

Modeling the Spatial Distribution of Cancer and Determining the Associated Risk Factors

Lijalem Melie Tesfaw  and Essey Kebede Muluneh

Department of Statistics, Bahir Dar University, Bahir Dar, Ethiopia.

Cancer Informatics
Volume 19: 1–14
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176935120939898



ABSTRACT

BACKGROUND: Cancer is the second leading cause of death globally. Despite developing countries including Ethiopia continuing to shoulder the greatest burden, insufficient research has been conducted to determine geographical and other characteristic effects. The main objective of this study was to assess the distribution and risk of cancer and determine the effects of some common clinical patient characteristics on current patient status by taking into account the spatial effect.

METHODS: The data for this study were obtained from the oncology ward of Felege Hiwot Referral Hospital, Bahir Dar, Ethiopia. About 415 cancer patients were included in the study. Spatial mixed ordinal logistic regression model was used to explore the geographical patterns of the incidence of cancer and identify the risk factors.

RESULTS: The findings of this study show that only 1.45% of patients were cured and 46.02% improved, whereas the rest have shown no change and even worse status after treatment. The estimated odds of patients who received chemotherapy was 4.284 times the estimated odds of patients who received palliative care. Prognostic factor (stage of cancer tumor), complication of cancer such as anemia during diagnosis, and treatment of patients given in the hospital had significant effect on the patient status.

CONCLUSION: Patients without anemia were more likely to be cured and improved than patients with anemia during diagnosis. Most of the patients had advanced stage (IV) of cancer tumor, which dismantles the capability of the treatment to be less effective. There was negative spatial effect on the incidence of cancer, indicating that districts with higher cancer incidence were usually surrounded by districts with lower incidence.

KEYWORDS: Incidence of cancer, patient status, spatial mixed OLR

RECEIVED: June 11, 2020. **ACCEPTED:** June 16, 2020.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Lijalem Melie Tesfaw, Department of Statistics, Bahir Dar University, P.O. Box 79, Bahir Dar 6000, Ethiopia. Email: lijalemmelie@gmail.com

Background

Cancer is a genetic disease caused by changes to genes that control the way our cells function, especially how they grow and divide, which leads to unregulated growth and division of cells that form malignant tumor and invade the nearby parts of our body. The cause of cancer, genetic changes, can be inherited from parents and also during a person's life time as a result of errors that occur as cells divide by certain environmental exposures. Cancer tumors are malignant, which can spread and invade nearby tissue, called metastatic cancer, which can start almost anywhere in the human body.^{1,2} Recently, the National Cancer Institute² showed that there are more than 100 types of cancer, usually named after the organs or tissues where the cancers form. Among these, only 20 types of cancers were considered in this study based on the registry of cancer patients in the oncology ward of Felege Hiwot Referral Hospital (FHRH) in the Western Amhara region of Ethiopia.

According to the Global Burden of Disease Cancer result explored in 2015, cancer was the second leading cause of death globally, 8 million deaths, with cardiovascular diseases being the first.³ Based on the 2018 world cancer statistics (excluding nonmelanoma skin cancer), there were an estimated 18 million cancer cases around the world, of which 9.5 million were in

men and 8.5 million in women. Globally, lung cancer is the first most commonly diagnosed cancer that affects most cases (12.3% of the total). Breast (2 088 849 cases, 12.2% of the total), colorectal (10.00%), prostate (9.8%), and stomach (5%) are within the top 5 most commonly diagnosed cancers, ranking second, third, fourth, and fifth, respectively.⁴

Cancer is a major public health burden in both developed and developing countries. About 72% of all cancer deaths in 2007 occurred in low- and middle-income countries.⁵ The same pattern of cancer deaths occurs in sub-Saharan African countries, particularly in Ethiopia. In Ethiopia, cancer accounts for about 5.8% of the total national mortality. According to the 2015 national cancer report of Ethiopia, among the entire adult population, breast cancer (30.2%), cervical cancer (13.4%), and colorectal cancer (5.7%) were the most prevalent cancers. As Ethiopia has diversified geographical area, in this study the spatial effect of cancer was also accounted for by determining the district where the patient resides. Assigning individuals to their place of residence also poses problems,⁶ although this is usually the only location information that is available. Often the goal of a geographic analysis is to identify a common environmental exposure in a population, but exposures that are occupational or recreational may not necessarily reveal themselves in a residential analysis.



A study by Jennifer et al⁷ reported that in Africa the number of cancer deaths is rising at an alarming rate, that is, the number of patients with cancer will increase almost 70% based on age demographics alone in 2030. Most cancers are diagnosed at an advanced stage in Africa mainly because of limited cancer treatment. Over 20% of African countries have no access to cancer treatments at all, whereas access is limited and sporadic in other countries.⁷ There are only 4 hospitals in Ethiopia that give treatment for a small number of patients with cancer, which is a very small number for a country that has more than 100 million people. The FHRH in Bahir Dar is 1 of the 4 hospitals that provide treatment for patients with cancer in Ethiopia. It was launched at the end of 2016. The growing population coupled with lifestyle changes means an increasing burden of cancer. However, oncology services are wholly inadequate to serve the entire country in the essential time. As a result of this, most of the patients are being treated only at the advanced stage of disease and die within short period or as soon as they arrive at the hospital.

The main objective of this study was to assess the distribution of the risk of cancer by taking into account the spatial effect on district and explore whether there is spatial autocorrelation of incidence of cancer between districts. It was also aimed to determine the effects of some common clinical patient characteristics, complications, and prognostic factors on current patient status.

Methods

Data

The data for this study were obtained from the registry of cancer patients in the oncology ward of FHRH. Felege Hiwot Referral Hospital is located in Bahir Dar, the capital city of Amhara region and a metropolitan city in the North Western region of Ethiopia. The oncology ward of the hospital is designed to register all cancer patients coming from different districts of Western Amhara region (see Figure 1). Nurses working in the oncology ward of the hospital were selected and trained to collect the cancer data of patients from September 2016 to January 2019.

Registries attempt to consolidate information by patient ID so that each case appears only once. Completeness of the data was checked and confirmed by a supervisor engaged in continuous supervision during data collection. For patients with cancer, a series of interventions, including psychosocial support or palliative care, surgery, and chemotherapy aimed at curing the disease or prolonging life considerably while improving the patient's quality of life, were carried out. Data were collected from the oncology ward of the hospital by retrospectively reviewing all new cancer patients in the cancer registry center report. A series of questions using oral and medical tools regarding risk factors, main symptoms, complications, comorbidities, treatment options (if they had used any before coming to the hospital), and prevention and early detection measures of

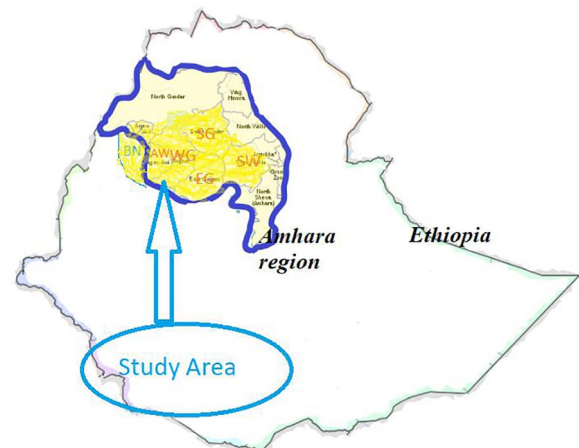


Figure 1. An overview map of the study region (yellow highlighted) in Amhara, Ethiopia, that consists of 6 zones. AW indicates Awi; BN, Benshangul; EG, East Gojjam; SG, South Gondar; SW, South Wollo; WG, West Gojjam.

cancer were asked to evaluate the cancer identity of each patient. After data collection, information was entered into Excel datasheets and then exported to other statistical software SAS 9.4 and R 3.5.2 for analysis. In this study, data from 415 cancer patients were collected. Of these patients, 285 (68.67%) have undergone chemotherapy and 120 (28.92%) have undergone both chemotherapy and surgery (combined), whereas the rest (10 [2.41%]) of the patients were given palliative care. Palliative care is an urgent humanitarian treatment given to relieve pain rather than cure when patients have advanced stages of cancer and little chance of cure.

Variable description

Before clinical assessment was provided at the oncology ward of FHRH, patients were asked to fill in their life history (for patient characteristics, see Table 1), and unique patient ID was assigned. Patient characteristics/risk factors such as age, gender, residence, and blood type; prognostic factors such as stage; complications such as anemia; and treatment and patient status indicated in Table 1 are variables considered in this study. Variables mentioned under risk factors are factors associated with causing a cancer and determined by looking at things that influence the incidence of new cancer cases, whereas prognostic factors can only be determined by following up people who already have cancer. Risk factors are patient characteristics associated with the risk of contracting cancer disease, which include age, gender, residence, and blood type. On the contrary, details of prognostic factor, complication, and treatment were collected from patients during treatment in the hospital, and finally, the patient status was recorded in their registry card after receiving treatment. The prognostic factor (stage of cancer) and the presence of complication (anemia) were registered at the beginning of the diagnosis, that is, as soon as they are registered at the hospital. Complications of cancer are unanticipated/unforeseen diseases that arise after and as a result of

Table 1. Description of variables considered in this study.

INDEPENDENT VARIABLES	FREQUENCY (%)
Characteristics and risk factors	
Age	
Gender	
Female (0)	257 (61.93)
Male (1)	158 (38.07)
Residence blood type	
A+	106 (25.54)
A-	18 (4.34)
AB+	17 (4.10)
AB-	4(0.96)
B+	101 (24.34)
B-	16 (3.86)
O+	121 (29.16)
O-	32 (7.71)
Prognostic factor	
Stage	
I	64 (15.42)
II	98 (23.62)
III	114 (27.47)
IV	139 (33.49)
Complication	
Anemia at diagnosis	
1 =yes	227 (54.70)
0=no	188 (45.30)
Treatment	
Chemotherapy	285 (68.67)
Chemotherapy and surgery	120 (28.92)
Palliative care	10 (2.41)
DEPENDENT VARIABLE	
Patient status	
0=cured	6 (1.45)
1=improved	191 (46.02)
2=same	108 (26.02)
3=deteriorate	85 (20.45)
4=death	25 (6.02)

Key: Residence=expressed in districts from 6 zones: West Gojjam (WG), East Gojjam (EG), Awi (AW), South Gondar (SG), South Wollo (SW), and Benshangul (BN).

cancer such as anemia in our study. Based on the information of patients in their registry card, the following variables were considered in this study (see Table 1).

Statistical Modeling

Spatial analysis

Public health data, such as the cancer data under analysis, often show differences or patterns across different geographical areas. Ignoring such data during analysis may give faulty results and conclusions. Spatial analysis is an analysis which includes the influence of space into the analysis. All statistical methods of spatial data have to take the spatial arrangement, and the resulting correlations, of observations into consideration to provide accurate and meaningful conclusions based on the analysis.^{8,9}

Spatial data are distinguished by observations that are obtained at spatial locations s_1, s_2, \dots, s_i , where s_i are coordinates⁸ in the plane \mathbb{R}^2 and rarely in the space \mathbb{R}^3 . Spatial data have different features such as point, line, area, and volume. Point is a precise location, s , in space indicated by a dot on a map; Line is a sequential collection of connected points like road and rivers; and Area is a region enclosed by lines like counties, states, and districts, 1 feature of spatial data that has been considered in this study. Weighted matrix, sometimes called Contiguity matrix, describes the relationship between districts i and j in the specified area. The (i, j) th element of a spatial proximity matrix \mathbf{W} , denoted by w_{ij} , quantifies the spatial dependence between regions i and j , and collectively, w_{ij} defines a neighborhood structure over the entire area. The spatial correlation parameter $\mathbf{W} = (w_{ij})$ is a neighborhood matrix for the areal units,¹⁰ which can be defined as

$$w_{ij} = \begin{cases} 1, & \text{if districts } i \text{ and } j \text{ share} \\ & \text{a common boundary, } i \neq j \\ 0, & \text{Otherwise} \end{cases} \quad (1)$$

In this case, the symmetric properties of \mathbf{W} are established because $w_{ji} = w_{ij}$, and its diagonal elements are equal to zero; being the similarity of the i th region with itself, $w_{ii} = 0$.

Spatial autocorrelation, covariation, or correlation between neighboring observations of variables gives us information on similarity between observations and similarity among locations. Based on the collection of weighted matrix, spatial proximity matrix, there are different measures of similarity that define different index classes. The 2 standard statistics that are used to measure the strength of spatial association among areal units are Moran's I and Geary's C .¹⁰ Moran's I is widely used,⁹ and variations of it relate to likelihood ratio tests and best invariant tests for particular models of correlation for normally distributed random variables. Having the spatial proximity matrix constructed, Moran's I statistics as a measure of global indexes of spatial autocorrelation with spatial proximity and similarity between areas i and j can be written as follows:

$$I = \left(\frac{1}{S^2} \right) \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}} \quad (2)$$

$$= \frac{1}{S^2} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} Y_{ij}}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}}$$

where Y_i is the total number of incidence of cancer diagnosed in the i th district, $\bar{Y} = \sum_{i=1}^N Y_i / N$ is the overall mean, and S^2 is the sample variance observed in Y_i 's that can be computed as $S^2 = (1/N) \sum_{i=1}^N (Y_i - \bar{Y})^2$.

In testing the presence of spatial autocorrelation, the null hypothesis states that the nearby districts do not affect one another, which implies that there is no dependence and spatial randomness in the data. In contrast, the alternative hypothesis states that there is spatial association or dependence among the districts. The research hypothesis for this study states that the nearby districts in Western Amhara region have an association with or dependence on the risk of cancer. Spatial autocorrelation in the nearby areas is considered to be present when the test statistic such as Moran's I takes on a larger value, compared with what would be expected under the null hypothesis of no spatial association.^{9,10} Spatial variable $W_{ij}Y_{ij}$ is a variable of a product of weighted matrices W_{ij} and Y_{ij} , a cross product of values in district i and j that deviated from the average value \bar{Y} (equation (2)). Y_{ij} is needed to measure the proximity, which means the distance between the observed i value and the neighboring j values. Although there are several types of distance-based methods,¹¹ the most common distance method, Euclidean distance, $Y_{ij} = (Y_i - \bar{Y})(Y_j - \bar{Y})$, was used in this study.

Ordinal logistic regression model

Ordinal logistic regression analysis deals with the association of a dependent variable with independent variables when the dependent variable has more than 2 categories having natural order or rank.¹² The dependent variable Y is assumed to have an ordinal scale with J categories, and $\mathbf{X} = x_1, x_2, \dots, x_p$ is the vector of explanatory variables. Then the chances of the variable response of the j th category of explanatory variable \mathbf{X} in particular can be expressed by $P[Y = j | \mathbf{x}] = \pi_j(\mathbf{x})$.

When response categories are ordered, the logits can use the ordering, which results in greater power and simple interpretation. The cumulative probability for Y is the probability that Y falls at or below a particular outcome category j and is given by

$$P(Y \leq j) = \pi_1(\mathbf{x}) + \dots + \pi_j(\mathbf{x}), \quad j = 1, 2, \dots, J \quad (3)$$

where J is the number of categories for the response variable Y .

The cumulative logit model^{13,14} is given as follows:

$$\text{logit}(P[Y \leq j | \mathbf{x}]) = \log \left(\frac{P[Y \leq j | \mathbf{x}]}{1 - P[Y \leq j | \mathbf{x}]} \right)$$

$$= \log \left(\frac{\pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) + \dots + \pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x}) + \pi_{j+2}(\mathbf{x}) + \dots + \pi_J(\mathbf{x})} \right) = \alpha_j + X'\beta \quad (4)$$

where $j = 1, 2, \dots, J-1$, and the probability $P[Y = j | \mathbf{x}]$ can be estimated as:

$$P[Y \leq j | \mathbf{x}] = \frac{\exp(\alpha_j + X'\beta)}{1 + \exp(\alpha_j + X'\beta)} \quad (5)$$

The cumulative probabilities do not use the final one, $P(Y \leq J)$, because it necessarily equals 1. The parameter β is a vector of regression coefficients describing the effect of the corresponding independent variable \mathbf{X} on the log odds of response in category j or below. When this model fits well, it requires a single parameter rather than $J-1$ parameters to describe the effect of \mathbf{X} , because the model assumes that the effect of \mathbf{X} is identical for all $J-1$ cumulative logits. This is known as the proportional odds model.

Random effects test

The random effects model is commonly used to detect whether the variable intended has random effect or not, beyond the fixed effect that is encompassed in the model.^{15,16} Random factors are not restricted to linear mixed models. Researchers want to incorporate random factors into nonlinear models to build a model that accommodates correlated data or to consider the levels of a factor as selected from a population of levels to make inference to that population.¹⁵ Common questions in mixed modeling are whether variance components are zero, whether random effects are independent, and whether rows (columns) can be added or removed from the covariance matrix. The effect of cancer type was detected by including in the model as a random effect. The patient status within one cancer type is likely to be correlated with each other. The goal here is to make inference for the population in the study area of cancer types. This could be accommodated by incorporating the cancer type into the model.

The random test can be conducted in 2 ways: (1) by adding the random effect to the model that contains the fixed effects and perceiving whether there is pragmatic change in the estimated parameter or (2) using mixture χ^2 statistic.^{13,15}

Spatial mixed ordinal logistic regression model

Spatial mixed ordinal logistic regression (SMOLR) is an analysis which incorporates spatial effects into the mixed ordinal logistic regression model. Scholars in the study by Muhammad and Tuti Purwaningsih^{14,17} used different methods to account for spatial effect and estimate its effect in their model. The spatial logistic regression model in Muhammad and Tuti Purwaningsih¹⁴ accounted for the spatial effect by including weighting of the ij th location through $w_{ij} = 1/h_{ij}$, where h_{ij} is the Euclid distance between the districts i, j and estimated existence of event. On the contrary, the SMOLR model is established to handle the spatial relationship proposed by Muhammad and Tuti Purwaningsih,¹⁷ and the specific model used in this study is given by

$$\begin{aligned} \text{logit}(P[Y_{ik} \leq j | x, \mu_k]) &= \log \left(\frac{P[Y \leq j | x, \mu_k]}{1 - P[Y \leq j | x, \mu_k]} \right) \\ &= \alpha_j + \mathbf{X}\beta + \rho \mathbf{W}\mathbf{y} + \mu_k \end{aligned} \quad (6)$$

where j is the j th category of the dependent variable and \mathbf{W} is a weighting matrix that represents the spatial proximity of the region; $w_{ij} = 0$ if regions i and j are not adjacent districts directly, whereas $w_{ij} = 1$ if districts i and j are immediately adjacent. Spatial weighting matrix (\mathbf{W}), which has been obtained, is multiplied by the vector \mathbf{y} and the results will be considered as a new variable; the so-called spatial variable will be used in ordinal logistic regression analysis.¹⁷ The parameter vectors β and ρ refer to the effect of the independent and spatial variable, respectively.

Goodness of fit of the model

Before making any inference, we need to select the best model out of the proposed spatial ordinal logistic regression (SOLR) and SMOLR models. The likelihood and pseudo-likelihood methods were used to estimate the parameters in the SOLR and SMOLR models, respectively. To choose the best model, Bayesian information criterion (BIC) and Akaike information criterion (AIC) were used.^{13,14}

Results

The mean age of patients at diagnosis was 43.5 years with a standard deviation of 15.54, and the median age was 45 years. The minimum age of patients included in the study was 2 years, whereas the maximum was 82 years. Most of the patients (72%) were between 30 and 60 years old, 16% were less than 30 years old, and the rest (12%) were more than 60 years old.

One-fourth of the patients (104 cases, 25.06% of the total) had advanced stages (IV) of cancer during diagnosis. This late detection and treatment of cancer leads to the substantial aggravation of the burden of cancers and bad outcomes of patient status. All patients took appropriate treatment based on physicians' prescription. Of those patients who have taken treatment, only 1.45% were cured and 46.02% have improved, whereas in the rest of patients there was no change and they became even worse after treatment. This shows that it is less likely that the patient can be cured, but that the effect of diseases can be improved if effective measures are put in place to control risk factors, detect cases early, and offer good care to those with the disease. The response variable, patient status, is constructed by categorizing the severity of cancer into 5 types: cured, improved, same, deteriorate, and death (see Table 1). This type of categorization is used by the hospital to manifest the general status of patients after treatment. Nearly half (47.5%) of the patients have improved or have been cured, and one-fourth remained in the same status after treatment. The distribution of men and women within different cancer types is different. Figure 2 reveals that a large proportion of female

patients suffer from breast (93.8%), gastric (88.9%), hepatocellular carcinoma (58.3%), rectal (61.5%), and ovarian cancer (97.2%). On the contrary, colonic (70.0%), Hodgkin lymphoma (100.0%), non-Hodgkin lymphoma (72.9%), pancreatic (75.0%), round blue cell tumor (66.7%), and sarcoma cancer (60.0%) were more predominant in male than female patients.

There were 20 cancer types available and considered in this study based on the oncology ward cancer registry at the FHRH (see Table 2). This table revealed all types of cancer in the study and their corresponding number of patients and the percentage share that belongs to each cancer type among a total of 415 patients. Breast cancer was the highest cancer type (19.28% of the total) diagnosed, followed by NHL (16.87%).

Forty-one districts from 6 zones were considered in the study (Table 3). There are 13, 8, 7, and 10 districts from West Gojjam, East Gojjam, Awi, and South Gondar zones, respectively, whereas South Wollo and Benshangul are the only zones that consist of only 1 district. Patients belonging only to these districts were available in oncology cancer registry ward of the hospital.

Address-matching of cancer registry records was obtained from the zone administration office of the districts, which was carried out using a commercial geographical positioning system (GIS) product to assign approximate latitude-longitude coordinates to a patient's reported district of residence. The geographical coordinates were recorded by collaborating with GIS experts in the zone administration office. There were differential variations in patterns across the 41 districts. The highest incidence of cancer was recorded in Bahir Dar. However, the population at risk is not in the standardized form, in a sense that the highest incidence of cancer in Bahir Dar might be due to the high number of inhabitants in the district/city. This suspected problem was handled by standardizing the population at risk per 100 000 per district. Finote Selam district was then found to have the highest incidence of cancer (19.59 per 100 000 population) relative to other districts in the study (Table 4).

Figure 3 shows the incidence of cancer within different age groups irrespective of the sex of patients. Even if the population at risk within each age group was not standardized, the incidence of cancer among women and men is different. The incidence in women was high than men between the age group of 21 and 70 years, and vice versa in the age group below 21 and above 70 years. The incidence of cancer increases with increasing rate until it reaches age 50, using the average cubic estimated reference line. After age 50, the incidence of cancer declines with increasing rate.

For the sake of simplicity, to explore the incidence of cancer with age, the variable was constructed by categorizing into different age groups in which each group contains 5 years. This type of categorization is consistent with a previous report.² In Figure 3, the incidence of cancer in each age group was obtained by summing up the number of patients within the specified age group.

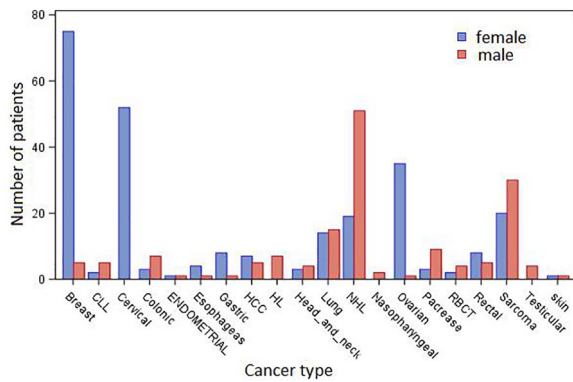


Figure 2. Distribution of the number of female (blue color) and male (red color) patients within different cancer types. CLL indicates chronic lymphocytic leukemia; HCC, hepatocellular carcinoma; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; RBCT, round blue cell tumor.

Table 2. Cancer types and respective number of patients.

CANCER TYPE	NO. OF PATIENTS	PERCENT
Breast	80	19.28
CLL	7	1.69
Cervical	52	12.53
Colonic	10	2.41
Endometrial	2	0.48
Esophagus	5	1.2
Gastric	9	2.17
HCC	12	2.89
HL	7	1.69
Head and neck	7	1.69
Lung	29	6.99
NHL	70	16.87
Nasopharyngeal	2	0.48
Ovarian	36	8.67
Pancreas	12	2.89
RBCT	6	1.45
Rectal	13	3.13
Sarcoma	50	12.05
Testicular	4	0.96
Skin (SCCA)	2	0.48

Abbreviations: CLL, chronic lymphocytic leukemia; HCC, hepatocellular carcinoma; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; RBCT, round blue cell tumor; SCCA, squamous cell carcinoma antigen.

The description of the incidence of cancer described in Table 3 did not account for the population at risk in each district, which leads to biased conclusion. The highest incidence of cancer was recorded in Finote Selam district with an

incidence rate of 19.5929, followed by Chagni district with an incidence rate of 14.6985, whereas the lowest incidence of cancer was recorded in Enemay district with an incidence rate of 1.1117, followed by Smada district with an incidence rate of 1.2120. Hence, to determine the incidence of cancer for each district, the population at risk should first be considered to have plausible description of the incidence rate. The latitude and longitude which show where each district is located in the Western Amhara region are illustrated graphically in Figure 4, which shows the incidence of cancer in each district.

The categorical covariates have different number of categories based on the nature of the covariate. The bivariate analysis using the χ^2 statistic, presented in Table 5, shows the relationship between categorical independent variables and the dependent variable (categories of patient status—cured, improved, same, deteriorate, and death). The χ^2 statistic was used to compare the actual frequencies in a bivariate table with the frequencies that would have been expected if there was no relationship between the variables. We can see from the table that the only variables not significantly associated with patient status were sex and blood group, meaning that the patterns of responses across both sex and different blood groups were essentially the same for all categories of the independent variable.

Table 5 shows that there is not any cancer patient with anemia who was cured. This might show how a complication like anemia would aggravate the severity of cancer. In parallel, of those patients who died of cancer, the higher proportion of patients (68.0%) had anemia and the remaining (32%) did not. From this, we can highly suspect that the patient status has a significant association with anemia. The corresponding P value of .001 indicates that anemia has significant association with patient status. The patterns of proportion of cancer patients with and without anemia seem to vary by patient status. Patients without anemia were more likely to enter into the lowest scale patient status (100% vs 0.00% for cured and 61.8% vs 38.2% for improved) than patients with anemia and less likely than patients with anemia to enter into the highest scale patient status (44.7% vs 55.3% for deteriorate and 32.0% vs 68.0% for death). The reverse is true for patient with versus without anemia.

There was a significant association ($P=0.001$) between treatment given for patients and patient status. Among patients of the same status, 75.9% took chemotherapy, whereas 20.4% have taken both chemotherapy and surgery (combined). The stage of patients had a significant association with patient status. Most of the patients who died of cancer were in stages III and IV (20% and 64%, respectively). This shows that the odds of dying of cancer increases with the stage of patients, meaning the patients in advanced stages are more likely to die.

Spatial autocorrelation, Moran's I test

The weighted matrix \mathbf{W} was defined to evaluate the existence of spatial autocorrelation. The matrix allows the measurement

Table 3. Districts within zones, respective number of patients, and total number inhabitants (population at risk).

ZONE	DISTRICT	NO. OF PATIENTS (%)	POPULATION AT RISK
WG	North Achefer	20 (4.82)	138 114
	South Achefer	14 (3.37)	119 578
	Burie	9 (2.17)	109 389
	Bahir Dar	46 (11.08)	620 900
	Bahir Dar Zuria	28 (6.75)	182 794
	Dembecha	10 (2.41)	100 897
	Finote Selam	8 (1.93)	28 993
	Gonji Kolela	10 (2.41)	98 654
	Mecha	26 (6.27)	287 569
	Merawi	5 (1.20)	32 318
	Sekela	6 (1.45)	114 425
	Woberma	5 (1.20)	85 283
	Jabi Tehnan	15 (3.61)	140 012
	Yilmanadensa	17 (4.10)	195 613
EG	Bibugn	6 (1.45)	83 243
	Debre Markos	5 (1.20)	91 340
	Enebse Sarmidr	5 (1.20)	106 630
	Enemay	2 (0.48)	140 905
	Goncha Siso enese	5 (1.20)	121 133
	Hulet ejun enese	11 (2.65)	251 845
	Machakel	4 (0.96)	108 270
	Mota	7 (1.69)	54 057
AW	Ankasha Guagusa	15 (3.61)	187 144
	Banja	9 (2.17)	104 599
	Chagni	5 (1.20)	34 017
	Dangla	13 (3.13)	131 897
	Fagita Lekoma	5 (1.20)	117 249
	Guangua	4 (0.96)	203 196
	Jawi	6 (1.45)	66 012
SG	Alefa	5 (1.20)	129 215
	Libo Kemkem	15 (3.61)	209 760
	Debre Tabor	5 (1.20)	62 989
	Dera	18 (4.34)	217 513
	Estie	10 (2.41)	308 929
	Ebinat	9 (2.17)	192 235
	Farta	6 (1.45)	201 790
	Fogera	9 (2.17)	196 682
	Gaynt	5 (1.20)	250 518
	Smada	3 (0.72)	207 532
SW	Dessie	6 (1.45)	118 372
BN	Pawi	3 (0.72)	47 829

Abbreviations: AW, Awi; BN, Benshangul; EG, East Gojjam; SG, South Gondar; SW, South Wollo; WG, West Gojjam.

Table 4. Incidence of cancer disease per 100 000 standardized population at risk of each district.

DISTRICT	LONGITUDE	LATITUDE	INCIDENCE OF CANCER	POPULATION AT RISK
Finote Selam	10.71	37.26	19.59	28 993
Chagni	10.95	36.50	14.70	34 017
Merawi	11.25	36.50	13.40	32 318
Mota	11.08	37.87	12.95	54 057
Bahir Dar Zuria	11.25	37.17	12.51	182 794
North Achefer	11.60	37.03	10.63	138 114
Gonji Kolela	11.45	37.67	8.65	98 654
South Achefer	11.83	37.17	8.26	119 578
Debre Tabor	11.85	38.22	7.94	62 989
Mecha	11.50	37.00	7.87	287 569
Jabi Tehnan	10.80	37.17	7.73	140 012
Banja	11.17	36.25	7.61	104 599
Dangla	11.42	36.67	7.56	131 897
Dembecha	10.67	37.17	7.10	100 897
Debre Markos	10.23	37.32	7.01	91 340
Jawi	11.75	36.42	6.96	66 012
Ankasha Guagusa	11.00	36.67	6.91	187 144
Yilmanadensa	11.50	37.33	6.86	195 613
Bibugn	11.00	37.58	6.80	83 243
Bahir Dar	11.60	37.38	6.75	620 900
Dera	11.75	37.50	6.73	217 513
Pawi	11.33	36.33	6.27	47 829
Libo Kemkem	12.33	37.68	6.01	209 760
Burie	10.70	37.06	5.76	109 389
Woberma	10.37	35.75	4.53	85 283
Sekela	11.17	37.00	3.99	114 425
Ebinat	12.17	38.08	3.88	192 235
Dessie	11.13	38.53	3.78	118 372
Fogera	11.97	37.68	3.65	196 682
Hulet ejun enese	11.25	37.75	3.64	251 845
Fagita Lekoma	11.33	36.75	3.64	117 249
Enebse Sarmidr	11.08	38.25	3.41	106 630
Machakel	10.67	37.33	3.12	108 270
Goncha Siso enese	11.17	38	3.10	121 133
Alefa	12.25	36.33	2.79	129 215
Estie	11.67	38.17	2.79	308 929
Farta	12.00	38.00	2.38	201 790
Guangua	11.00	35.25	1.64	203 196
Gaynt	12.00	38.33	1.49	250 518
Smada	11.50	38.25	1.21	207 532
Enemay	10.67	38.00	1.11	140 905

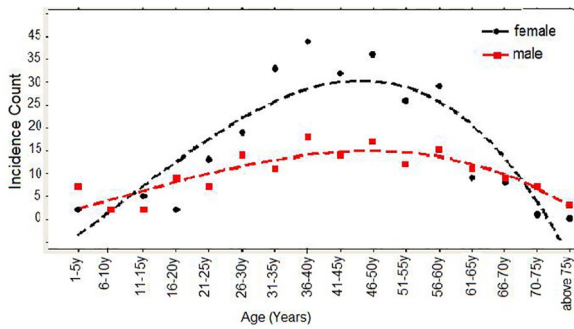


Figure 3. Incidence of cancer in different age groups for women (black color) and men (red color).

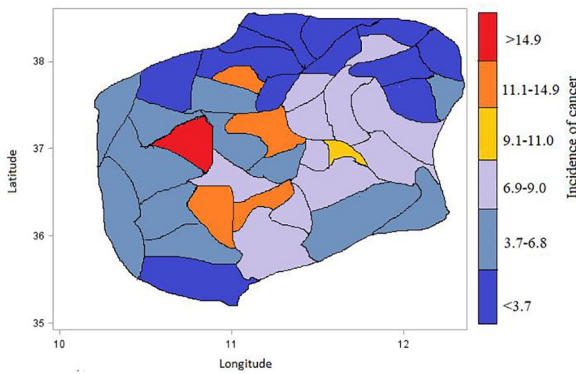


Figure 4. Spatial distribution of the incidence of cancer (per 100000 population) in the respective location specified by longitude and latitude for each districts.

of the nonrandom association between the value of incidence of cancer observed in a given district and the value of incidence of cancer observed in neighboring district units which is used to compute Moran’s *I* index. Moran’s *I* index is a univariate analysis that measures the spatial autocorrelation, in a sense that the index allows to detect whether or not the incidence of cancer in a given district may be similar to that of neighboring districts. The estimated Moran’s *I* statistic index is -0.0491 with a *P* value of .0022 (Table 6). The *P* value is found to be less than .05, suggesting significant evidence of unexplained spatial autocorrelation in the incidence of cancer. Negative Moran’s *I* values indicate negative spatial association. On the contrary, to identify whether there is spatial correlation, in addition to looking at the *P* value, we can simply compare the observed Moran’s *I* statistic index with the expected Moran’s statistic index $E(I) = -1/(n-1)$. The observed value of Moran’s $I = -0.0491$ is less than its expected value $E(I) = -0.0024$, showing a negative spatial autocorrelation or a clustered pattern, which means that observations from nearby districts tend to be less alike than observations from districts farther apart.

To make the incidence of cancer comparable among different districts, the population at risk was standardized into 100000 population. As a result, the values of incidence of cancer indicate the incidence of cancer out of a total of 100000 population at risk in the given districts.

The Moran’s *I* index used to test the spatial autocorrelation can be described graphically (Figure 4) and can determine whether the decision made in the Moran’s *I* statistic was plausible and persistent. It also verifies whether there is evidence of clustering of incidence of cancer or whether there are districts that have unusual clusters of cancer so that simply we can perceive causal explanations by seeing whether there is spatial patterning of incidence of cancer.

The standardized incidence of cancer for each district is displayed in the form of maps in Figure 4. The color from warmer blue color to warmer red color shows districts with lower to higher incidence of cancer disease, respectively. The high value of incidence of cancer (19.5929) was found in Finote Selam district and the low incidence of cancer (1.1117) was found in Enemay district as described (Table 4). Enemay and Smada districts are found to have the lower incidence of cancer. One possible reason for this may be that both Enemay and Smada are districts that are very far from the treatment center and hence it will be difficult to come for diagnosis. Moreover, it might be because most of the people are living in rural areas and there is lack of awareness about the disease. On the contrary, both districts are highland with relatively cold weather condition compared with the other neighbor districts, which might lead the districts to have lower likelihood of the incidence of cancer.

Model Fitting and Estimating Parameters

Prior to directly fitting the model, some preliminary analysis was done to select the best fit model with appropriate covariates. The spatial variable (S_g) was created using a multiplication of spatial weighting matrix **W** and incidence of cancer (*Y*) in each district, which can simply be denoted by *Sv* instead of *Wy*. The spatial weighting matrix **W** looks at the closeness between districts, which is 1 if adjacent and 0 if not directly adjacent (see equation (1)). Hence, for model building, spatial variable (simply) was used instead of districts.

Variables were selected using stepwise method of selection. In fact, it was also done manually using the backward and forward selection method, which results in the same set of variables. During stepwise variable selection, the selection was done by 15% entry and 20% stay significance level, as also mentioned in the study by Hosmer et al.¹³ For all selected variables, all possible pairwise interaction effects were examined in the model. However, there was no any interaction effect between all selected covariates with 15% entry and 20% stay significance level. As numerous studies in different areas and eras¹⁸⁻²³ reported that cancer is highly interrelated with age, we realized that maintaining the covariate age in our model (Table 7) is advisable. To perceive the effect of independent variables at the beginning, intercept-only ordinal logistic regression (OLR) model without any independent variables was fitted. Then, the OLR model that includes all independent variables including the spatial variable was fitted and model comparison was done

Table 5. Association between patient status and covariates (N=415).

COVARIATES	PATIENT STATUS					P VALUE
	CURED, NO. (%)	IMPROVED, NO. (%)	SAME, NO. (%)	DETERIORATE, NO. (%)	DEATH, NO. (%)	
Anemia						
No	6 (100)	118 (61.8)	57 (52.8)	38 (44.7)	8 (32.0)	.001
Yes	0 (0.0)	73 (38.2)	51 (47.2)	47 (55.3)	17 (68.0)	
Blood group						
A-	0 (0.0)	8 (4.2)	5 (4.6)	5 (5.9)	0 (0.0)	.89
A+	1 (14.29)	47 (24.6)	24 (22.2)	25 (29.4)	9 (36.0)	
AB-	0 (0.0)	2 (1.0)	2 (1.9)	0 (0.0)	0 (0.0)	
AB+	0 (0.0)	8 (4.2)	5 (4.6)	3 (3.5)	1 (4.0)	
B-	0 (0.0)	7 (3.7)	3 (2.8)	6 (7.1)	0 (0.0)	
B+	1 (14.29)	49 (25.7)	26 (24.1)	19 (22.4)	6 (24.0)	
O-	2 (28.57)	13 (6.8)	7 (6.5)	9 (10.6)	1 (4.0)	
O+	3 (42.86)	57 (29.8)	36 (33.3)	18 (21.2)	8 (32.0)	
Treatment						
Chemotherapy	3 (50.0)	124 (64.9)	82 (75.9)	58 (68.2)	18 (72.0)	.001
Combined	2 (33.3)	66 (34.6)	22 (20.4)	26 (30.6)	4 (16.0)	
Palliative	1 (16.7)	1 (0.5)	4 (3.7)	1 (1.2)	3 (12.0)	
Sex						
Female	3 (50.0)	112 (58.6)	65 (60.2)	59 (69.4)	18 (72.0)	0.346
Male	3 (50.0)	79 (41.4)	43 (39.8)	26 (30.6)	7 (28.0)	
Stage						
I	3 (50)	34 (17.80)	17 (16.04)	11 (12.94)	1 (4.00)	.000
II	1 (16.67)	62 (32.46)	25 (23.58)	7 (8.24)	3 (12.00)	
III	1 (16.67)	66 (34.55)	30 (28.30)	12 (14.12)	5 (20.00)	
IV	1 (16.67)	29 (15.18)	34 (32.08)	55 (64.71)	16 (64.00)	

key: Combined=both chemotherapy and surgery.

using AIC and BIC. The estimated fit statistic of AIC (BIC) was 1056.067 (1072.180) and 924.848 (981.244) for intercept-only OLR and SOLR models, respectively. The dummy variable and the reference (*) variables indicated in Table 7 and the estimated SOLR model can be written by

$$\begin{aligned} \text{logit}(P[Y_i \leq j]) = & \alpha_j + \beta_1 \text{Age}_i + \beta_{20} \text{Anemia}_i \\ & + \beta_{30} \text{chemotherapy}_i + \beta_{31} \text{combined}_i + \beta_{40} \text{Stage_I}_i \\ & + \beta_{41} \text{Stage_II}_i + \beta_{42} \text{Stage_III}_i \end{aligned} \quad (7)$$

where α_j is the intercept and $j=0, 1, 2,$ and 3 for cured, improved, same, and deteriorate status of patient, respectively, and $i=1, 2, \dots, 415$. $\beta_1, \beta_2, \dots,$ and β_m are parameters of the

corresponding independent variables to be estimated, which describe the effect of the corresponding variable on the log odds of patient status at j or below category. In the parameter β_s , there is no subscript j because the model assumes proportional odds, and hence the effect of independent variables on the patient status is identical for all $J-1$ cumulative logit models.

The requirement of random effects of cancer type in the model was tested using a mixture of χ^2 distributions. These tests were done by adding cancer type as the random effect in the SOLR model which becomes the SMOLR model. The need for random effect was significant ($\chi^2_{0.1} = 6.28; P = 0.0254$), which indicates that the model without cancer-type random

Table 6. Indicator of spatial autocorrelation.

INDICATOR	STATISTIC	P VALUE
Moran's <i>I</i>	-0.0491	.0022

effects does not fit the data well. The estimated variance of random effects of cancer type was 0.2997 by assuming constant variance-covariance working correlation structure.

To estimate the parameters of the SMOLR model, estimation methods such as commutative logistic mixed model (clmm) and generalized linear mixed model (glmmix) using R and SAS statistical software, respectively, were used. Nevertheless, at the end, the estimated parameters for the models described in Table 7 were obtained using the glmmix SAS procedure.

Model comparison was done based on AIC and BIC, the model with the smallest AIC and BIC being preferred. The fit statistics AIC (BIC) for intercept-only OLR model, SOLR model, and SMOLR model were 1056.067 (1072.180), 924.848 (981.244), and 923.040 (937.970), respectively (Table 7). The SMOLR model having the smallest AIC (BIC) is therefore the best model for the data.

Therefore, the final estimated model can be written by

$$\begin{aligned} \text{logit}(P[Y_i \leq j]) = & \hat{\alpha}_j - 0.003 \text{Age}_i + 0.6034 \text{Anemia}_i \\ & + 1.4549 \text{Chemotherapy}_i + 1.9799 \text{Combined}_i \\ & - 0.6460 \text{Stage}_i - 1.2438 \text{Stage_II}_i \\ & - 0.7285 \text{Stage_III}_i \end{aligned} \quad (8)$$

As patient status has 5 categories ($J=5$), the model has 4 intercepts such as $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2$ and $\hat{\alpha}_3$ with estimated values of -6.2469, -1.6466, -0.03957, and 2.1793, respectively. Usually, these are not of interest except for estimating the probability of patient status that falls at category j or below, $P[Y \leq j]$.

The SMOLR model (Table 7) shows anemia at diagnosis, treatment, stage at diagnosis, and spatial variables have significant effects ($P \leq 0.05$) on log odds of probability of patient status category at j or below, but not age. The ordered logit (log odds) for patients with no anemia at diagnosis being in a lower patient status category was 0.6034 greater than patients who had anemia at diagnosis. The estimated odds among patients without anemia at diagnosis is 1.828 ($= e^{0.6034}$) times greater than the odds among patients who had anemia at diagnosis, which reflects that the estimated odds of patients who did not have anemia at diagnosis to fall in the lower direction of patient status category was higher by 82.28% of estimated odds of patients who had anemia, keeping other variables constant. This means the patients without anemia were less likely to fall in the higher patient status category than patients with anemia.

The estimated odds of patients who received chemotherapy treatment and patients who received both chemotherapy and

surgery (combined) treatment was 4.284 ($= e^{1.4549}$) and 7.242 ($= e^{1.9799}$) times the estimated odds of patients who received palliative care, respectively. Hence, patients who received chemotherapy and patients who received both chemotherapy and surgery were likely to fall in the lower direction of patient status category than patients who received palliative care. However, patients who received both chemotherapy and surgery were more likely to fall in the lower direction of patient status category than patients who took chemotherapy and palliative care (Table 5).

Stage of patients at diagnosis also has significant effect on the log odds of probability of patient status at j or below category. The estimated odds of patients who had stage I cancer tumor at diagnosis to fall in the lower direction of patient status category was 0.524 ($= e^{-0.6460}$) times the estimated odds of patients who had stage IV cancer tumor at diagnosis, which indicates that the estimated odds of patients who had stage I cancer tumor at diagnosis to fall in the lower patient status category was lower by 47.6% of the estimated odds of patients who had stage I cancer tumor at diagnosis.

It is noted that there is spatial autocorrelation between districts. In Table 7, $P=.0126$ also shows that there was spatial correlation of levels of patient status between districts. The spatial variable correlation with patient status was a negative value, -0.0208, which indicates that districts with lower levels of patient status are usually surrounded by districts with higher levels of patient status.

The goodness of fit of the model was assessed using the Hosmer-Lemeshow test.¹³ The Hosmer-Lemeshow goodness-of-fit test statistic $\chi^2=36.6198$ with its corresponding $P=0.3935$ shows that in the goodness-of-fit model, the data are held.

Discussion

This article investigated the distribution of incidence of cancer disease across districts in the Western Amhara region, Ethiopia, and identified the factors that affect patient status by considering the limited number of patient characteristics from the patient registry card in FHRH. Twenty types of cancer were considered in the study, and of those cancer type, breast cancer was the highest diagnosed cancer type. A similar study conducted in Addis Ababa, Ethiopia, also found that breast cancer was the most common malignant neoplasm among women.²⁴ Among patients who were included in the study and received treatment, only 1.45% were cured and 46.02% improved, whereas the rest have no change in status because of treatment and were even worse after getting treatment.

The incidence of cancer within each district was assessed. To make the incidence of cancer comparable between districts, the population at risk was standardized¹⁰ into the same 100 000 population at risk per district. Incidence of cancer showed significant variation across the different districts included in the study (19.59 per 100 000 population in Finote Selam to 1.11

Table 7. Parameter estimates using intercept-only OLR model, spatial OLR model, and spatial mixed OLR model.

PARAMETER	INTERCEPT-ONLY OLR MODEL		SPATIAL OLR MODEL			SPATIAL MIXED OLR MODEL		
	ESTIMATE (SE)	P VALUE	ESTIMATE (SE)	P VALUE	OVERALL P VALUE	ESTIMATE (SE)	P VALUE	OVERALL P VALUE
Intercept								
Cure	-4.222 (0.4112)	<.0001	-4.3954 (0.5702)	<.0001		-6.2469 (0.9357)	<.0001	
Improved	-0.1013 (0.0983)	.3028	0.1292 (0.4159)	.7561		-1.6466 (0.8322)	.0625	
Same	1.0198 (0.1112)	<.0001	1.6861 (0.4242)	<.0001		-0.03957 (0.8276)	.9624	
Deteriorate	2.7473 (0.2063)	<.0001	3.8532 (0.4754)	<.0001		2.1793 (0.8518)	.0192	
Death*	0.0000		0.0000			0.0000		
Age	-	-	-0.0028 (0.0064)	.6612	.6612	-0.0030 (0.0067)	.6570	.6570
Anemia	-	-						
No			0.3270 (0.0989)	.0009	.0009	0.6034 (0.2029)	.0031	.0031
Yes*			0.0000			0.0000		
Treatment	-	-						
Chemotherapy			0.3329 (0.2194)	.1292		1.4549 (0.6084)	.0173	
Combined			0.8145 (0.2409)	.0007	.0031	1.9799 (0.6336)	.0019	.0035
Palliative*			0.0000			0.0000		
Stage	-	-						
I			-0.1649 (0.2845)	.5622	.0001	-0.6460 (0.5684)	.2564	.0001
II			-0.8235 (0.2352)	.0005		-1.2438 (0.5360)	.0208	
III			-0.3421 (0.2165)	.1141		-0.7285 (0.5293)	.1695	
IV*			0.0000			0.0000		
Spatial variable	-	-	-0.0201 (0.0082)	.0141	.0141	-0.0208 (0.0083)	.0126	.0126
Variance component								
Sigma ₁₁ (σ_{11}^2)	/		/			0.2997 (0.2339)		
AIC (BIC)	1056.067 (1072.180)		924.848 (981.244)			923.040 (937.970)		

Abbreviations: -, the corresponding variable was not included in the model; /, not applicable in the model; *, reference categories; AIC, Akaike information criterion; BIC, Bayesian information criterion; OLR, ordinary logistic regression.

per 100 000 population in Enemay). The spatial autocorrelation, association between the value of incidence of cancer observed in a given district unit and the value of incidence of cancer observed in neighboring district units, was checked using Moran's *I* index and showed a negative autocorrelation. This means that a district in the neighborhood of high incidence of cancer is expected to have low incidence.

It was found that one-third of patients visited the hospital with advanced stage (IV) of cancer. This result which is in line with results obtained in Africa⁷ and specifically in Ethiopia²⁵ tells us that as cancer treatment centers and health service delivery are limited in the country and due to the lack of

awareness of the people about the disease, most cancer patients were diagnosed at an advanced stage. Treating the disease at advanced stage becomes difficult with the available technology and very limited cancer specialists in poor countries, which ultimately leads to most patients becoming incurable from the disease and needing palliative care.

Numerous studies^{18,26} have reported that cancer is a disease of aging, in a sense that the occurrence of cancer becomes high when people get older. However, this was not the case in this study as the incidence of cancer was predominantly concentrated in age between 41 and 55 years. It could be due to that the life expectancy in developed country, which the study by

Bray et al⁴ shows, is quite different from the life expectancy in developing countries like Ethiopia. Mostly in Ethiopia and particularly in our study area, West Amhara people mostly go to religious place while they become old instead of going to hospital to be cured from diseases. In fact, it might be also due to the fact that the population at risk within the age group was not standardized as the population at risk within each age group was not known to make it standardized. On the contrary, age did not have a significant association with patient status, which indicates that the malignancy of cancer tumor is not dependent on the age group to which the patient belongs.

It was found in our study that anemia complication with cancer has a significant effect on patient status. For patients with advanced stage of cancer, complication with anemia was a common manifestation in some other studies such as Danaei et al²⁶ and Benson et al.²⁷ The estimated odds of patients who have no anemia at diagnosis to fall in the lower patient status category equals 1.828 times the estimated odds of patients who have anemia at diagnosis. It means that patients with no anemia complication at diagnosis fall in the lower patient status category compared with those patients who have anemia complication during diagnosis. It was also found that the prognostic factor stage during diagnosis has its own effect on the patients' status after they were given appropriate treatment based on the physician prescription. Different studies such as that by Pakzad¹⁸ repeated that when stage (extent of tumor spread) increased during diagnosis, the patient status will become more likely very bad as curing or improving a patient with advanced stage is difficult. It is advisable to have early detection of the disease so that it becomes simple to decrease the malignancy of the cancer tumor in the given treatment.

ABO blood type is associated with increased risk of lung cancer based on the study by Urun et al.²⁸ In contrast, the study conducted in China²⁹ on gastric cancer explained that there is no significant association between blood group and cancer risk, which is consistent with our study. Hence, it can be recognized that the effect of blood group on cancer risk might depend on the cancer type.

The choice of treatment depends on where the cancer tumor is located, stage of the disease, and health status,²⁷ which was partially consistent with our study that the treatment type has significant effect on the patient status after taking treatment.

This study reveals that the spatial variable has negative significant effect where districts with lower levels of patient status were usually surrounded by districts with higher levels of patient status and that districts with higher incidence of cancer were usually surrounded by districts with lower incidence of cancer, which is in line with studies by Pakzad et al¹⁹ and Shen et al.²⁰ This demonstrated that spatial epidemiology offers insight into ways that individual characteristics, community attributes, and physical environments interact to produce distinctive risk, illness, and disease management patterns.

Conclusion

Patients who did not have anemia complication during diagnosis were more likely to be cured and improved than those patients who had anemia complication. Most of the patients had advanced stage (IV) of cancer tumor, which dismantles the capability of the treatment to be less effective and increases the resistance of malignancy of the cancer tumor, and hence it was perceived that the stage of patients during diagnosis has its own impact on patient status after treatment. The age of the patient when the presence of cancer was disclosed has not significant relationship with the current status of the patient after taking treatment. If the patients are capable of undergoing surgery, on average giving both chemotherapy and surgery treatment was effective for patients to be in a better status. There was significant spatial correlation of incidence of patients with cancer between districts in which districts with high incidence of cancer were usually surrounded by districts with low incidence of cancer and vice versa.

Acknowledgements

We thank the personnel in the hospital who helped us collect the data and provided expert assistance. We would also like to acknowledge Dr Teshome Alene, staff of Medical Health at College of Medicine and Health Science, for technical and professional assistance.

Author Contributions

LMT conceived the idea of the research, conducted data analysis, interpreted the data, and prepared all figures and tables by verifying the analytical method. Both authors discussed the results and wrote the manuscript.

Availability of Data and Material

The data sets of generated analyses during the study were collected from Felege Hiwot Referral Hospital (FHRH), Bahir Dar, Ethiopia, by personnel in the hospital by initiation of Department of Statistics on behalf of Bahir Dar University. Anyone can access the data from the Hospital by proposing reasonable proposal request.

Ethical Consideration

Permission to undertake the study was obtained from Felege Hiwot Referral Hospital (FHRH) by the initiation of Department of Statistics on behalf of Bahir Dar University. The patients/participants were informed that participation was confidential and private information would be protected. Identification of patients/participants was done only through numerical codes, and data collection of the study groups was carried out only when privacy was ensured based on Ministry of Health Legislation in the FHRH.

ORCID iD

Lijalem Melie Tesfaw  <https://orcid.org/0000-0002-2555-8559>

REFERENCES

1. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet.* 2010;11:685–696.
2. National Cancer Institute at the National Institutes of Health. <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>
3. Center M, Siegel R, Jemal A. *Global Cancer: Facts and Figures.* Atlanta, GA: American Cancer Society; 2015.
4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394–424.
5. De Flora S, La Maestra S. Epidemiology of cancers of infectious origin and prevention strategies. *J Prev Med Hyg.* 2015;56:E15–E20.
6. Boscoe FP, Ward MH, Reynolds P. Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. *Int J Health Geogr.* 2004;3:28.
7. Jennifer D, Cathyrne K, Milner D, et al. Africa's emerging cancer crisis: a call to action. <https://bvgh.org/wp-content/uploads/2017/07/Africas-Emerging-Cancer-Crisis-A-Call-to-Action.pdf>. Updated 2017.
8. Gelfand AE, Diggle P, Guttorp P, Fuentes M, eds. *Handbook of Spatial Statistics.* Boca Raton, FL: CRC Press; 2010.
9. Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data* (Vol. 368). Hoboken, NJ: John Wiley & Sons; 2004.
10. Carlin BP, Gelfand AE, Banerjee S. *Hierarchical Modeling and Analysis for Spatial Data.* Boca Raton, FL: Chapman and Hall/CRC Press; 2014.
11. Betts MG, Diamond AW, Forbes GJ, Villard MA, Gunn JS. The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. *Ecol Model.* 2006;191:197–224.
12. Midi H, Sarkar SK, Rana S. Collinearity diagnostics of binary logistic regression model. *J Interdiscip Math.* 2010;13:253–267.
13. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied Logistic Regression* (Vol. 398). Hoboken, NJ: John Wiley & Sons; 2013.
14. Muhammad N, Tuti Purwaningsih S. *Weighted Distance and Optimum Intersection Spatial Logistic Regression for Prediction of Rural Poverty Status in West Java.* Scientific and Academic publishing; 2009.
15. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* New York, NY and Park, CA: Springer-Verlag and SAGE; 2000.
16. Seltman H. *Experimental Design and Analysis.* John Wiley and Sons; 2015.
17. Muhammad N, Tuti Purwaningsih S. Modeling spatial ordinal logistic regression and the principal component to predict poverty status of districts in Java Island. *Int J Statis Appl.* 2013;3(1):1–8.
18. Pakzad R, Ghoncheh M, Pournamdar Z, et al. Spatial analysis of skin cancer incidence in Iran. *Asian Pac J Cancer Prev.* 2016;17:33–37.
19. Pakzad R, Khani Y, Pakzad I, et al. Spatial analysis of stomach cancer incidence in Iran. *Asian Pac J Cancer Prev.* 2016;17:27–32.
20. Shen X, Wang L, Zhu L. Spatial analysis of regional factors and lung cancer mortality in China. *Cancer Epidemiol Biomarkers Prev.* 2017;26:569–577.
21. Brown P, Jiang H, Ezzat S, Sawka AM. A detailed spatial analysis on contrasting cancer incidence patterns in thyroid and lung cancer in Toronto women. *BMC Public Health.* 2016;16:950.
22. Roshandel G, Boreiri M, Sadjadi A, Malekzadeh R. A diversity of cancer incidence and mortality in West Asian populations. *Ann Glob Health.* 2014;80:346–357.
23. Gregorio DI, Cromley E, Mrozinski R, Walsh SJ. Subject loss in spatial analysis of breast cancer. *Health Place.* 1999;5:173–177.
24. Hadg E, Seifu D, Tigneh W, et al. Breast cancer in Ethiopia: evidence for geographic difference in the distribution of molecular subtypes in Africa. *BMC Women's Health.* 2018;18:40.
25. Woldeamanuel YW, Girma B, Teklu AM. Cancer in Ethiopia. *Lancet Oncol.* 2013;14:289–290.
26. Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M; for Comparative Risk Assessment Collaborating Group. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet.* 2005;366:1784–1793.
27. Benson K, Balducci L, Aapro M. Anemia and cancer. In: Balducci L, Ershler WB, Bennett JM, eds. *Anemia in the Elderly.* Boston, MA: Springer; 2008:99–113.
28. Urun Y, Utkan G, Cangir AK, et al. Association of ABO blood group and risk of lung cancer in a multicenter study in Turkey. *Asian Pac J Cancer Prev.* 2013;14:2801–2803.
29. Xiao S, Feng F, Sun L, et al. Blood type AB predicts promising prognosis in gastric cancer patients with positive preoperative serum CEA. *Medicine (Baltimore).* 2017;96:e8496.