DOI:10.31557/APJCN.3144.20260915

RESEARCH ARTICLE

Converging Cancer Profiles in Bangladesh: An Integrated Statistical Analysis of Blood, Lung, Breast, and Cervical Cancer Datasets

S M Nazmuz Sakib[1, 2, 3, 4, 5]

[1]Graduate of LLB (Hon's), Faculty of Law, Dhaka International University, Satarkul Rd, Dhaka - 1212, Bangladesh. [2]Member, Bangladesh English Language Teachers Association (BELTA), Bangladesh. [3]Associate Member, Bangladesh Computer Society, Bangladesh. [4]Fellow, Scholars Academic and Scientific Society, H.No-204, Borhawor, P.S-Murajhar, Dist- Hojai, Assam-782439, India. [5]Member, International Association of Engineers (IAENG), Hong Kong.

Abstract

Background: Bangladesh faces a growing cancer burden that remains inadequately characterized because of limited population-based data and fragmented cancer surveillance systems. This study integrates four publicly available datasets from Bangladeshi clinical and academic settings, comprising 19,516 records across blood, lung, breast, and cervical cancers. This integrated approach provides a unique opportunity to identify cross-cancer patterns within a low- and middle-income country (LMIC) setting, where comprehensive clinical data remain limited. **Materials and Methods:** Descriptive statistics, chi-square tests of independence, one-way analysis of variance (ANOVA), and independent-samples t-tests were used to summarize patient characteristics, examine associations between clinical variables, and compare outcomes across four independent datasets: (1) a multivariate blood cancer clinical dataset (n = 2,295); (2) a lung cancer survival dataset (n = 1,782) collected from four tertiary hospitals in Dhaka; (3) a breast cancer molecular and clinical dataset (n = 15,054); and (4) a cross-sectional survey of cervical cancer awareness and HPV vaccine uptake among 385 female undergraduate students. **Results:** Significant associations were identified between cancer subtype and treatment outcome in the blood cancer cohort ($\chi^2 = 2307.76$, $p < 0.001$), disease stage and survival in the lung cancer cohort ($\chi^2 = 375.26$, $p < 0.001$), histologic grade and tumour size in the breast cancer cohort ($F(2,15045) = 140.66$, $p < 0.001$), and cervical cancer awareness and HPV vaccine knowledge ($\chi^2 = 32.88$, $p < 0.001$). A substantial knowledge-to-behaviour gap was observed in cervical cancer prevention: although 75.1% of unvaccinated participants expressed willingness to receive the HPV vaccine, only 8.3% had been vaccinated. Smoking status was not significantly associated with survival among patients with lung cancer ($p = 0.762$), and no individual genetic marker (BCR-ABL, FLT3, or TP53) was significantly associated with treatment outcomes in the blood cancer cohort. **Conclusion:** Disease stage emerged as the most consistent prognostic determinant across the cancer cohorts, underscoring the importance of early detection and timely diagnosis. The substantial gap between HPV vaccination willingness and actual vaccine uptake, together with the absence of a significant smoking-related survival effect in the lung cancer cohort, highlights important structural barriers to cancer prevention and healthcare delivery in Bangladesh. These findings support strengthening early detection programmes, expanding HPV vaccination strategies, and improving equitable access to cancer diagnosis and treatment.

Keywords: Bangladesh, blood cancer, breast cancer, cancer epidemiology, cervical cancer, HPV vaccination

Asian Pac J Cancer Nursing, 283-293

Submission Date: 06/27/2026 Acceptance Date: 08/15/2026

Corresponding Author:
Dr. S M Nazmuz Sakib
[1]Graduate of LLB (Hon's), Faculty of Law, Dhaka International University, Satarkul Rd, Dhaka - 1212, Bangladesh. [2]Member, Bangladesh English Language Teachers Association (BELTA), Bangladesh. [3]Associate Member, Bangladesh Computer Society, Bangladesh. [4]Fellow, Scholars Academic and Scientific Society, H.No-204, Borhawor, P.S-Murajhar, Dist- Hojai, Assam-782439, India. [5]Member, International Association of Engineers (IAENG), Hong Kong.
Email: sakibpedia@gmail.com

Introduction

Cancer remains one of the leading public health challenges worldwide. According to the Global Cancer Statistics 2024 (GLOBOCAN), approximately 20.6 million new cancer cases and 9.8 million cancer-related deaths occurred globally in 2024, corresponding to a lifetime risk of approximately one in five individuals developing cancer [1]. The burden is disproportionately greater in low- and middle-income countries (LMICs), where increasing cancer incidence coincides with limited healthcare resources, inadequate screening programmes, delayed diagnosis, and restricted access to treatment [3].

In Bangladesh, cervical cancer is one of the most common malignancies among women, whereas lung cancer remains the leading cause of cancer-related mortality among men [2]. Despite the growing burden of disease, the national cancer profile remains incompletely characterized because of limited population-based cancer registries and fragmented surveillance systems. Recent estimates indicate approximately 167,256 new cancer cases and 116,598 cancer-related deaths annually in Bangladesh, with lung, breast, and oral cancers accounting for a substantial proportion of this burden [3]. Nevertheless, comprehensive studies integrating multiple cancer datasets from Bangladesh remain scarce.

To address this gap, the present study integrates four publicly available datasets derived from Bangladeshi clinical and academic settings. Together, these datasets include 19,516 individual records collected between 2018 and 2023 and encompass hematological malignancies, lung cancer, breast cancer, and cervical cancer prevention. This integrated approach enables comparisons across diverse cancer types using a consistent statistical framework.

The objectives of this study were to: (1) describe the demographic and clinical characteristics of each cancer cohort; (2) identify significant associations between clinical variables and patient outcomes within each dataset; (3) identify common patterns across different cancer types that may inform national cancer control strategies; and (4) provide evidence to support cancer prevention and healthcare planning in Bangladesh.

By integrating evidence from multiple cancer datasets, this study offers a comprehensive perspective on cancer epidemiology in Bangladesh. The findings provide valuable insights into clinical outcomes, preventive behaviours, and healthcare gaps, thereby supporting data-driven public health policies aimed at improving cancer prevention, early detection, and patient management in resource-limited settings.

Materials and Methods

Datasets

Blood Cancer Clinical Profiles

The first dataset (DOI: 10.17632/d3vs7sbpt2.1) comprised 2,295 confirmed cases of blood cancer collected from multiple medical institutions across Bangladesh and published in the Mendeley Data repository in July 2025 [4]. The dataset included 15 demographic, clinical, and hematological variables, including age, sex, cancer subtype (AML, ALL, CLL, CML, lymphoma, and multiple myeloma), treatment modality, bone marrow aspiration findings, white blood cell (WBC) count, serum protein electrophoresis, platelet count, treatment outcome (cured, ongoing treatment, or deceased), genetic markers (BCR-ABL, FLT3, and TP53), and treatment-related side-effect severity. One record containing duplicated column-header information was identified during data cleaning and removed, resulting in a final analytical sample of 2,294 patients.

Lung Cancer Mortality and Survival

The second dataset (DOI: 10.17632/wzzxjs8gww.1), published in February 2026 [5], included clinical and demographic information for 1,782 patients with lung cancer treated at four tertiary hospitals in Dhaka, Bangladesh: the National Institute of Cancer Research and Hospital, Evercare Hospital Dhaka, United Hospital Limited, and Square Hospital Ltd. The dataset contained 18 variables, including age, sex, disease stage (Stages I–IV), family history of cancer, smoking status, body mass index (BMI), cholesterol level, comorbidities (hypertension, asthma, cirrhosis, and other cancers), treatment modality, and survival status.

Breast Cancer Molecular and Clinical Profiles

The third dataset (DOI: 10.17632/dbz42w9x8h.1), published in September 2025 by researchers from the University of Dhaka, consisted of 15,054 breast cancer cases [6]. Based on the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) framework, the dataset included 34 clinicopathological and molecular variables, including estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) status, PAM50 molecular subtype, histologic grade, tumour size, tumour stage, lymph node status, mutation count, overall survival, relapse-free survival, vital status, and treatment information (chemotherapy, hormone therapy, and radiotherapy).

Cervical Cancer Awareness and HPV Vaccine Survey

The fourth dataset consisted of a cross-sectional survey of 385 female undergraduate students enrolled at private universities in Bangladesh and was publicly available through the Figshare repository as supplementary data to a PLOS Global Public Health publication by Shah and Alam [7]. Data were collected between January and July 2023 using a structured self-administered questionnaire. The dataset included 27 variables covering demographic characteristics, awareness of cervical cancer and its risk factors, knowledge and uptake of HPV vaccination, awareness of cervical cancer screening, and behavioural intentions related to cancer prevention.

Statistical Analysis

All statistical analyses were performed using Python

with the pandas, NumPy, SciPy, and pyreadstat libraries. Data visualization was performed using Matplotlib.

Categorical variables are presented as frequencies and percentages, whereas continuous variables are summarized as means and standard deviations (SD).

Associations between categorical variables were evaluated using the chi-square test of independence. One-way analysis of variance (ANOVA) was used to compare mean tumour size across histologic grade categories in the breast cancer cohort. Where appropriate, statistically significant ANOVA results were followed by post hoc multiple-comparison analyses.

Independent-samples t-tests were performed to compare continuous variables between two independent groups, including comparisons of age and body mass index between surviving and non-surviving patients in the lung cancer cohort.

Pearson's correlation coefficient was calculated to assess the relationship between tumour size and mutation count in the breast cancer dataset.

All statistical tests were two-sided, and a p-value < 0.05 was considered statistically significant. Because large datasets may produce statistically significant results despite small effect sizes, effect size estimates were reported whenever appropriate. Missing data were handled using listwise deletion for each individual analysis.

Results

Blood Cancer: Distribution, Haematological Characteristics, and Treatment Outcomes

Cancer Subtype and Demographic Distribution

The blood cancer dataset included 2,294 patient records representing six malignancy subtypes. Chronic lymphocytic leukaemia (CLL) was the most common subtype (18.0%, n = 413), followed by acute lymphoblastic leukaemia (ALL) (17.5%, n = 401), chronic myeloid leukaemia (CML) (16.7%, n = 383), lymphoma (16.4%, n = 376), acute myeloid leukaemia (AML) (16.0%, n = 366), and multiple myeloma (15.5%, n = 355). The distribution of subtypes was relatively balanced across the cohort (Figure 1a). Treatment outcomes were similarly distributed, with 34.5% of patients classified as cured, 33.0% receiving ongoing treatment, and 32.5% deceased (Figure 1b). Side-effect severity was reported as severe in 26.0%, moderate in 24.3%, and mild in 24.2% of cases, while 15.5% of records contained missing values due to conditional data collection procedures.

Genetic Markers

Among the documented genetic alterations, BCR-ABL was the most frequently reported marker (n = 628, 27.4%), followed by FLT3 (n = 566, 24.7%) and TP53 (n = 564, 24.6%) (Figure 1a).

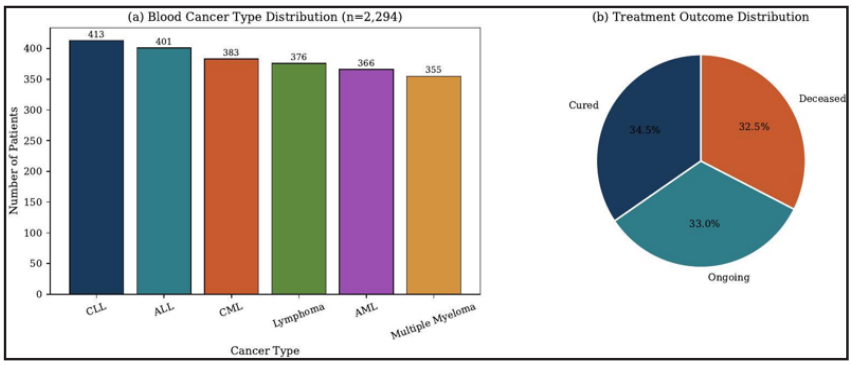

Figure 1. Distribution of Blood Cancer Subtypes and Treatment Outcomes. (a) Frequency of the six blood cancer subtypes included in the study. (b) Distribution of treatment outcomes (cured, ongoing treatment, and deceased).

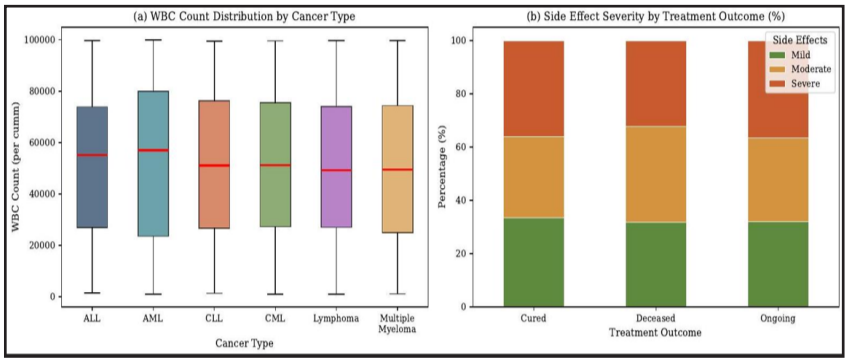

Figure 2. Hematological Parameters and Treatment-related Side Effects in Patients with Blood Cancer. (a) Distribution of white blood cell (WBC) counts according to blood cancer subtype. (b) Distribution of treatment-related side-effect severity according to treatment outcome.

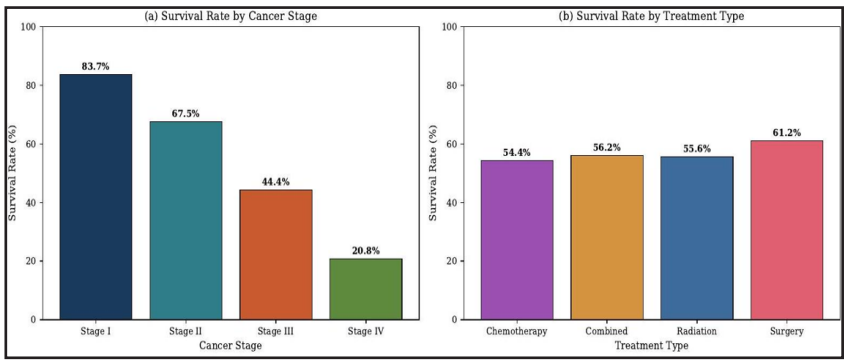

Figure 3. Survival According to Disease Stage and Treatment Modality in Lung Cancer. (a) Survival rates across disease stages. (b) Survival according to treatment modality.

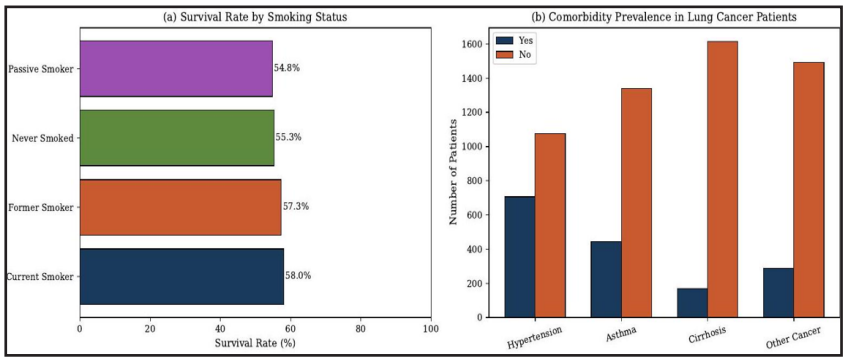

Figure 4. Smoking Status, Comorbidities, and Lung Cancer Outcomes. (a) Survival according to smoking status. (b) Distribution of major comorbidities among patients with lung cancer.

WBC Count According to Cancer Subtype

Mean white blood cell (WBC) counts varied across cancer subtypes. AML demonstrated the highest mean WBC count (53,357 cells/mm³; SD = 30,233), followed by ALL (51,900 cells/mm³; SD = 27,997), CLL (50,927 cells/mm³; SD = 28,029), CML (50,469 cells/mm³; SD = 28,159), lymphoma (50,325 cells/mm³; SD = 28,046), and multiple myeloma (49,883 cells/mm³; SD = 28,531) (Figure 2a).

Association Between Cancer Subtype and Treatment Outcome

A chi-square test of independence demonstrated a significant association between cancer subtype and treatment outcome ($\chi^2(18) = 2307.76$, $p < 0.001$) (Table 1). Among individual cancer subtypes, ALL showed the highest proportion of cured patients (38.0%), whereas AML demonstrated the highest proportion of deceased patients (37.7%).

Association Between Genetic Markers and Treatment Outcome

Treatment outcomes did not differ significantly according to the presence of the major genetic markers BCR-ABL, FLT3, and TP53. The distribution of cured, ongoing treatment, and deceased outcomes was comparable across these genetic groups ($p > 0.05$).

Lung Cancer: Survival, Disease Stage, and Clinical Characteristics

Cohort Characteristics

The lung cancer dataset included 1,782 patients with a mean age of 56.5 years (SD = 18.9; range: 25–89 years). Most patients were male. Overall, 1,012 patients (56.8%) were classified as survivors, whereas 770 (43.2%) were

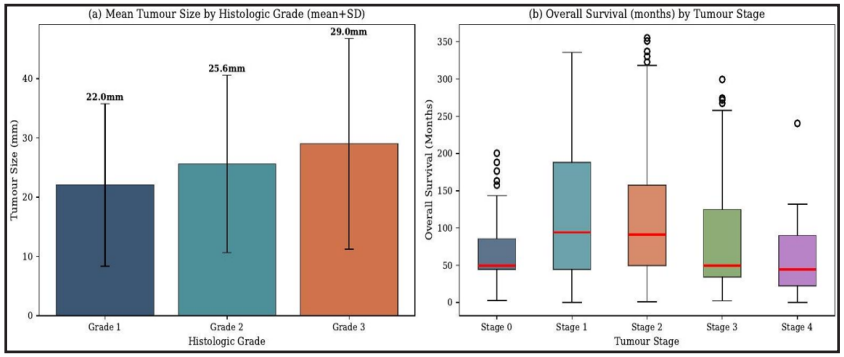

Figure 5. Histologic Grade, Tumour Size, and Survival in Breast Cancer. (a) Mean tumour size according to histologic grade. (b) Overall survival according to tumour stage.

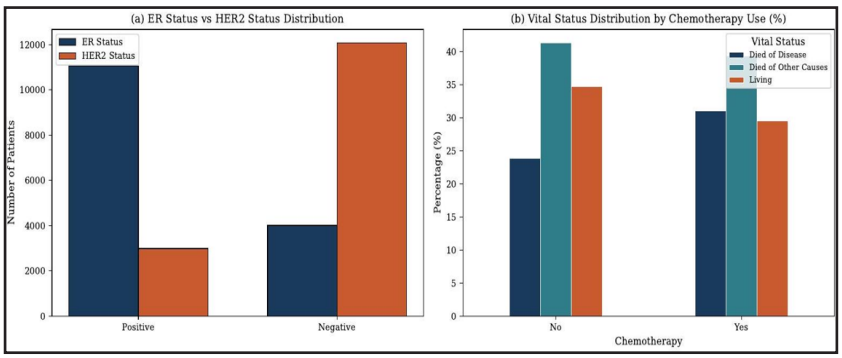

Figure 6. Molecular Characteristics and Treatment Patterns in Breast Cancer. (a) Distribution of ER and HER2 molecular subtypes. (b) Vital status according to chemotherapy use.

classified as non-survivors. Regarding disease stage, Stage II was the most frequent category (30.2%, n = 539), followed by Stage I (25.8%, n = 459), Stage III (24.0%, n = 428), and Stage IV (20.0%, n = 356).

Association Between Disease Stage and Survival

Disease stage demonstrated a strong association with survival outcomes. Survival decreased progressively with advancing stage: 83.7% among Stage I patients (384/459), 67.5% among Stage II patients (364/539), 44.4% among Stage III patients (190/428), and 20.8% among Stage IV patients (74/356). The association between stage and survival was statistically significant ($\chi^2(3) = 375.26$, $p < 0.001$), as shown in Figure 3a (Table 2).

Association Between Treatment Modality and Survival

Treatment modality was not significantly associated with survival outcomes ($\chi^2(3) = 4.70$, $p = 0.195$; Figure 3b). Survival rates were 61.2% among patients receiving surgery, 56.2% among those receiving combined treatment, 55.6% among those receiving radiotherapy, and 54.4% among those receiving chemotherapy.

Association Between Smoking Status and Survival

Smoking status was not significantly associated with survival (χ^2 (3) = 1.16, $p = 0.762$; Figure 4a). Survival rates were 58.0% among current smokers, 57.3% among former smokers, and 55.3% among never-smokers.

Comorbidities

Hypertension was the most common comorbidity, affecting 706 of 1,782 patients (39.6%), followed by asthma, cirrhosis, and other malignancies (Figure 4b). In this cohort, the presence of comorbidities was not significantly associated with overall survival, suggesting that tumour stage remained the predominant prognostic determinant.

Breast Cancer: Molecular Subtypes, Tumour Biology, and Survival

Cohort Characteristics

The breast cancer dataset included 15,054 patients with a mean age at diagnosis of 60.4 years (SD = 13.0; range, 21.9–96.3 years). At the end of follow-up, 5,028 patients (33.3%) were alive, 3,876 (25.7%) had died from breast cancer, and 6,138 (40.8%) had died from other causes. The relatively high proportion of deaths from

Table 1. Summary of the Principal Statistical Analyses Performed Across the Four Cancer Datasets.

Dataset	Variables Tested	Test	p-value	Interpretation
Blood Cancer	Cancer Type vs. Outcome	χ^2	<0.001	Significant
Blood Cancer	Genetic Marker vs. Outcome	χ^2	<0.001*	Artefact; not clinically meaningful
Blood Cancer	WBC Count by Subtype	Descriptive	-	AML highest; MM lowest
Lung Cancer	Stage vs. Survival	χ^2	<0.001	Highly significant
Lung Cancer	Smoking Status vs. Survival	χ^2	0.762	Not significant (null finding)
Lung Cancer	Treatment Type vs. Survival	χ^2	0.195	Not significant
Lung Cancer	Age vs. Survival	t-test	0.169	Not significant
Lung Cancer	BMI vs. Survival	t-test	0.609	Not significant
Breast Cancer	Grade vs. Tumour Size	ANOVA	<0.001	Highly significant
Breast Cancer	ER Status vs. Vital Status	χ^2	<0.001	Significant
Breast Cancer	Chemo vs. Vital Status	χ^2	<0.001	Significant (indication bias)
Breast Cancer	Tumour Size vs. Mutation Count	Pearson	0.632	Near-zero correlation (r = -0.004)
Cervical Cancer	CC Awareness vs. HPV Vaccine Awareness	χ^2	<0.001	Highly significant
Cervical Cancer	HPV Vaccine Uptake	Descriptive	-	8.3% (n=32/385)
Cervical Cancer	HPV Vaccine Willingness	Descriptive	-	75.1% (n=265/353)

Table 2. Overall Survival according to Disease Stage in Patients with Lung Cancer.

Stage	Total	Survived (n)	Not Survived (n)	Survival Rate (%)
Stage I	459	384	75	83.7
Stage II	539	364	175	67.5
Stage III	428	190	238	44.4
Stage IV	356	74	282	20.8
Total	1,782	1,012	770	56.8

χ^2 (3)=375.26,p<0.001

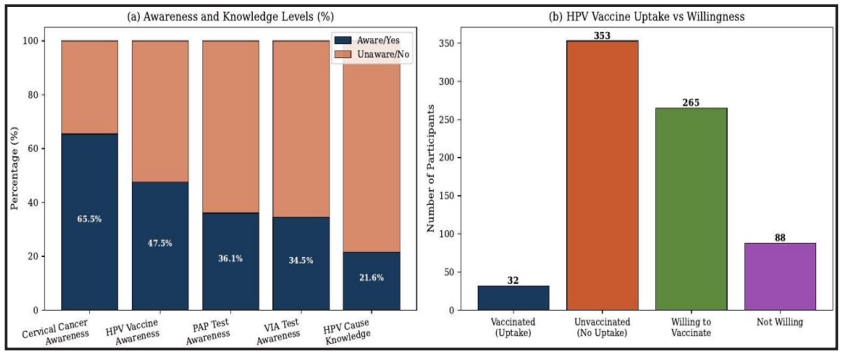

Figure 7. Cervical Cancer Awareness and HPV Vaccination. (a) Awareness of cervical cancer, HPV infection, and screening methods among female undergraduate students. (b) HPV vaccine uptake and willingness to receive vaccination.

other causes reflects the extended follow-up period of the METABRIC cohort.

Molecular Subtype Distribution

ER-positive tumours accounted for 73.4% of the cohort (n = 11,046), whereas HER2-positive tumours represented 19.8% (n = 2,988) (Figure 6a). Stage II disease was the most common stage at diagnosis, comprising 57.6% of cases (n = 8,670). These distributions are consistent with previously published findings from the METABRIC cohort [8, 9].

Histologic Grade and Tumour Size

Tumour size increased significantly with increasing histologic grade (ANOVA: F (2, 15045) = 140.66, p < 0.001) (Table 3). The mean tumour size was 22.0 mm (SD = 13.7) for Grade 1 tumours, 25.6 mm (SD = 15.0) for Grade 2 tumours, and 29.0 mm (SD = 17.8) for Grade 3 tumours (Figure 5a). This progressive increase indicates that higher-grade tumours tend to present with greater local tumour burden, consistent with established evidence in breast oncology [10].

Overall Survival by Tumour Stage

Overall survival differed markedly according to tumour stage. Patients with Stage I disease had the longest mean survival (117.7 months), followed by those with Stage II (108.9 months), Stage III (82.9 months), and Stage IV disease (62.1 months). Interestingly, patients with Stage 0 disease had a mean survival of 76.4 months, which may reflect diagnostic heterogeneity or the inclusion of in situ lesions with prolonged follow-up and late events (Figure 5b).

Chemotherapy and Vital Status

Chemotherapy administration was significantly associated with vital status (χ^2 (2) = 83.39, p < 0.001; Figure 6b). Among patients who did not receive chemotherapy, 33.5% were alive at the end of follow-up, compared with 29.6% of those who received chemotherapy. This finding most likely reflects treatment selection, whereby chemotherapy was preferentially administered to patients with more advanced or biologically aggressive disease.

Cervical Cancer: Awareness, Knowledge, and the HPV Vaccine Uptake Gap

Cervical Cancer Awareness

Among the 385 female undergraduate students

Table 3. Tumour Size according to Histologic Grade in the Breast Cancer Cohort.

Grade	n	Mean Size (mm)	SD (mm)	ANOVA F-statistic
Grade 1	1,326	22.04	13.70	
Grade 2	6,168	25.61	14.97	F(2,15045)=140.66
Grade 3	7,554	29.01	17.75	
Overall	15,048	26.52	16.27	p<0.001

Table 4. Summary of Cervical Cancer Awareness, HPV-related Knowledge, Vaccination Uptake, and Preventive Behavioural Intentions among Female Undergraduate Students.

Variable	Yes / Aware (n)	Percentage (%)
Cervical cancer awareness	252	65.5
HPV as cause of cervical cancer	83	21.6
HPV vaccine awareness	183	47.5
PAP test awareness	139	36.1
VIA test awareness	136	35.3
HPV vaccine uptake (actual)	32	8.3
HPV vaccine willingness (unvaccinated)	265	75.1
Screening willingness	249	64.7
Parental communication willingness	295	76.6
Vaccine recommendation willingness	-	-

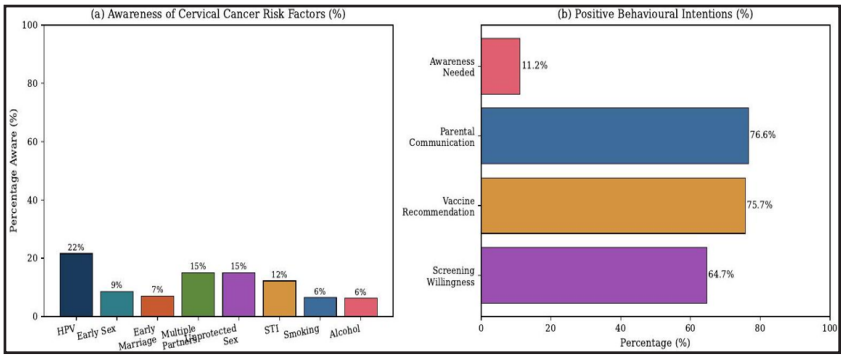

Figure 8. Knowledge of Cervical Cancer Risk Factors and Preventive Behaviours. (a) Awareness of major cervical cancer risk factors. (b) Behavioural intentions regarding cervical cancer prevention, including screening, HPV vaccination, and communication with parents.

included in the study, 252 (65.5%) reported having heard of cervical cancer. However, awareness of human papillomavirus (HPV) as the primary causative agent was considerably lower, with only 83 participants (21.6%) correctly identifying HPV as the cause of cervical cancer (Figure 7a). Knowledge of cervical cancer screening methods was similarly limited: 36.1% of participants were aware of the Papanicolaou (Pap) test, whereas awareness of visual inspection with acetic acid (VIA) was substantially lower.

Knowledge of Risk Factors

Among participants who were aware of cervical cancer (n = 252), knowledge of specific risk factors remained incomplete (Figure 8a). Unprotected sexual intercourse was the most frequently identified risk factor (23.7%), followed by multiple sexual partners (22.9%) and a history of sexually transmitted infections (21.3%). Awareness of HPV infection as the primary etiological factor, early sexual debut, and early marriage was comparatively lower, indicating substantial gaps in knowledge that should be addressed through targeted health education initiatives.

HPV Vaccine Uptake and Willingness to Receive Vaccination

A substantial gap was observed between HPV vaccine uptake and willingness to receive vaccination. Only 32 participants (8.3%) had received the HPV vaccine (Figure 7b). In contrast, among the 353 unvaccinated participants, 265 (75.1%) expressed willingness to be vaccinated. This marked discrepancy suggests that barriers to HPV vaccination in this population are more likely to be structural than motivational. These findings are consistent with previous studies conducted among university students in Bangladesh [11-13].

Screening Intentions and Preventive Behaviours

Overall, 249 participants (64.7%) expressed willingness to undergo cervical cancer screening. Similar proportions indicated willingness to recommend HPV vaccination to others, while 295 participants (76.6%) reported willingness to discuss cervical cancer prevention with their parents (Figure 8b). Chi-square analysis demonstrated a significant association between awareness of cervical cancer and knowledge of the HPV vaccine ($\chi^2(1) = 32.88$, $p < 0.001$), indicating that general awareness of cervical cancer is strongly associated with knowledge of preventive vaccination.

A summary of cervical cancer awareness, HPV-related knowledge, vaccination uptake, and preventive behavioural intentions is presented in Table 4.

Cross-Disease Synthesis

Age Profiles Across Cancer Types

The mean age at diagnosis differed across the cancer

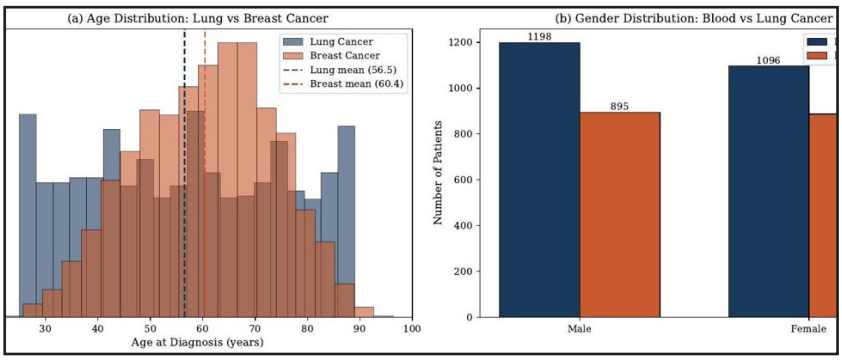

Figure 9. Comparison of Demographic Characteristics across Cancer Cohorts. (a) Age distribution at diagnosis for lung and breast cancer. (b) Gender distribution in the blood and lung cancer cohorts.

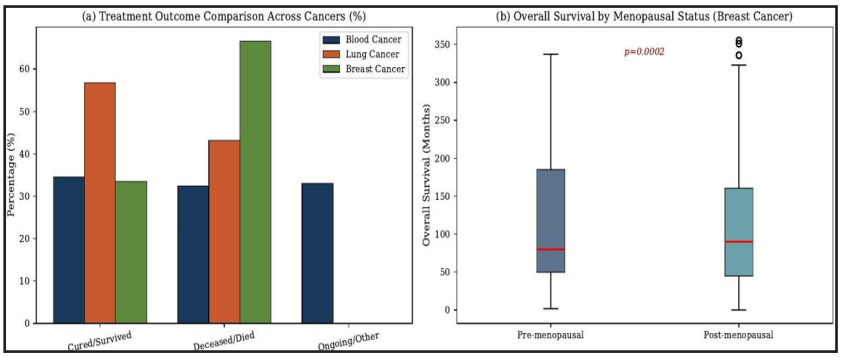

Figure 10. Cross-cancer Comparison of Survival Outcomes. (a) Comparison of treatment outcomes across blood, lung, and breast cancer cohorts. (b) Overall survival according to menopausal status in breast cancer.

cohorts (Figure 9a). Patients with lung cancer had a mean age at diagnosis of 56.5 years, whereas those with breast cancer had a mean age of 60.4 years. These findings are broadly consistent with global epidemiological patterns. However, the relatively younger age of lung cancer patients in the Bangladeshi cohort may reflect earlier tobacco exposure, occupational and environmental risk factors, and a higher proportion of patients presenting with advanced disease at diagnosis.

Stage as the Universal Prognostic Determinant

Across all four datasets, disease stage or, in the case of cervical cancer, the level of awareness and preventive behaviour emerged as the strongest determinant of clinical outcomes. Among patients with blood cancer, acute myeloid leukemia (AML) exhibited the highest mortality. In the lung cancer cohort, the survival rate declined to below 21% among patients with Stage IV disease. Similarly, patients with Stage IV breast cancer had a mean overall survival of only 62.1 months. In the cervical cancer cohort, limited awareness and poor preventive knowledge were associated with extremely low HPV vaccine uptake (Figure 10a). Collectively, these findings emphasize the critical importance of early diagnosis and preventive interventions across the cancer continuum.

Effects of Treatment Modality

Across both the blood cancer and lung cancer datasets, treatment modality alone was not a statistically significant predictor of clinical outcomes. In the lung cancer cohort, patients who underwent surgery demonstrated numerically better survival than those receiving other treatment modalities; however, the difference was not statistically significant ($p = 0.195$). These findings likely reflect the inherent limitations of observational data, in which treatment allocation is influenced by disease severity and clinical characteristics. Consequently, patients with more advanced disease are more likely to receive intensive treatment, potentially obscuring the true effect of therapeutic interventions. These observations highlight the need for more detailed clinical registries capable of accounting for treatment selection bias in future oncology research in Bangladesh.

Postmenopausal patients demonstrated significantly longer overall survival than premenopausal patients (Figure 10b). This apparent paradox may be explained by the higher prevalence of less aggressive, hormone receptor-positive tumours among older women, particularly the Luminal A molecular subtype, which is generally associated with a more favourable prognosis [15].

Discussion

The Predominant Role of Stage at Diagnosis

The most consistent finding across all four datasets was the strong influence of disease stage on patient outcomes. In the lung cancer cohort, survival declined dramatically from 83.7% in Stage I disease to 20.8% in Stage IV disease, representing more than a fourfold reduction. A similar pattern was observed in breast cancer, where mean overall survival decreased progressively with

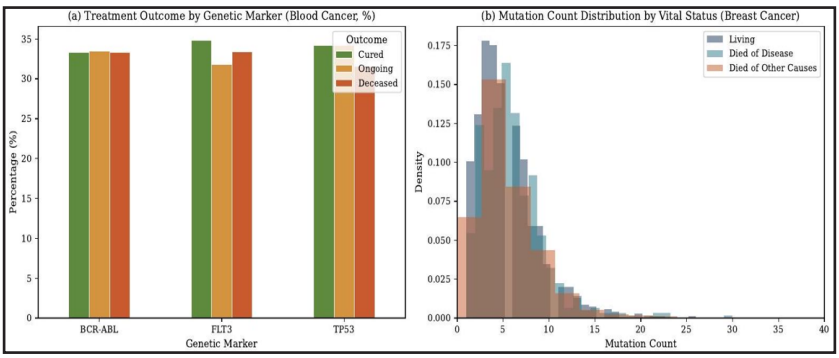

Figure 11. Genetic Alterations in Blood and Breast Cancer. (a) Treatment outcomes according to genetic markers in blood cancer. (b) Distribution of mutation burden according to vital status in breast cancer.

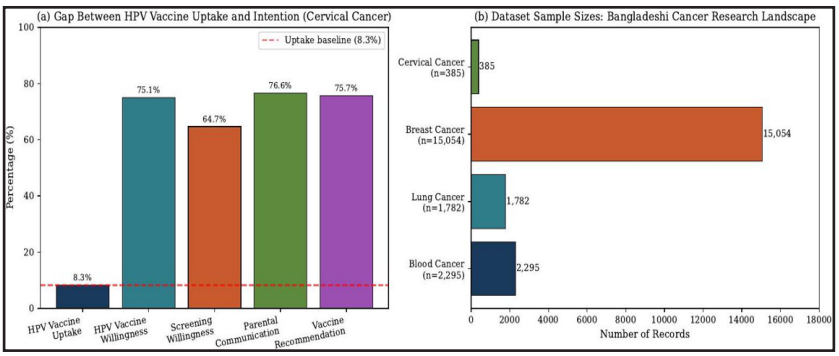

Figure 12. Integrated Overview of Cancer Prevention Indicators in Bangladesh. (a) Comparison of HPV vaccine uptake, willingness to receive vaccination, screening intention, and parental communication. (b) Sample size distribution across the four study datasets.

advancing stage. Although the prognostic importance of stage is well established, the consistency of this finding across four independent datasets strengthens its relevance as a system-level observation within the Bangladeshi healthcare setting. Bangladesh currently lacks a comprehensive national cancer screening program, and delayed diagnosis remains a major challenge in many low- and middle-income countries (LMICs) [14]. These findings provide further evidence supporting investment in early detection strategies, including population-based screening programs, community awareness initiatives, and improved access to diagnostic services.

The Smoking–Survival Paradox in Lung Cancer

The absence of a significant association between smoking status and survival in the lung cancer cohort ($p = 0.762$) warrants careful interpretation. In high-income countries, cigarette smoking is a well-established predictor of poorer lung cancer outcomes [1]. Several factors may explain the lack of association observed in the present study. First, the high prevalence of tobacco exposure, including passive and environmental smoking, may have reduced variability in smoking status, thereby limiting its discriminatory value. Second, disease stage at diagnosis may mediate much of the adverse effect of smoking on survival, diminishing the independent contribution of smoking status. Third, potential underreporting or the use of broad smoking categories (current, former, never, and passive smokers) may have reduced the ability to detect meaningful differences. These findings emphasize the importance of validating epidemiological associations within local populations rather than assuming that observations from high-income settings are universally applicable.

Genetic Markers and Clinical Outcomes in Blood Cancer

The observation that BCR-ABL, FLT3, and TP53 mutations were not significantly associated with treatment outcomes differs from evidence reported in clinical trials demonstrating substantial benefits of genotype-directed therapies, particularly tyrosine kinase inhibitors for BCR-ABL-positive chronic myeloid leukemia [15]. This discrepancy is likely attributable to the structure of the available dataset. The genetic marker variable identified only the principal mutation, whereas treatment outcomes were recorded using broad categories (cured, ongoing treatment, or deceased) without information linking specific therapies to individual genetic alterations. Consequently, the dataset did not permit evaluation of genotype-specific treatment effectiveness. Future oncology registries in Bangladesh should incorporate detailed molecular and treatment data to facilitate meaningful pharmacogenomic analyses and support precision oncology. Mutation burden across vital status groups is shown in Figure 11a and Figure 11b.

The Knowledge–Behaviour Gap in Cervical Cancer Prevention

One of the most important findings of this study is the substantial discrepancy between willingness to receive HPV vaccination (75.1%) and actual vaccine uptake (8.3%) among female undergraduate students. This difference of

nearly 67 percentage points suggests that low vaccination coverage is driven primarily by structural barriers rather than vaccine hesitancy. Potential contributing factors include the high cost of HPV vaccines in the private sector, the relatively recent implementation and limited coverage of the national HPV vaccination programme, restricted vaccine availability within university health services, and persistent sociocultural barriers surrounding discussions of reproductive health [11, 13, 16]. These findings indicate that expanding vaccine accessibility and strengthening public health education may substantially improve HPV vaccination coverage among young women.

This finding is consistent with previous studies. A 2023 cross-sectional survey involving 800 university students in Bangladesh reported an HPV vaccination rate of approximately 8.3%, despite a similarly high willingness to receive the vaccine [11]. Likewise, a 2025 study conducted among medical students in Khulna found an HPV vaccine uptake of only 11.23% [12]. Bangladesh's Expanded Programme on Immunization (EPI), introduced in late 2023, primarily targets school-aged girls and does not routinely include university students, leaving this population underserved [16]. The present findings provide further evidence supporting the expansion of the national HPV vaccination programme to include young women enrolled in tertiary education. An integrated overview of prevention indicators and dataset sample sizes is provided in Figures 12a and 12b, respectively.

Chemotherapy and Breast Cancer Outcomes: Interpretation of Counterintuitive Associations

The observed association between chemotherapy use and a lower proportion of surviving patients at follow-up ($p < 0.001$) is most likely attributable to indication bias. Chemotherapy was administered to 3,858 patients (25.6%), predominantly those with higher-grade and biologically more aggressive tumours. Given the observational nature of the dataset, causal relationships cannot be established. Therefore, these findings should not be interpreted as evidence that chemotherapy adversely affects survival; rather, they illustrate the methodological challenges of interpreting treatment effects using non-randomized real-world clinical data.

Limitations

This study has several limitations. First, all four datasets were derived from cross-sectional or registry-based sources, precluding causal inference. Second, the blood cancer dataset contained potential data-entry artefacts, including column headers embedded within the dataset, which may have inflated certain chi-square statistics. Third, although the breast cancer cohort was large, it originated from the METABRIC framework and may not fully represent the demographic and clinical characteristics of the Bangladeshi population. Fourth, the blood and lung cancer datasets lacked detailed time-to-event information, preventing survival analyses using Kaplan–Meier curves or Cox proportional hazards models. Finally, the cervical cancer survey employed convenience sampling among students from a private university, limiting the generalizability of its findings. Future research should prioritize prospective, longitudinal, registry-linked datasets with comprehensive treatment histories and standardized clinical variables.

In conclusion, this study presents an integrated cross-disease analysis of blood, lung, breast, and cervical cancer using four publicly available Bangladeshi datasets comprising 19,516 records. Several important findings emerged from the analyses.

Disease stage at diagnosis was the strongest and most consistent predictor of clinical outcomes across the four cancer datasets, highlighting early detection as the highest priority for cancer control strategies. Smoking status and genetic mutation type were not significantly associated with survival in the lung and blood cancer cohorts, respectively, suggesting that findings derived from high-income settings may not always be directly applicable to the Bangladeshi context. In the cervical cancer dataset, a substantial knowledge-to-behaviour gap was identified, with willingness to receive HPV vaccination exceeding actual vaccine uptake by nearly 67 percentage points, indicating that structural barriers rather than vaccine hesitancy remain the primary obstacle to vaccination. In the breast cancer cohort, higher histologic grade was significantly associated with larger tumour size, supporting the established relationship between tumour aggressiveness and disease progression. Furthermore, treatment modality was not independently associated with improved outcomes in the observational lung and blood cancer datasets, reflecting confounding by indication and the inherent limitations of non-randomized clinical data.

Collectively, these findings support the implementation of comprehensive national cancer control strategies in Bangladesh, including expansion of population-based screening programmes, broader access to HPV vaccination for young women enrolled in tertiary education, integration of molecular profiling with treatment records, and the development of a nationally coordinated cancer registry. The findings also demonstrate the value of integrating epidemiological and data science approaches to generate meaningful evidence for healthcare policy, while emphasizing the importance of recognizing the methodological limitations inherent in secondary data analyses.

Acknowledgements

The author gratefully acknowledges the researchers and institutions that made the datasets used in this study publicly available, including the contributors to the Mendeley Data repositories (DOIs: 10.17632/d3vs7sbpt2.1, 10.17632/wzzxjs8gww.1, and 10.17632/dbz42w9x8h.1), the Figshare repository associated with the PLOS Global Public Health publication by Shah and Alam, and the AIUB Institutional Review Board for overseeing the ethical conduct of the cervical cancer survey. No external funding was received for this study.

Conflict of Interest

The author declares that there are no conflicts of interest.

Data Availability

The datasets analysed during the current study are publicly available from the following repositories:

•Blood Cancer Dataset: https://doi.org/10.17632/d3vs7sbpt2.1

•Lung Cancer Dataset: https://doi.org/10.17632/wzzxjs8gww.1

•Breast Cancer Dataset: https://doi.org/10.17632/dbz42w9x8h.1

•Cervical Cancer Survey Dataset: Figshare (associated with PLOS Global Public Health by Shah and Alam).

References

1. Sung H, Filho AM, Laversanne M, Ferlay J, Siegel RL, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. CA: a cancer journal for clinicians. 2026;76(4):e70090. https://doi.org/10.3322/caac.70090
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2024;74(3):229-263. https://doi.org/10.3322/caac.21834
3. Iqbal J. National Cancer Registries in Bangladesh and Social Determinants of Health: Challenging Cancer Disparities. Cancer Control : Journal of the Moffitt Cancer Center. 2025 02 27;32:10732748251323757. https://doi.org/10.1177/10732748251323757
4. Siam M, Momin M, Syeid S, Hasan SMM. Comprehensive Hematological and Clinical Profiles Related to Blood Cancer in Bangladesh: A Multivariate Dataset for Epidemiological and Clinical Research Applications. 2025 07 09;2. https://doi.org/10.17632/d3vs7sbpt2.2
5. Paul T, Khan A. A Multi-Hospital Clinical Dataset for Lung Cancer Mortality and Survival Analysis in Dhaka, Bangladesh. 2026 02 23;1. https://doi.org/10.17632/wzzxjs8gww.1
6. Hira MIK. Breast Cancer Diagnostic Dataset (Tabular) for Machine Learning Classification. 2025 09 19;1. https://doi.org/10.17632/dbz42w9x8h.1
7. Rahman MM, Sharif AB, Uddin BMM, Bari S, Khan SI, Jerin IA, Rahman F, Orthi SU. Knowledge, attitude and practice related to cervical cancer prevention among medical students in Bangladesh: a cross-sectional study. Discover Education. 2025 08 10;4(1):273. https://doi.org/10.1007/s44217-025-00675-4
8. Prat A, Carey LA, Adamo B, Vidal M, Tabernero J, Cortés J, Parker JS, Perou CM, Baselga J. Molecular Features and Survival Outcomes of the Intrinsic Subtypes Within HER2-Positive Breast Cancer. JNCI: Journal of the National Cancer Institute. 2014 08 01;106(8):dju152. https://doi.org/10.1093/jnci/dju152
9. Cobec IM, Juhasz-Böss I, Seropian P, Huwer S, Varzaru VB, Rempen A, Moatar AE. Molecular Subtypes and Survival Patterns in Female Breast Cancer: Insights from a 12-Year Cohort. Medicina. 2025 Oct 16;61(10):1858. https://doi.org/10.3390/medicina61101858
10. Wu X, Baig A, Kasymjanova G, Kafi K, Holcroft C, Mekouar H, Carbonneau A, Bahoric B, Sultanem K, Muanza T. Pattern of Local Recurrence and Distant Metastasis in Breast Cancer By Molecular Subtype. Cureus.8(12):e924. https://doi.org/10.7759/cureus.924
11. Emon HH, Kowbi RN, Chowdhury MN, Teep TR, Habib MA, Tayeeba F, Hossain MM. Assessing knowledge, attitudes, and barriers toward HPV cervical cancer and vaccination among female university students in Bangladesh. Scientific Reports. 2025 Dec 24;15(1):44529. https://doi.org/10.1038/s41598-025-28112-4
12. Promi FT, Al-Abrar S, Nirob NCS, Mallick S, Islam KMD, Islam A, Akhter F. Knowledge, Attitude, and Practice Regarding Human Papillomavirus Infection Among Medical College Students in Khulna, Bangladesh. Epidemiology and Health System Journal. 2025 Dec 30;12(4):191-200. https://doi.org/10.34172/ehsj.26569
13. Rahman MM, Sharif AB, Uddin BMM, Bari S, Khan SI, Jerin IA, Rahman F, Orthi SU. Knowledge, attitude and practice related to cervical cancer prevention among medical students in Bangladesh: a cross-sectional study. Discover Education. 2025 08 10;4(1):273. https://doi.org/10.1007/s44217-025-00675-4
14. Islam MR, Hasan ATMK, Khatun N, Ridi IN, Rasheed MMO, Islam SMA, Karim MN. Demographic differentials of lung cancer survival in Bangladeshi patients. PLOS ONE. 2021 Dec 10;16(12):e0261238. https://doi.org/10.1371/journal.pone.0261238
15. Shahid AM, Vainchenker W, Constantinescu SN. JCMM Annual Review on Advances in Biotechnology for the Treatment of Haematological Malignancies: A Review of the Latest In-Patient Developments 2024–2025. Journal of Cellular and Molecular Medicine. 2025 07 06;29(13):e70700. https://doi.org/10.1111/jcmm.70700
16. "HPV vaccine can prevent cervical cancer in Bangladesh," Bangladesh Sangbad Sangstha (BSS) News, 2023. [Online]. Available: https://www.bssnews.net/news/206082.