Study of prognostic factors in gastric cancer: Application of a cox model and logistic regression

Idrissa SY, M Bousso, AI Correa, MA Loum, A Diop, K Toure, B Traore, AT Diallo and M Dieng

DOI: https://doi.org/10.22271/maths.2022.v7.i5b.887

Abstract

Introduction: Gastric cancer has a poor prognosis. It is the fifth most common cancer in terms of incidence and the fourth most common cancer in terms of mortality worldwide. The aim of our study is to identify the clinical and pathological factors associated with survival and predictive of death.

Method: Our study included 262 patients over a 13-year period from 2007 to 2020 at Aristide Le Dantec Hospital in Senegal and Conakry Hospital in Guinea. The survival period was established from one month after treatment to the date of death or last news. We first used the Cox proportional hazard model to identify independent factors associated with survival. Then we applied logistic regression to these factors to develop the prognostic equation.

Results: The overall survival at 6 months, 1 year and 5 years after treatment was respectively 74%, 57%, 38%. The probability of occurrence of death after 5 years can be estimated with 93% accuracy, by: log \( \text{π} = -19,983 + 2,119 \times \text{Métastases} + 1,268 \times \text{Ulcère gastrique} + 1,100 \times \text{Tabac} + 37,646 \times \text{Epigastralgie} + 37,563 \times \text{Cardiopathie} + 37,220 \times \text{Mucineux}. \)

Conclusion: Our approach contributes to determine the impact of clinical and pathological factors on the survival of patients and the eventual occurrence of death related to stomach cancer.

Keywords: Cancer, stomach, survival, cox regression, logistic regression

Introduction

Stomach cancer or gastric cancer is a public health problem according to the World Health Organization (WHO), \([1]\). It develops in the gastric wall and is generally manifested by weight loss, epigastric pain and vomiting. According to GLOBOCAN data \([2]\), it is responsible for more than one million new cases in 2020 and about 769,000 deaths, ranking fifth in incidence and fourth in mortality worldwide. Its incidence in Senegal and Guinea was estimated at 597 and 271 new cases respectively, with a mortality of 70% and 90% \([2]\). The potentially curative treatment is surgery exclusively or combined with chemotherapy. The standard method for prognostic remains the tumor stage, the presence of lymph nodes or metastases (TNM classification) \([3]\). However, there is no consensus reference neither on survival nor on the occurrence of death, constraining the involvement of clinical and pathological factors. This situation favors a delayed diagnosis and management and very low survival rates, in the order of 20 to 30%, at 5 years \([4, 5, 21]\).

Researchers have made extensive use of logistic regression \([6]\) to identify factors predictive of the occurrence of medical events such as confirmation of pathology \([7, 8]\), the appearance of complications such as metastases or tumor recurrence \([9, 10]\), and death \([11, 12, 13]\). While the Cox proportional hazard model \([14]\) was applied in the presence of survival data to determine the independent, associated parameters \([15, 16, 10, 17, 18]\). The explanatory power of statistical models relies heavily on the identification of the most relevant factors with respect to the event studied. These factors are supposed to be able to explain almost all of the phenomenon. Thus, the adequacy between the research question and the methodology is decisive.

Although logistic regression can predict the occurrence of death, and its model is parametric, it can only be used to predict discrete functions. Since survival times are continuous, this model...
is not sufficient to link the covariates to the instantaneous risk of death. As for the Cox proportional hazard model, it expresses the probability of occurrence of the event studied at a precise moment, knowing that it had not occurred just before (instantaneous risk). However, this model is only partially specific, because the basic risk (in the absence of any explanatory factors) is not defined and it constitutes a nuisance parameter. The objective of our present study is to develop a specific model to predict death from gastric cancer at five years post-treatment, with factors that have a significant and independent impact on survival.

Methods
Type, duration and setting of study: We conducted a retrospective, prognostic study. It spanned a period of 13 years, from 2007 to 2020. The study setting was the general surgery and cancer departments of Aristide Le Dantec Hospital in Senegal and the University Hospital of Conakry in the Republic of Guinea. These are reference facilities in each country with a very large capacity.

Study population and follow-up
The study population was that of patients with gastric cancer. All patients with confirmed gastric cancer who had received treatment (exclusive surgery or combined with chemotherapy) were included in the study. Patients who died within 30 days after treatment were excluded from the study, because death is considered to be related to the surgical procedure and not to the cancer. The database contained 262 patients who met the inclusion criteria. The survival period was established from one month after treatment to the date of death or last news.

Data collection
The data collected were based on the information found in the patients’ files. The parameters studied were: Age, sex, history (Epi gastralgia, Gastric ulcer, Gastric tumor, Tobacco, salt, and alcohol consumption, Arterial hypertension, Diabetes, Heart disease), reasons for consultation (Epigastralgia, Vomiting, Abdominal mass, Constipation), duration of evolution of symptoms, general signs (pale mucous membranes, colored mucous membranes, dehydration, malnutrition), physical signs (epigastric sensitivity, epigastric mass, stasis stomach), biological data (Hemoglobin, Hematocrit, Platelets, Creatine), macroscopic appearance of the tumor (stenosing, ulcerated, budding, hemorrhagic, Infiltrating), differentiation of adenocarcinoma, presence of adenopathy in the abdomen, regional invasion, metastases (Hepatic, Pulmonary, Peritoneal), Preoperative resuscitation (Blood transfusion, Parenteral nutrition, Nasogastric tube), Treatment (Exclusive surgery, Surgery associated with chemotherapy), Death, and survival time at 6 months, 1 year and 5 years.

Statistical analysis
After a description of the general characteristics of the study population. We used the Cox proportional hazard model to identify independent factors associated with survival. Through the hazard function h, defined at each time t by

\[ h(t) = \lim_{h \to 0} \frac{(P(t \leq T < t + h | T \geq t))}{h} \]

This risk function represents the probability that death from stomach cancer will occur at time t knowing that it has not been observed before t (instantaneous risk), and is modeled as follows

\[ h(t; X_1, ..., X_p) = h_0(t)e^{\beta_1 X_1 + \cdots + \beta_p X_p} = h_0(t)e^{\beta' x} \]

where \( h_0 \) is the hazard base function, \( \beta = (\beta_1, ..., \beta_p)^T \in \mathbb{R}^p \) is an unknown regression parameter measuring the association between potential predictors \( X_1, ..., X_p \) and the hazard risk function.

The Cox partial likelihood function is given by:

\[ L(\beta) = \prod_{i=1}^{n} \frac{e^{X_i \beta}}{\sum_{j \in R_i} e^{X_j \beta}} \]

The maximum likelihood estimator of the parameter \( \beta \) is defined as the solution of the problem \( \max_{\beta \in \mathbb{R}^p} L(\beta) \). The appropriateness of the model is assessed by the Likelihood Ratio Test and the Wald Test [14]. For the time-dependent variables, we used the interaction technique [14] to fit them. The proportionality of the risks over time is evaluated by the Schoenfeld residuals test [14]. Finally, we applied logistic regression to the factors retained in the Cox model to predict the occurrence of death five years after treatment.

For each individual \( i = 1 \ldots n \), \( Y_i \) is a binary response variable that indicates: \( Y_i = 1 \) if individual i dies of stomach cancer, and \( Y_i = 0 \) otherwise. The logistic regression model for individual i to die of stomach cancer assumes that the conditional probability \( \pi_i = P(Y_i = 1 | X_i) \) is given by

\[ \text{Logit}(\pi_i) = \log \left( \frac{\pi_i}{1 - \pi_i} \right) = \beta_1 + \beta_2 X_{i1} + \cdots + \beta_p X_{ip} = X_i' \beta \]

where \( \beta \), is a regression parameter that measures the association between the predictors and the occurrence of death. Taking the exponential of equation \( \text{Logit}(\pi_i) \) yields the odds ratios for individual i, given by:

\[ \frac{\pi_i}{1 - \pi_i} = e^{X_i' \beta} \Rightarrow \pi_i = \frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}} \]

This expression defines a multiplicative model for the odds. For example, if the j-th covariate increases by one (or is present) while holding all other constants constant, the risk of death is multiplied its odds.

The log-likelihood for \( \beta \) calculated from the sample \((X_1, Y_1) \cdots (X_n, Y_n)\) is given by:

\[ L(\beta) = \sum_{i=1}^{n} [Y_i \log(\pi_i) + (1 - Y_i)\log(1 - \pi_i)] \]

We proceeded for an external validation of the model. Indeed, 70% (183 patients) of our randomly selected study population was devoted to training and the remaining 30% to validation. The quality of the final model is evaluated by the sensitivity, specificity and the area under the ROC curve (AUC).

Results
We collected 262 patients with gastric cancer who met our inclusion criteria. The average age was 52±11 years. Male patients represented 52%, i.e. a sex ratio of 1.06. The most frequent symptoms were vomiting (77%), malnutrition (76%) and epigastric pain (68%). Gastric ulcer (69%), heart disease (62%) and smoking (50%) were the main antecedents. Adenocarcinoma was mucinous in 30% of cases, adenopathy in the abdomen was observed in 47% and metastasis in 35% of patients. All patients have received treatment, surgery
exclusively for 62%, for the others (38%), surgery was associated with chemotherapy. The overall survival at 6 months, 1 year and 5 years after treatment was respectively 74%, 57%, 38%.

Following a variable selection based on the Wald test at the 5% threshold, the Cox regression model was reduced to the variables shown in Table 1.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>Hazard Ratio (HR)</th>
<th>Wald-stat.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metastases</td>
<td>1.022</td>
<td>2.978</td>
<td>22.126</td>
<td>0.000</td>
</tr>
<tr>
<td>Gastric ulcer</td>
<td>0.993</td>
<td>2.784</td>
<td>17.689</td>
<td>0.001</td>
</tr>
<tr>
<td>Tobacco</td>
<td>0.979</td>
<td>2.661</td>
<td>16.588</td>
<td>0.000</td>
</tr>
<tr>
<td>Epigastralgia</td>
<td>0.839</td>
<td>2.315</td>
<td>12.247</td>
<td>0.000</td>
</tr>
<tr>
<td>Heart disease</td>
<td>0.609</td>
<td>1.838</td>
<td>7.411</td>
<td>0.006</td>
</tr>
<tr>
<td>Mucinous</td>
<td>0.436</td>
<td>1.546</td>
<td>3.966</td>
<td>0.046</td>
</tr>
<tr>
<td>Malnutrition</td>
<td>0.431</td>
<td>1.539</td>
<td>3.779</td>
<td>0.046</td>
</tr>
</tbody>
</table>

Table 1: Estimation of Cox regression model parameters

At a significance level of 5% in the logistic regression model we keep the following variables recorded in Table 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>Odd Ratio (OR)</th>
<th>Standard Error</th>
<th>z-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metastases</td>
<td>2.119</td>
<td>8.322</td>
<td>1.0339</td>
<td>1.979</td>
<td>0.000</td>
</tr>
<tr>
<td>Gastric ulcer</td>
<td>1.268</td>
<td>3.555</td>
<td>0.8617</td>
<td>1.416</td>
<td>0.041</td>
</tr>
<tr>
<td>Tobacco</td>
<td>1.1</td>
<td>3.003</td>
<td>1.112</td>
<td>3.138</td>
<td>0.008</td>
</tr>
<tr>
<td>Epigastralgia</td>
<td>37.646</td>
<td>2.23</td>
<td>174.94</td>
<td>0.014</td>
<td>0.02</td>
</tr>
<tr>
<td>Heart disease</td>
<td>37.563</td>
<td>2.058</td>
<td>1622.2</td>
<td>2.582</td>
<td>0.006</td>
</tr>
<tr>
<td>Mucinous</td>
<td>37.22</td>
<td>1.46</td>
<td>906.54</td>
<td>4.568</td>
<td>0.001</td>
</tr>
<tr>
<td>Intercept (β₀)</td>
<td>-19.983</td>
<td>1.0089</td>
<td>1.857</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Estimation of cox regression model parameters

The equation for the predictive model for the occurrence of death can be written as:

\[
\log \left( \frac{\pi_1}{(1 - \pi_1)} \right) = -19.983 + 2.119* \text{Métablastes} + 1.268* \text{Ulèce gastrique} + 37.646* \text{Epigastralgie} + 37.563* \text{Cardiopathie} + 37.22* \text{Mucineux}.
\]

The quality of the model is measured by the following indicators in Table 3:

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>94.70%</td>
</tr>
<tr>
<td>Specificity</td>
<td>92.50%</td>
</tr>
<tr>
<td>AUC</td>
<td>93.3%</td>
</tr>
</tbody>
</table>

Table 3: Model quality

Discussion

Cancer is one of the leading causes of death and a major obstacle to increasing life expectancy in every country in the world. According to the estimates of the World Health Organization (WHO), in a report on global cancer statistics, [19], there were about 19.3 million new cancer cases and 10 million deaths worldwide in 2020. According to the same report, the global cancer burden is expected to be 28.4 million cases in 2040, an increase of 47% from 2020, with a greater increase in developing countries (64% to 95%) compared to developed countries (32% to 56%).

Stomach cancer, for its part, remains an important cancer in the world and is responsible for more than 1 million new cases in 2020 and about 769,000 deaths, ranking fifth in incidence and fourth in mortality worldwide [19]. In African regions, mortality rates are sometimes higher than the global average while incidence rates are the lowest [20].

Our study included 262 patients with gastric cancer. One hundred and eighty-nine (189), or 72% of the patients were received at the Aristide le Dantec Hospital in Senegal and 73 at the University Hospital of Conakry in Guinea. The absence of national cancer registries in most African countries suggests a low incidence of gastric cancer on the continent. Studies have been conducted in Senegal with 136 histologically confirmed cases in 10 years [21] and 36 cases in 6 years [4], in Guinea Conakry with 13 cases in 6 years [22], in Cameroon with 120 cases in 6 years [15] and in Congo Brazzaville with 41 cases in 9 years [5]. According to GLOBOCAN estimates, stomach cancer incidences in Senegal and Guinea were 597 and 271 new cases respectively in 2020 [2]. The lack of a reliable data collection system in this area makes it possible to under estimate these incidences. The mean age of the patients in our series was 52 years with extremes ranging from 26 to 82 years. The variability according to age, with a standard deviation of 7 years, was not significant, which explains that subjects of almost all age groups are concerned by gastric cancer. The same is true for the sex ratio which was 1.06. Men and women are affected without significant difference. This trend is also found in African series [15, 21, 4, 22, 5]. On the other hand, in Western and Asian series, stomach cancer occurs most often in patients of advanced age and a strong male predominance [23, 24].

To identify independent parameters, associated with survival, we used the Cox proportional hazard model. The $\beta$ coefficients or Hazard ration (HR) measure the association between each covariate studied and the instantaneous risk function for death. This approach is similar to that used by G.A. Bang et al. [15] in their study on the clinical epidemiology and risk factors for mortality of gastric cancer in sub-Saharan Africa. They used the Cox model to determine the respective hazard ratios (HR) with 95% confidence intervals. Results showed that variables independently associated with survival included: WHO stage 3 performance status (HR = 1.935, p = 0.042), palpable epigastric mass on examination (HR = 1.909, p = 0.042), and liver metastases (HR = 2.310, p = 0.012). G. Capelli et al. [16] took a different approach with the Cox model. Their aim was to evaluate the performance of a nomogram to distinguish the parameters associated with survival. The performance of the nomogram was calculated by Harrell's C-index and calibration plots [16]. The results showed that tumor location, lymphnode ratio, and pT stage were associated with overall survival of gastric cancer...
cancer patients (Index-C 0.75).

After identifying the factors independently and significantly associated on survival of patients in our series, we applied the predictive model \[26\] of logistic regression to them. This allowed us to develop a prognostic equation for the occurrence of death related to gastric cancer with five (5) years of follow-up. Indeed, for each potential predictor, the probability of death is estimated by the odds ratio. This corresponds to the number of times that death is likely to occur in the presence of the factor considered. Thus, the most relevant predictors are selected to constitute the prognostic equation. M. Sylla et al. \[11\], adopted the same procedure to identify and evaluate pre-, intra- and post-operative histo prognostic factors in terms of occurrence of local recurrence, metastasis and overall survival with or without recurrence for patients with gastric cancer. The results showed that at the 5\% threshold, the occurrence of metastasis was an unfavorable prognostic factor \(\text{OR}=27.36\), whereas reinvention increased the chances of survival \(\text{OR}=0.07\). Choney Zangmo et al. \[25\] also sought to identify factors that influence the occurrence of death for cancer patients (all cancers combined). They used logistic regression. The best fitted model was obtained from the deviance analysis at the 5\% threshold. Poor prognostic factors: age, length of stay and cancer site.

It should be noted that logistic regression can be performed in a variety of other ways depending on the research objectives and the data available.

Direct logistic regression is used when no specific assumptions are made about the order or importance of the predictor variables. In fact, this type only considers the contribution of the top predictors \[21\].

Descriptive logistic regression allows the importance of a phenomenon to be measured and its profile to be traced according to a number of variables. As a result, this model tends to suggest hypotheses rather than confirm them, given the exploratory nature of the study \[30\].

Automated logistic regression ("Stepwise") is used to generate hypotheses. Indeed, it is mainly used when the field of research is less explored and the knowledge of possible predictors is limited \[25\].

The main feature of our methodology lies in the application of logistic regression to predict death from gastric cancer on factors already identified by the Cox model as independently associated with survival. The variables retained in our final equation thus provide two pieces of information at once. S. Xu et al. \[30\] adopted a similar methodology to investigate the prevalence of Hashimoto’s thyroiditis in adult papillary thyroid cancer patients and its association with cancer recurrence and outcome. The primary outcome was the association of Hashimoto’s thyroiditis (HT) with papillary thyroid cancer (PTC) mortality, assessed using Cox proportional hazards regression models. The secondary outcome was the association of HT with aggressive features and structural recurrence of PTC, assessed using logistic regression with and without adjustment for related factors. According to Cox proportional hazards regression models, HT was associated with decreased PTC mortality after adjustment for sex, age, primary tumor size, ETE, metastasis, extra ganglionic extension \(\text{HR}, 0.19; 95\% \text{CI}, 0.05-0.76; \text{ P}=0.02\). Logistic regression showed that HT was negatively associated with frequencies of primary tumor size of 4 cm or more \(\text{OR}, 0.16; 95\% \text{CI}, 0.10-0.33; \text{ P value} < 0.001\), ETE \(\text{OR}, 0.79; 95\% \text{CI}, 0.71-0.88; \text{ P value} < 0.001\), gross ETE \(\text{OR}, 0.36; 95\% \text{CI}, 0.30-0.44; \text{ P value} < 0.001\), lateral neck metastases \(\text{OR}, 0.74; 95\% \text{CI}, 0.66-0.84; \text{ P}\ <$ 0.001\), extra ganglion extension \(\text{OR}, 0.52; 95\% \text{CI}, 0.43-0.62; \text{ P}\ <$ 0.001\), and distant metastases \(\text{OR}, 0.11; 95\% \text{CI}, 0.03-0.43; \text{ P}=0.002\). The results suggest that autoimmune Hashimoto thyroiditis has a protective role in association with thyroid cancer.

Our study identified gastric ulcers, smoking and heart disease as antecedent or antecedent factors leading to gastric cancer, with a significant impact on survival and poor prognosis. Our results are supported by those found by J.-D. Korwin et al. \[27\] who showed in their study that patients with gastric ulcers most often develop gastric cancer. The common denominator of these two pathologies is the infection with the bacterium Helicobacter pylori. Chronic gastritis caused by the bacterium can lead to ulcers and in the long term, transform the mucosa and induce mutation of it scells \[27\].

Analyzing the role of echocardiography in the detection and monitoring of left ventricular dysfunction in patients undergoing cancer treatment, C. Charbonnel et al. \[24\] found that heart disease was a comorbid factor that negatively affects the quality of life and prognosis of cancer patients. Smoking is considered a contributing and aggravating factor for all cancers \[4, 5, 15, 29, 30\].

The symptomatology of gastric cancer is very polymorphic, and often a source of diagnostic delay and poor prognosis \[10\].

Epigastric pain is an independent sign associated with long-term survival of patients in our series. The prognosis is poor in the presence of this symptom. Our study also revealed a significant impact of under nutrition on survival, which is not the case for its association with prognosis. In the series of G.A. Bang \[15\], the frequency of a palpable epigastric mass on examination was highlighted in patients with very short survival times. In the series of B. Diop et al. \[14\] of patients followed for gastrectumor, the symptoms were mainly weight loss, signs of pyloricstenosis and hematemesis.

Gastric adenocarcinoma is the most common histological type of gastric cancer \[21\]. The occurrence of metastases is identified by our study as adversely affecting survival and is a very poor prognostic factor. This is characteristic of a very advanced state of the disease. Hepatic, pulmonary or peritoneal invasion is noted. Research on cancer all agree on this subject \[4, 5, 15, 21, 11, 29, 30\].

Age and sex were not retained in our study as independent factors, significantly associated neither with survival nor with the occurrence of death. This explains why the survival of patients of both sexes and of all ages is affected in an almost similar way. The study conducted by M. Sylla et al. \[11\], on the histoprognostic factors of gastric cancer, showed on the other hand that an advanced age negatively affected the survival of patients with gastric cancer.

Table 4: General characteristic of the study population

<table>
<thead>
<tr>
<th>Variables</th>
<th>Average</th>
<th>[Min-Max]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>52</td>
<td>[16 - 82]</td>
</tr>
<tr>
<td>Duration of evolution of symptoms (months)</td>
<td>7</td>
<td>[1-24]</td>
</tr>
<tr>
<td>Blood platelets (×1000)</td>
<td>364</td>
<td>[10-911]</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td>Percentage</td>
</tr>
</tbody>
</table>

“111”
Table 5: General characteristic of the study population

<table>
<thead>
<tr>
<th>Test</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute death [2 - 6 months]</td>
<td>69</td>
</tr>
<tr>
<td>69</td>
<td>44</td>
</tr>
<tr>
<td>69</td>
<td>113</td>
</tr>
<tr>
<td>193</td>
<td>149</td>
</tr>
<tr>
<td>Survival rate 692/262 (74%)</td>
<td>149/262 (57%)</td>
</tr>
</tbody>
</table>

Table 6: Results of the tests of significance of the cox model

<table>
<thead>
<tr>
<th>Test</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood ratio test</td>
<td>2e-16</td>
</tr>
<tr>
<td>Wald test</td>
<td>2e-16</td>
</tr>
</tbody>
</table>

Conclusion

The prognosis of gastric cancer is poor. The overall survival at 5 years was 38% for the patients of our survey. Our methodology consisted in the use of the Cox proportional hazard model to identify the independent parameters associated with survival. We then applied logistic regression to these parameters to predict the occurrence of death from gastric cancer at 5 years after treatment. Our approach helps to determine the impact of clinical and pathological factors on survival and the eventual occurrence of death for patients with gastric cancer.

However, a good data collection system would provide more information and will contribute to the implementation of a benchmark.

References

1. OMS, Organisation Mondiale de la santé. [en ligne], 2022 [consulté le 30 juin 2022]. Disponible sur: https://www.who.int


