Cox Proportional Hazard Regression Model in Prediction of Chronic Renal Failure Progression – Preliminary Study

Krzysztof Szajowski¹, Sławomir C. Zmonarski², Marian Klinger²
Anna Puziewicz-Zmonarska ³

¹Instytut of Mathematics, Wrocław University of Technology
²Department of Nephrology, Wrocław University of Medicine,
³Department of Pediatric Nephrology, Wrocław University of Medicine.

ABSTRACT
Our topic was to verify the usability of Cox proportional hazard regression model in evaluation of the probability that an individual chronic renal failure (CRF) patient would require renal replacement therapy within a specified period of time. Material: database that included 225 records describing progression of CRF in 15 patients from the time of their first contact with our nephrologist to the outset of severe renal insufficiency including 13 cases at the beginning of renal replacement therapy. Our analysis started from the estimation of separate clinical and laboratory parameters and proved their known usefulness. The analysis of combined prognostic value of different laboratory and clinical parameters gives the possibility to verify their relative prognostic power and offers a potential support to the nephrologist. Reliability of such models strongly depends on the number of parameters included.

INTRODUCTION
Renal replacement therapy still offers a nearly unique possibility to maintain patients with severe renal failure alive. But before they reach the moment when their renal function needs to be substituted they receive standard care at an outpatient clinic. The regimen includes regular examination by a nephrologist and a set of lab tests. Taken together, this data provide the physician with a limited amount of information concerning blood pressure, blood cell count, serum creatinine, urea, electrolytes, total cholesterol and routine urine analysis with daily protein excretion. Because a progressive loss of renal function is often observed even when the underlying disease is no longer active, the treatment is focused on neutralising the influence of other factors that contribute to disease progression. On the other hand, it is quite well known that the rate of CRF progression may be differently or even not sensitive to such therapeutic interventions and these patients in a period of months may require renal replacement therapy.

Our topic was to verify the usability of Cox proportional hazard regression model in evaluation of the probability that an individual CRF patient would require renal
replacement therapy within a scheduled period of time.

**MATERIAL**

The data taken into consideration were collected from patient records stored at Outpatient District Nephrology Clinic. They provided a limited amount of information about chronic renal failure progression in 17 cases. In 13 (6-men; 7 – women) cases the data covered a period from their first visit to the beginning of renal replacement therapy. No patient received erythropoietine. In 10 cases renal failure was caused by chronic glomerular disease and in 3 cases by chronic interstitial disease.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>First symptoms of disease (age) [years]</td>
<td>32.8</td>
<td>16.3</td>
</tr>
<tr>
<td>First examination [years]</td>
<td>37.6</td>
<td>17.9</td>
</tr>
<tr>
<td>Age at the moment of first dialysis [years]</td>
<td>50.8</td>
<td>15.0</td>
</tr>
<tr>
<td>Time of observation [months]</td>
<td>124.2</td>
<td>67.8</td>
</tr>
</tbody>
</table>

**METHOD PRESENTED**

The method named "survival analysis" is used in the paper. More precisely, one of model constructed to the survival analysis is adopted. The term "survival analysis" originated in the study and analysis of time elapsed till death (i.e., survival times) for medical patients diagnosed with some fatal disease. Survival analysis is now a well-developed field of statistical research and methodology pertaining to modelling and testing hypotheses of failure time data for humans as well as animals, machines, electronic equipment, automobile components, etc. Hence, the methodology is far more general than the analysis of survival time. In fact, fields of study other than medicine have given other names to the identical methodology discussed here. Other names of the survival analysis are analysis of failure time data, reliability analysis, event history analysis. Roughly speaking, such models could be described as follows. Let T be the time from some starting moment to the characteristic event.

This time depends on some covariates. Our aim is to find a model to describe the covariates influence on the distribution of T. Such modelling of survival times is based on two distinct approaches - parametric and nonparametric. The nonparametric methods, widely used in clinical trials, include Kaplan-Meier estimates of the survival distribution (i.e. P(T>t)), Cox proportional hazards regression models. The mentioned models take into account the pertinent data as well. In some generalised model we have time-dependent covariates or time dependent strata.

In the described data we have some information concerning patients. The aim is to construct the mathematical model, on the basis of which we are able to predict the moment of the first dialysis (the fatal disease). The Cox regression model with time dependent covariates will be used. For fixed time of observation, without censoring, one can adopt the logistic regression.

Let T be the time from beginning of observation to the moment of the first dialysis. The clinical data gives a possibility to estimate the probability that the first dialysis will take place no earlier than after 75 weeks. Let p=\( P(T>75) \). Based on the regression model \( p=f(q_1,q_2,\ldots,q_k) \) one can obtain the tool for estimation the probability of the first
dialysis not earlier than after 75 weeks for new patients. By mathematical reason the function of $p$ is estimated. The most frequently used function is

$$\text{logit}(p) = \log\left(\frac{p}{1 - p}\right)$$

The more complex approach is based on the survival analysis models. The survival function is defined as follows:

$$S(t \mid x) = P\{T \geq t \mid x\}$$

$x$ is the vector of covariates. It is an equivalent approach to determine and estimate the hazard function

$$h(t) = \frac{f(t)}{S(t)}$$

where $f(t)$ is the density function of the survival time.

The survival function is used to predict the probability that the first dialysis will take place not earlier than after moment $t$.

**DETAILED DESCRIPTION OF THE MODEL USED,**

The data for the survival analysis have the particular form. Among observed values is time variable - the time elapsed till death, development of a particular symptom, or relapse after remission of disease. The observation of the patient can be finished by a defined event (death, the first dialysis) or by a lack of contact with the patient caused by other reasons (such case is named censoring of the data). As it was mentioned above, we would like to know the survival function as the regression dependence on the covariates, which could be constant or time-dependent. The time dependence of covariates makes additional difficulties in the forecast of time of the first dialysis.

The particular character of the data for the survival analysis is:

- the survival time of a group of similar individuals will tend to be positively skewed
- the survival time is frequently censored

The analysis of such data can be done by

- the estimation of the survival function (equivalently - the hazard function);
- the construction of the regression model for the survival function with covariates (creatinine level, blood pressure);

$\Rightarrow$ constant in time;

$\Rightarrow$ time dependent.

For the forecast of the first diagnosis moment one can consider the logistic regression model as well as the Cox proportional hazards model with time dependent covariates.

**LOGISTIC REGRESSION**

Let $Y$ be a random variable with two values: 0 when $T < t_0$ and 1 otherwise. We are interested event if the survival time is longer than $t_0$. Precisely, we would like to find the dependence of the probability of the event that $T > t_0$ on the observed covariates. Let $x=(x_1, x_2, ..., x_n)$ be the set of covariates of the patient and $p=P(Y=1\mid x)=S(t_0 \mid x)$ denote the probability that the survival time is longer than $t_0$. The transformed value of $p$ is
modelled as follows.
\[
\logit(p) = \log\left(\frac{p}{1 - p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n
\]

The value of \( p \) we obtain from relation.
\[
p = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n}}
\]

When the parameters \( \beta_1, \beta_2, \ldots, \beta_n \) are known and the values of \( x_1, x_2, \ldots, x_n \) for the patient are given, the probability \( P(T > t_0) \) can be calculated.

**THE COX PROPORTIONAL HAZARD MODEL**

The survival function can be described as follows:
\[
S(t) = \exp\{-H(t)\}
\]

where
\[
H(t) = \int_0^t h(u)du
\]

The Cox proportional hazard model is the regression model for the survival time data. It assumes that the hazard function \( h(t, Z_i) \) for \( i \)-th patient has the form:
\[
h(t, Z_i) = h_0(t) r_i(t)
\]

where \( r_i(t) = e^{\beta Z_i(t)} \), and \( h_0(t) \) are the estimated baseline hazard functions (see [2], 149-198, 223-236).

The estimation of the parameters of the model can be done by *coxph* procedure of S-Plus package. The forms of the data for this procedure are given in Table 2 and the result of calculations are in Table 3. The analysis of models with many covariates needs a high performance computer.

Table 2. The order of data required for performing coxph (Splus) procedure.

<table>
<thead>
<tr>
<th>id</th>
<th>SP</th>
<th>Name</th>
<th>Sex</th>
<th>start</th>
<th>stop</th>
<th>l-st visit.</th>
<th>Glom.RD</th>
<th>Interst.RD</th>
<th>RRsyst.</th>
<th>Dialysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 171 8 C.M.</td>
<td>0</td>
<td>0</td>
<td>204</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>130</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 172 8 C.M.</td>
<td>0</td>
<td>204</td>
<td>207</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>140</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 173 8 C.M.</td>
<td>0</td>
<td>207</td>
<td>215</td>
<td>11</td>
<td>0</td>
<td>1</td>
<td>150</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 174 8 C.M.</td>
<td>0</td>
<td>215</td>
<td>219</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>160</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 175 8 C.M.</td>
<td>0</td>
<td>219</td>
<td>221</td>
<td>17</td>
<td>0</td>
<td>1</td>
<td>160</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 176 8 C.M.</td>
<td>0</td>
<td>221</td>
<td>226</td>
<td>22</td>
<td>0</td>
<td>1</td>
<td>140</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 177 8 C.M.</td>
<td>0</td>
<td>226</td>
<td>231</td>
<td>27</td>
<td>0</td>
<td>1</td>
<td>120</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. One of coxph (Splus) calculation products that refers toward I model.

Urea >15

<table>
<thead>
<tr>
<th>coef</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creatinine</td>
<td>0.812</td>
<td>2.25</td>
<td>0.356 2.28 0.023</td>
</tr>
<tr>
<td>exp(coef)</td>
<td>exp(-coef)</td>
<td>lower .95 upper .95</td>
<td></td>
</tr>
<tr>
<td>Creatinine</td>
<td>2.25</td>
<td>0.444</td>
<td>1.12 4.53</td>
</tr>
</tbody>
</table>

n= 11

<table>
<thead>
<tr>
<th>coef</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creatinine</td>
<td>0.783</td>
<td>2.19</td>
<td>0.792 0.989 0.32</td>
</tr>
<tr>
<td>exp(coef)</td>
<td>exp(-coef)</td>
<td>lower .95 upper .95</td>
<td></td>
</tr>
<tr>
<td>Creatinine</td>
<td>2.19</td>
<td>0.457</td>
<td>0.463 10.3</td>
</tr>
</tbody>
</table>

Figure 1. Relative prognostic value at serum urea concentration in a two component model (incl. serum urea and creatinine). The relative influence of creatinine concentration on the possibility achieve in the moment of first dialysis declines along the increase of urea concentration). The horizontal axis presents time [months].
Figure 2. Relative prognostic value of different ranges of serum creatinine concentration in three component model (blood hemoglobin - systolic pressure – serum creatinine. Horizontal axis presents time [months].

COMMENT

Our analysis started from the estimation of separate clinical and laboratory parameters. At this point the results were not surprising. They proved the predictive value of serum levels of creatinine, erythrocyte blood count, haemoglobin, serum urea nitrogen and marginally hyperpotasemia. As opposed to statistical analysis performed by Locatelli (published last year in Nephrology, Dialysis and Transplantation) we can not attribute the same importance to the degree of hypertension but in our study we analysed systolic and diastolic pressure but not mean blood pressure. The same problem was found in the case of proteinuria, but the outcome was probably strongly influenced by inaccurate estimation of proteinuria provided by routine general urine tests. Due to a low number of patients we could not estimate the role of underlying renal disease.

All the tested variables may act as cofounders of one another; therefore the analysis of the effect of all the possible factors should simultaneously take many or even all of them into account. Our first choice was an analysis of combined predictive value of different ranges of serum urea with the full range of serum creatinine. Generally, serum creatinine is believed to have a higher prognostic value than urea nitrogen. The significance of prognostic value of urea nitrogen increases in parallel to the increase of its serum concentration. When the serum concentration of urea becomes higher than
30 mmol/l, urea becomes a more important prognostic factor than creatinine. The mean follow-up of our patients was 74.4 months with standard deviation 65 months. It means that in this model the probability that a particular patient who has been observed due to CRF for 75 months will require renal replacement therapy is modified strongly by the serum urea level. In the area of urea concentration higher than 15 mmol/l the probability is near 60% and when the urea concentration is higher than 30 mmol/l the probability is over 80%.

In the second model the full range of blood haemoglobin and systolic pressure together with different ranges of creatinine level has been taken into account. This three-element model depicts the predictive significance of creatinine level. In this model the probability that a particular patient followed-up 75 months will require hemodialysis changes meaningfully from 60% in moderate up to 75-85% in advanced renal failure. The third model that includes the full range of serum creatinine, systolic pressure and different ranges of serum potassium shows a very weak influence of potassium concentration on the prognosis of renal failure progression.

CONCLUSIONS

1. There are several routine laboratory parameters that can be used independently as predictive factors of renal failure progression. With decreasing order of confidence they are as follows: creatinine (p=0.0036), erythrocyte blood count (p=0.004), blood hemoglobin (p=0.0061), urea nitrogen (p=0.052). Due to a low number of patients the predictive value of blood pressure could not be confirmed.

2. The confidence of prognosis of disease progression increases when a simultaneous analysis of different variables is performed. This makes, the prognosis of CRF progression more precise. These models also help to estimate the relative predictive value of separate constituents.

3. The estimation error of survival function increases with time. It was shown on projections by dotted lines.

REFERENCES


