Evaluation of parametric models by the prediction error in colorectal cancer survival analysis

Ahmad Reza Baghestani¹, Mahmood Reza Gohari², Arezoo Orooji¹, Mohamad Amin Pourhoseingholi³, Mohammad Reza Zali³

¹Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
²Department of Biostatistics, Iran University of Medical Sciences, Tehran, Iran.
³Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran

ABSTRACT

Aim: The aim of this study is to determine the factors influencing predicted survival time for patients with colorectal cancer (CRC) using parametric models and select the best model by predicting error’s technique.

Background: Survival models are statistical techniques to estimate or predict the overall time up to specific events. Prediction is important in medical science and the accuracy of prediction is determined by a measurement, generally based on loss functions, called prediction error.

Patients and methods: A total of 600 colorectal cancer patients who admitted to the Cancer Registry Center of Gastroenterology and Liver Disease Research Center, Taleghani Hospital, Tehran, were followed at least for 5 years and have completed selected information for this study. Body Mass Index (BMI), Sex, family history of CRC, tumor site, stage of disease and histology of tumor included in the analysis. The survival time was compared by the Log-rank test and multivariate analysis was carried out using parametric models including Log normal, Weibull and Log logistic regression. For selecting the best model, the prediction error by apparent loss was used.

Results: Log rank test showed a better survival for females, BMI more than 25, patients with early stage at diagnosis and patients with colon tumor site. Prediction error by apparent loss was estimated and indicated that Weibull model was the best one for multivariate analysis. BMI and Stage were independent prognostic factors, according to Weibull model.

Conclusion: In this study, according to prediction error Weibull regression showed a better fit. Prediction error would be a criterion to select the best model with the ability to make predictions of prognostic factors in survival analysis.

Keywords: Parametric model, Prediction error, Apparent loss, Colorectal cancer.

Introduction

Colorectal cancer is the third most common cancer and cause of cancer death worldwide (1). The incidence and mortality of colorectal cancer are rising rapidly in Asian countries (2-4).

Gastrointestinal cancers are the most important causes of mortality in Iran, which is located in the Middle East, Asia, and the burden of these cancers are increasing (5, 6) and CRC is not an exception (7). The prognosis of CRC is relatively good in terms of survival. Over the past 60 years, numerous claims have been made of variables being related to survival of colorectal cancer. Several studies have also considered independent prognostic factors,
including age at diagnosis (8), sex (9), stage (10), histological grade (8) etc. Survival models or failure time models are statistical techniques to estimate the overall time up to specific events and find the related factors or predict the outcome. Prediction is important in medical science, because doctors need to estimate the survival of patients to choose the best treatment and it helps one to know about the disease condition in the future (11). The accuracy of prediction is determined by a measurement, generally based on loss functions, called prediction error. Recently, this technique developed to estimate the prediction error in survival analysis, in order to find the best model for analyzing the prognosis factors (12, 13). The aim of this study is to determine the factors influencing predicted survival time for patients with colorectal cancer using parametric models and select the best model by predicting error’s technique.

**Patients and Methods**

The data belongs to registered patients with colorectal carcinoma who admitted to the Cancer Registry Center of Gastroenterology and Liver Disease Research Center, Taleghani Hospital, Shahid Beheshti University of Medical Sciences; Tehran, Iran, in the period between 2002 to 2007. All patients were followed until January 1, 2007 (as failure time) from their diagnosis by telephone contact (14, 15). Each patient was informed by a consent form for documenting his/her information in the Cancer Registry Center. The data of 600 patients who were followed at least for 5 years and have completed information selected for this study. Body Mass Index (BMI), Sex, family history of CRC, tumor site (colon, rectum), stage of disease (early, advanced) and histology of tumor (Mucinous, others) included in the analysis. The survival time was compared by the Log-rank test and multivariate analysis was carried out using parametric models including Log normal, Weibull and Log logistic regression. For selecting the best model, the prediction error by the apparent loss (12) was used in which smaller error indicates a better model. P<0.05 was considered as statistically significant and all analysis carried out using R software (16).

**Table 1.** The result of log rank test for Univariate analysis

<table>
<thead>
<tr>
<th>Prognostic Factors</th>
<th>Number of death</th>
<th>Number of patients</th>
<th>Mean (SE)</th>
<th>Test statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 25</td>
<td>113(31.2)</td>
<td>362(60.3)</td>
<td>71.565(5.055)</td>
<td>28.55</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>More than 25</td>
<td>38(16)</td>
<td>238(39.7)</td>
<td>128.032(9.719)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>96(23)</td>
<td>418(69.7)</td>
<td>109.166(6.99)</td>
<td>4.936</td>
<td>0.026</td>
</tr>
<tr>
<td>Rectum</td>
<td>55(30.2)</td>
<td>182(30.3)</td>
<td>77.32(6.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family history</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>88(25)</td>
<td>345(57.5)</td>
<td>74.557(6.523)</td>
<td>2.034</td>
<td>0.154</td>
</tr>
<tr>
<td>No</td>
<td>63(24.7)</td>
<td>255(42.5)</td>
<td>115.178(6.812)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>57(22.3)</td>
<td>256(42.7)</td>
<td>111.503(9.348)</td>
<td>3.87</td>
<td>0.049</td>
</tr>
<tr>
<td>Female</td>
<td>94(27.3)</td>
<td>344(57.3)</td>
<td>88.114(8.334)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early</td>
<td>57(18.8)</td>
<td>303(50.5)</td>
<td>117.436(7.768)</td>
<td>25.537</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Advance</td>
<td>94(31.6)</td>
<td>297(49.5)</td>
<td>68.118(4.938)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucinous</td>
<td>18(28.1)</td>
<td>64(10.7)</td>
<td>105.113(14.006)</td>
<td>0.766</td>
<td>0.381</td>
</tr>
<tr>
<td>Others</td>
<td>133(24.8)</td>
<td>536(89.3)</td>
<td>99.799(6.369)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

Among 600 patients, 344 were men (57.3%) and 256 were women (42.7%). Among 151 patients who died, 62.3% were men. The mean of survival for patients was 105.08 months (95% CI: 950.5-115.1) and the median was 94.5 months (95% CI: 58.6-130.4). Log rank test showed a better survival for females, BMI more than 25, patients with early stage at diagnosis and patients with colon tumor site (Table 1).

All factors included in parametric models (Log normal, Weibull and Log logistic censored regression) and prediction error by apparent loss was estimated for each model consequently, which resulted as the 1.46 for Log logistic, 1.49 for Log normal and 1.28 for Weibull. Therefore, Weibull model was the best model among these parametric models (Table 2) and revealed that BMI and stage of disease were independent prognostic factors of CRC survival. The relative risk of death for patients in the advanced stage of disease is 2.27 times more compared to patients is in the early first stage and patients with low BMI (less than 25) were at higher risk of death, compared to those with BMI more than 25. The other variables were not significant.

Table 2. The results of Weibull Censored Regression for Multivariate analysis

<table>
<thead>
<tr>
<th></th>
<th>RR</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>2.2</td>
<td>0.16</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Tumor site</td>
<td>1.17</td>
<td>0.14</td>
<td>0.2</td>
</tr>
<tr>
<td>Family history</td>
<td>1.12</td>
<td>0.14</td>
<td>0.4</td>
</tr>
<tr>
<td>Sex</td>
<td>1.13</td>
<td>0.14</td>
<td>0.3</td>
</tr>
<tr>
<td>Stage</td>
<td>1.96</td>
<td>0.14</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Histology</td>
<td>1.08</td>
<td>0.21</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Relative Risk, 2. Colon as the reference group, 3. Others were as the reference group.

Discussion

In this study, BMI and stage of the disease were prognostic factors of CRC survival, according to parametric regression model and the apparent loss prediction error indicated that Weibull model was the best option among parametric models to analyze the survival of CRC patients who admitted in Taleghani hospital.

However, in the Log rank test, sex was a significant factor, multivariate model showed no relation between sex and survival of CRC patients. A population study on about 165,000 CRC patients in Germany reported a better survival for women (17).

According to the histology type of tumor, Log rank test and Weibull model showed no difference in survival. This result is consistent to some studies (18, 19).

The Log rank analysis revealed a better survival of colon cancer, compared to rectal’s. However, this result was not significant in multivariate analysis. Other studies reported a better survival for colon cancer (20, 21). People with rectal cancer tend to be older and may have other serious health issues. Therefore, it would be the reason of different survival.

Family history of CRC was another risk factor in our analysis. Although individuals with a family history of colorectal cancer are diagnosed more often than the general population (22), the study suggests that survival from colorectal cancer may not be worse (23) and the result of this analysis in both univariate and multivariate confirmed that.

BMI was a prognostic factor of CRC survival in both Log rank and Weibull analysis and the patients with higher BMI had a better survival. A similar study suggested that underweight and obese women with colon cancer were at increased risk of death (24).

In multivariate and univariate analysis, the effect of the cancer’s stage was significant on survival time. A similar study indicated that patients whose cancer is in the early stage have a better survival time (25).

In this study, we used a parametric model to analyze the survival rate of patients with CRC and select the best-appropriated model using prediction error. Parametric models are more flexible than Cox semi parametric model (26-28).
Besides, prediction error would be criteria to select the best model with the ability to make predictions of prognostic factors in survival analysis.

Acknowledgment

This paper is a result of Ms Thesis in Biostatistics, ShahidBeheshti University of Medical Sciences.

References

23. Kirchhoff AC, Newcomb PA, Trentham-Dietz A, Nichols HB, Hampton JM. Family history and

24. Doria-Rose VP, Newcomb PA, Morimoto LM, Hampton JM, Trentham-Dietz A. Body mass index and the risk of death following the diagnosis of colorectal cancer in postmenopausal women (United States). Cancer Causes Control 2006;17:63-70.


