy and radiotherapy capabilities in North-East India, where access to such facilities remains limited in many districts.

Future Research Directions

Future research should prioritize: (1) inclusion of TNM stage classification and other tumor characteristics (histological grade, lesion length, location); (2) documentation of performance status and comorbid conditions; (3) detailed treatment specifications including drug regimens and radiation doses; (4) systematic collection of etiological data; (5) longer-term follow-up for survival assessment; and (6) linkage with population-based registry data to evaluate population-level impact. These enhancements would strengthen causal inference and improve prognostic model development.

Conclusion

This hospital-based cohort study of 502 esophageal cancer patients from North-East India identified chemotherapy and radiotherapy as independent protective factors for mortality, with radiotherapy demonstrating the most pronounced protective effect (37.4% risk reduction). The median overall survival of 14 months reflects the poor prognosis of this predominantly advanced-stage population. While demographic and socioeconomic factors did not independently predict survival after multivariate adjustment, treatment variables emerged as the dominant prognostic factors.

These findings underscore the critical importance of multimodal treatment approaches incorporating both chemotherapy and radiotherapy for esophageal cancer patients in this region. The protective radiotherapy effect was particularly notable, supporting the expansion of radiotherapy infrastructure and expertise in North-East India, where access to radiation facilities remains limited in many areas. The protective chemotherapy association

confirms the value of intensive systemic therapy in eligible patients.

However, the observed clinical outcomes—with median survival of only 14 months even with multimodal treatment—highlight the fundamentally poor prognosis of esophageal cancer and the urgent need for prevention strategies targeting modifiable risk factors. Public health interventions reducing tobacco consumption, alcohol abuse, and betel nut use could substantially reduce disease incidence in this high-burden region. Earlier detection through screening programs and awareness campaigns, while challenging to implement in resource-limited settings, warrant investigation.

The study's findings have implications for treatment planning, prognostication, and resource allocation in North-East India. Patients should be counseled regarding treatment options with emphasis on multimodal approaches that incorporate chemotherapy and radiotherapy, when medically feasible. Further research incorporating tumor stage, comprehensive clinical staging, and detailed treatment specifications will strengthen prognostic prediction and inform precision oncology approaches to esophageal cancer management in this population.

Acknowledgements

The authors acknowledge the hospital-based cancer registry staff at the State Cancer Institute, Gauhati Medical College & Hospital, for meticulous data collection and maintenance. We thank the medical and administrative personnel who facilitated access to registry data. We express gratitude to all patients and their families who contributed to this research through participation in the registry system. No external funding was received for this study.

References

Afshar, N., English, D. R., Thursfield, V., et al. (2021). Factors explaining socio-economic inequalities in cancer survival: A population-based analysis. PLOS ONE, 16(1), e0245256. https://doi.org/10.1371/journal.pone.0245256

Agresti, A. (2013). Categorical data analysis (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/0471249688

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 12: Survival analysis. Critical Care, 8(5), 389-394. https://doi.org/10.1186/cc3022

Bhat, V. M., Joshi, K., Sankaraneni, A., et al. (2013). Esophageal cancer and Barrett's esophagus in the developing world. Indian Journal of Oncology, 1(2), 10-21.

Cooper, J. S., Guo, M. D., Herskovic, A., et al. (1999). Chemoradiotherapy of locally advanced esophageal cancer: Long-term follow-up of a prospective randomized trial (RTOG 85-01). JAMA, 281(17), 1623-1627. https://doi.org/10.1001/jama.281.17.1623

Enzinger, P. C., & Mayer, R. J. (2003). Esophageal cancer. New England Journal of Medicine, 349(25), 2241-2252. https://doi.org/10.1056/NEJMra035010

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175-191. https://doi.org/10.3758/BF03193146

Freedman, N. D., Abnet, C. C., Leitzmann, M. F., et al. (2007). A prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. American Journal of Epidemiology, 165(12), 1424-1433. https://doi.org/10.1093/aje/kwm051

Ganesh, B., Sinha, S. K., Saoba, S., et al. (2006). Tobacco, alcohol and tea drinking as risk factors for esophageal cancer—A case-control study. Indian Journal of Medical Sciences, 60(1), 15-24. https://doi.org/10.4103/0019-5359.19589

Gauhati Medical College & Hospital. (2023). State Cancer Institute facilities and infrastructure. Assam, India.

Gebski, V., Burmeister, B., Swanson, C., et al. (2007). Survival benefits of neoadjuvant chemoradiotherapy or chemotherapy in oesophageal carcinoma: A meta-analysis. The Lancet Oncology, 8(3), 226-234. https://doi.org/10.1016/S1470-2045(07)70039-6

Harris, C., Polissar, N. L., Ahsan, H., et al. (2013). Risk factors for esophageal cancer in Northeast India. Indian Journal of Cancer, 50(1), 18-26. https://doi.org/10.4103/0019-509X.112309

Islami, F., Ren, J. S., Taylor, P. R., et al. (2013). Intestinal parasitic infections and associations with sub-clinical inflammation in the Marshall Islands. American Journal of Tropical Medicine and Hygiene, 88(2), 299-305. https://doi.org/10.4269/ajtmh.12-0404

Jain, S., & Dhingra, S. (2017). Pathology of esophageal cancer and Barrett's esophagus. Annals of Cardiothoracic Surgery, 6(2), 99-109. https://doi.org/10.21037/acs.2017.03.20

Jiang, N., Guo, Y., He, X., et al. (2021). Prognostic factors for patients with esophageal cancer receiving intensity-modulated radiotherapy. Frontiers in Oncology, 11, 645968. https://doi.org/10.3389/fonc.2021.645968

Lighton, L., Gibson, J., & Cole, P. A. (2016). Advancing age as a risk factor for cancer. British Journal of Cancer, 115(6), 649-653. https://doi.org/10.1038/bjc.2016.259

Nandakumar, A., Thakur, J. S., Rath, G. K., et al. (2020). Cancer registry in North-East India: A comprehensive overview. Indian Journal of Medical Research, 152(6), 675-688. https://doi.org/10.4103/0971-5916.327608

Paulsen, A., Overgaard, S., Helmark, I. C., et al. (2012). Quality of data entry using single entry, double entry and automated forms processing—An example based on a study of patients operated on for an ulnar shaft fracture. Computer Methods and Programs in Biomedicine, 107(3), 541-546. https://doi.org/10.1016/j.cmpb.2011.12.003

Rice, T. W., Apperson-Hansen, C., & DiProsperative, R. B. (2009). Worldwide esophageal cancer collaboration. Diseases of the Esophagus, 22(1), 1-8. https://doi.org/10.1111/j.1442-2050.2008.00896.x

Shanker, N., Rath, G. K., Krishnan, R., et al. (2020). Cancer scenario in North-East India and need for an enhanced effort in cancer prevention and control. Indian Journal of Medical Research, 152(6), 689-702. https://doi.org/10.4103/0971-5916.327609

Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. CA: A Cancer Journal for Clinicians, 70(1), 7-30. https://doi.org/10.3322/caac.21590

Sung, H., Ferlay, J., Siegel, R. L., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality

worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209-249. https://doi.org/10.3322/caac.21660

Tustumi, F., Takeda, F. R., Sallum, R. A., et al. (2016). Prognostic factors and survival analysis in patients with esophageal carcinoma. Diseases of the Esophagus, 29(6), 613-619. https://doi.org/10.1111/dote.12373

van Hagen, P., Hulshof, M. C. C. M., van Lanschot, J. J. B., et al. (2012). Preoperative chemoradiotherapy for esophageal or junctional cancer. New England Journal of Medicine, 366(22), 2074-2084. https://doi.org/10.1056/NEJMoa1112088

Wang, Y., Springer, S., Zhang, G., et al. (2018). Detection of somatic mutations and HPV in esophageal cancers by next-generation sequencing. Nature Reviews Gastroenterology & Hepatology, 15(9), 541-556. https://doi.org/10.1038/s41575-018-0040-7