Mathematical Statistics
Stockholm University

Proportional hazards and additive regression analysis of survival for severe breast cancer

Anna Törner

Examensarbete 2004:3

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Examensarbete **2004:3**,
http://www.math.su.se/matstat

Proportional hazards and additive regression analysis of survival for severe breast cancer

Anna Törner*

February 2004

Abstract

The primary objective of this work was to investigate and compare the use of the Cox proportional hazards model and Aalen's additive model in analysing survival data.

Survival data from a study of 52 patients with advanced breast cancer was investigated using the Cox proportional hazards model. The model was optimized by examining different aspects by use of appropriate residual plots. Covariates judged not to improve the model significantly were removed. The model was stratified with regard to tumour size to account for different baseline hazards.

After optimizing the Cox model, the same data was used to fit an additive model according to Aalen. Plots of the martingale residual process and Arja's plot was used to check model fit and optimize model options.

The information gained from fitting of the two models is similar in some respects but also quite different in others. Both procedures resulted in the same covariates selected to remain in the model. The Cox model yield easily interpreted estimates of the covariates effects, but the assumption of proportional hazard is necessary to make these estimates valid. The additive model and the plots of the cumulative regressions functions give an appealing understanding of how the hazard profile is distributed. Most often however, these cumulative regression functions do not easily transform into a single numerical estimate of the covariate effect.

*Postal address: Swedish Institute for Infectious Disease Control, SE-171 82 Solna, Sweden. E-mail: anna.torner@smi.ki.se. Supervisor: Mikael Andersson.

Chapter 1

Introduction

Breast cancer is the most common malignant disease for females in northern Europe and North America, corresponding to an age-corrected annual incidence of 100 to 120 per 100.000 females. The median age for new breast cancer diagnosis is 60 to 64 years. One out of 11 women will develop breast cancer in their life time. Breast cancer primarily effects women older than 50 years but even if the absolute incidence in women aged 20- 40 years is low, breast cancer constitutes about 24 percent of new cancers in this age group [Blåboka 2003]. This present work has aimed to investigate the use of two statistical models to model survival of patients with locally advanced breast cancer. Some emphasis has also been put on trying to explore some information on mutations in the MDR - and GST gene complex. The work has been divided in the following parts:

- To build a traditional Cox proportional hazards model to describe which factors influence the survival of these patients
- To fit an additive regression model to the same data with the same purpose
- To compare the results derived using the Cox proportional hazards model with the results from fitting the additive model

Data from a clinical study has kindly been made available by the Department for Genetics at the Norwegian Radium hospital, Oslo.

Chapter 2

Pathophysiology

2.1 Prognostic Factors

A number of factors with great prognostic value for breast cancer have been identified. The investigation of theses factors are important for prognosis and for making treatment decisions. For a cancer with a bad prognosis more aggressive treatment regimes may be chosen and the patient may be willing to accept more severe side effects. Some of the most commonly discussed prognostic factors are:

Estrogen- and Progesterone receptors About 60 percent of primary breast cancers contain estrogen receptors and the levels are usually greater in postmenopausal women than in premenopausal. If the breast tumor has positive receptors for both estrogen and progesterone the chances of response are greater. Hormone receptor negative tumors generally have a more severe prognosis and more aggressive chemotherapy may be warranted.

The clinical and histopathological TNM staging , primary tumor size, regional lymph node status combined with presence or absence of distant metastases provides important prognostic information.

Tumor size is categorized in four groups, $T1 - T4$ with $T1$ being tumors smaller than 2cm. $T3$ are tumors larger than 5cm. $T4$ classification means tumor of any size but with direct penetration to skin or breast wall. Tumor size is evaluated by radiologic examination. The patients in this study all had tumors in group T3 or T4. One could say that the differences between T1, T2 and T3 is merely a quantitative difference compared to the difference to T4 which is more of a quali-

tative difference indicating a more severe diagnosis for the T4 patients [Blåboka 2003].

Metastatic disease either in lymph nodes (N+) or distant metastatic disease (M+) predict a more severe prognosis.

Histologic grade describes how well differentiated the tumor cells are. If the cells are highly differentiated it indicates that the cells are more mature and that the cell cycle is less out of control compared to tumors made up of highly undifferentiated cells. Histologic grade is evaluated by microscopic examination of cells.

P53 is a tumor suppressor gene and mutations in this gene has been associated with increased risk of cancer for a number of different cancer types. In a situation with DNA damage $p53$ will stop the cell cycle in $G1$ phase awaiting DNA repair or aphthosis (cell programmed death).

Multi Drug Resistance gene is implicated in drug resistance towards many drugs. The MDR gene is coding for trans-membranous proteins acting as "pumps" effectively transporting substances such as chemotherapy out of the target cells. Mutations in the MDR gene has therefore been hypothesized to predict an increased chance of effect of chemotherapy.

The GST gene complex is implicated in detoxification of exogenous substances (substances that does not belong to the biological system).

2.2 Classification in Stages

Breast cancer is classified in four stages based on occurrence of different risk factors. These stages are used internationally to classify the disease for treatment decisions and prognosis. Since this classification system is truly international it is convenient to compare different studies and regimes used in different countries. Stage I disease is the least advanced stage and 5 years survival, i.e. the proportion of patients being alive after 5 years, is about 90 percent. On the other end of the scale is the most advanced stage, stage IV, where five years survival is about 30 percent [Blåboka 2003].

To classify the disease in different stages, a number of prognostic factors are investigated and the overall distribution of these prognostic factors decided with stage disease is present. It is therefore not always true that a patient with a large tumor is a stage IV patient, other factors are also important to consider.

2.2.1 Locally Advanced Breast Cancer - Stage III Disease

Patients with Locally advanced breast cancer may have tumors of size $T3$ or $T4$ and/or disease with extensive metastases to the lymph nodes, but no distant metastases. Five year relapse free survival is about 30 percent and the prognosis severe [Blåboka 2003]. A treatment option is pre-treatment by chemotherapy followed by surgery and/or radiotherapy. This setting is characterized as neoadjuvant. The patients in the precent study were all patient with locally advanced disease.

Chapter 3

Description of the Data Set

The data set was originally made up of 92 patients with breast cancer, stage III. For reasons beyond control of the author of this work only data for 52 patients were made available for statistical analysis. This selection of the patient group may not be representative for the original group of 92 patients. If the purpose of this work had been to accurately describe the prognosis for these patients this possibly skewed selection of patients might have been disadvantageous. However, given the fact that the original group already is a highly selected group and that the purpose of this work is to compare the use of different statistical models, this possible non random selection of patients should not affect the statistical work adversely.

The characteristics of these patients are described in table 3.1 below.

All patients had large tumors ($T3/T4$) and were treated with chemotherapy pre-operatively to facilitate surgery and to get a more complete removal of tumorous tissue at the time of operation. After surgery, Relapse Free Survival (RFS), and Overall survival (OS) in months were recorded. Relapse free survival was defined as time without recurrence of disease. For estimation of RFS there are 25 right censored observations and 27 patients who have experienced relapse. The RFS times (uncensored) varied from $0-59$ months. The longest observation time is 92 months (censored observation). Information on prognostic variables are not complete and the number of patients included in each analysis is therefore a little bit different.

Table 3.1: Characteristics of patients in study, n=52

Covariate	**Label**	**Value**
Estrogen receptor value	ER	Mean=77 , range(0-733)
Progesterone receptor value	PgR	Mean=124, range(0-1150)
Total number of mutations in MDR gene	Mdr1mut	Mean=1.7, range(0-3)
Mutations in p53 gene	p53	Yes: n=23 No: n=29
Tumor stage T3 or T4	T3/T4	T3: n=37, T4: n=15
Histological grade 2	Hist2	Yes: n=22, No: n=30
Histological grade 3	Hist3	Yes: n=21, No: n=31

Chapter 4

Cox Proportional Hazard Regression Model

The Cox proportional Hazards model is probably the most widely used method for modelling survival data, Cox (1972). For data with one explanatory variable, i.e. one covariate, non-parametric methods like plotting of Kaplan-Meier survival probabilities may be adequate if the groups being compared are reasonably similar. Frequently however, the groups being compared differ in many respects. They may have different age distributions, different proportion of men and women, different smoking habits etc. These differences come in addition to the covariate we are really interested in, and the analysis must be adjusted to compensate for these other differences, which may otherwise confound the analysis. The Cox proportional hazards model is a semi-parametric model for fitting survival data.

The basic model is as follows:

$$h(t \mid \mathbf{Z}) = h_0(t) \cdot \exp(\beta^t \mathbf{Z})$$

where $h_0(t)$ is the baseline hazard which may vary arbitrarily over time, and $\mathbf{Z}$ is the covariate vector. The covariates may be time-dependent but are here assumed to be fixed at the start of study. $\beta = (\beta_1, ...\beta_p)$is a vector of covariate coefficients. The baseline hazard is treated non-parametrically, but the individual covariate effects $(\beta_p)$are assumed to be constant throughout the study, hence the notation semi-parametric. The model is often called the proportional hazards model because of this constant covariate effect throughout the study. If two individuals are compared that have covariate values $\mathbf{Z}$ and $\mathbf{Z}^*$ the ratio of their hazard rates at *any* time point simplifies to:

$$\frac{h(t|\mathbf{Z})}{h(t|\mathbf{Z}^*)} = \frac{h_0(t)\exp[\Sigma_{k=1}^p \beta Z_k]}{h_0(t)\exp[\Sigma_{k=1}^p \beta Z_k^*]} = \exp[\sum_{k=1}^p \beta_k (Z_k - Z_k^*)]$$

This ratio is constant or "proportional" throughout the study. This assumption greatly facilitates the interpretation of covariate effects, as the effect of a given covariate compared to the absence of that covariate is expressed as a single constant. This does not however imply that the absolute difference between the two individuals discussed above is constant; the exponentiated covariates act multiplicatively on a baseline hazard which may vary freely.

4.1 Univariate analysis

Univariate analysis revealed the following factors to be of significant importance ($p < 0.1$) in predicting survival (variables with smallest p-value listed first): Oestrogen receptor status, Mutations in gene $p53$, Progesterone receptor status, Indicator variable for histological grade 3, Tumor stage ($T3$ or $T4$).

Indicator variable for Histological grade 2 was not significant ($p = 0.21$), but will be investigated in the model together with indicator variable for histological grade 3.

Age, total number of mutations in the GST gene complex and total number of mutations in the $MDR1$ gene all had p-values above 0.1 and will not be considered for inclusion in the model. Age will often be an important predictor in breast cancer. The individuals in this data set are however a very selected group with other prominent risk factors, such as advanced disease, and this may the explanation for age being non-significant as prognostic factor. Mutations in the GST gene complex also appears to be a non-important predictor in this group and these two variables will also be excluded a priori from investigation in the additive model later.

4.2 Cox model with several covariates

Fitting of the multivariate Cox proportional hazards model was conducted by starting with a model with all variables listed as significant above. One by one the least significant variable was removed until only significant variables remained in the model. Data for relapse free survival and overall survival were modelled in the same way. Basically the same variables are important for predicting both relapse free survival and overall survival. Relapse is however the event occurring first and logically the explanatory variables would therefore be more closely related to relapse free survival compared to overall survival. Overall survival is more prone to be influenced by additional fac-

Table 4.1: Results of preliminary fitting of Cox model

Covariate	**exp(coef)**	**se(coef)**	**Z**	**p**
$T3/T4$	3.469	0.45822	2.71	0.0066
$Hist2$	3.874	0.79985	1.69	0.0900
$Hist3$	5.929	0.80480	2.21	0.0270
PgR	0.993	0.00317	-2.13	0.0330
$p53mut$	2.305	0.42739	1.95	0.0510

Table 4.2: Comparison of models with and without indicator variables for histologic grade 2 and histologic grade 3

Model	**Deviance**	**Number of explanatory variables**
Model with $Hist2$ and $Hist3$	24.4	5
Model without $Hist2$ and $Hist3$	17.6	3

tors occurring after relapse, e.g. treatment after relapse. Therefore only the results for the end point relapse free survival will be presented here, unless there are contradictory or exceptional findings for overall survival. The initial fitting of the Cox proportional hazard model yielded the following parameter estimates (table 4.1).

Histologic grade is coded as two indicator variables, $Hist2 = 1$ corresponds to histologic grade 2 and $Hist3 = 1$ corresponds to histologic grade 3. This preliminary model contains the variables tumor stage, histological grade, $p53$ mutation and progesterone receptor level. $Hist2$, the indicator variable for histological grade 2 is not significant in itself but $Hist3$ may not be included in the model without the inclusion of $Hist2$ also.

A model with the variables $Hist2$ and $Hist3$ may be compared directly by assessing two times the difference in partial log likelihood, here denoted as deviance, for the two models divided by the differences in the number of variables. This simple comparison is possible because the models are nested (the model without $Hist2$ and $Hist3$ is a simpler "case" of the richer model) [Klein & Moeschberger 1997].

Difference in deviance:$(24.4 - 17.6) = 6.8$. The difference in deviance is approximately χ^2 distributed with 2 degrees of freedom, i.e. $p = 0.0334$. The difference in deviance for each model is calculated as the deviance compared to a model with all β equal to 0. This indicates that the inclusion of $Hist2$

and $Hist3$ significantly improves the model. This corresponds well to the clinical setting where histological grade is usually considered to be indicative of tumor aggressiveness and prognosis.

4.2.1 Model checking - preliminary model (A)

This first model has been fitted without considering the best functional form of the continuous variables (Progesterone receptor level) and without questioning the underlying assumption of proportional hazard. The fit of this preliminary model was therefore investigated by examining the following residual plots.

- The functional form of the continuous variable Progesterone receptor was investigated by examining Martingale residuals.
- The proportional hazards assumption was investigated by examining scaled Schoenfeld residuals.
- Score residuals were used to investigate the influence of individual observations.

4.2.2 Functional form of continuous covariates

The functional form of the covariate Progesterone receptor level needs to be checked. From a medical viewpoint it would seem logical to transform the variable in some way to decrease the positive influence of very high values. It is beneficial to be receptor positive, i.e. having a relative high receptor count, but clinical experience indicates it is not 10 times as beneficial to have a value of 1000 as a value of 100. Introducing a cut-off point for being classified as "receptor positive" or a log transformation would be logical options. To approach the problem more methodically I have chosen to use Martingale residuals to try to determine the correct functional form of the progesterone covariate [Klein & Moeschberger 1997].

If the data are right-censored and all the covariates are fixed at the start of the study the martingale residual may be defined as:

$$\hat{M_j} = \delta_j - \hat{H}_0(T_j) \cdot \exp(\Sigma_{k=1}^{p} Z_{jk} b_k), j = 1, 2, ...n \tag{4.1}$$

$T(j)$is the time point, when censoring or event occurs, for the j:th individual. b(k) are the estimated covariate coefficients. $\hat{H}_0(t)$ is the estimator of the cumulative baseline hazard, δ is 1 or 0 depending on if the event of interest has occurred. The residuals may be interpreted as the difference over

time of the number of observed events minus the number of expected events. The residuals are constructed as follows. The covariate for which we are interested in investigating the functional form is assumed to be independent of the other covariates. A Cox model with all covariates excluding the covariate to be investigated is fitted. The martingale residuals are plotted against the covariate to be investigated.A LOWESS smooth may be used to reduce the noise level. This plot will reveal if and how the martingale residuals change with increasing values of the covariate that is investigated.If the plot is linear, no transformation is needed. Including the untransformed covariate will in the model together with the other covariates yield an appropriate regression coefficient. If however there appears to be a discrete time point where the slope changes, a dichotomized transformation of the covariate may be indicated [Klein & Moeschberger 1997].

Looking at a martingale residual plot for our data, for the progesterone receptor covariate, the smoothed curve is roughly linear up to about a progesterone value of about 100 and then levels off. This suggests that progesterone receptor level may be coded as an indicator variable. The indicator variable P is coded as follows:

$$P = \begin{cases} 0 & \text{if } PgR < \Theta \\ 1 & \text{if } PgR \geq \Theta \end{cases} \tag{4.2}$$

The cut-off value Θ is chosen from the values of (PgR) in the data set. A profile likelihood may be plotted for each PgR value in the data set and the Θ value yielding the highest value of the log-likelihood is chosen. Here values for PgR in the data set from 64 to 179 were tried "manually" as cut-off points for the indicator covariate P. A cut-off point of 85 yielded the smallest p-value for the covariate and the smallest p-value (log-likelihood) for the full model. After dichotomizing the Progesterone receptor covariate, 35 patients had $P = 0$ and 17 patients had $P = 1$.

The original covariate PgR in the model will therefore be substituted by an indicator variable P with cut-off point 85.

Details of covariate coefficients for the model with dichotomous progesterone values are displayed in table 4.5.

4.2.3 Propotional hazard assumption

The proportional hazard assumption was examined for the variables $p53$ mutations, tumor stage ($T3/T4$), Histological grade ($Hist2$ and $Hist3$) and the Progesterone indicator variable (P).

The S-plus default for checking the proportional hazards assumption is a formal test and plot of scaled Schoenfeld residuals[Venables & Ripley 1999].

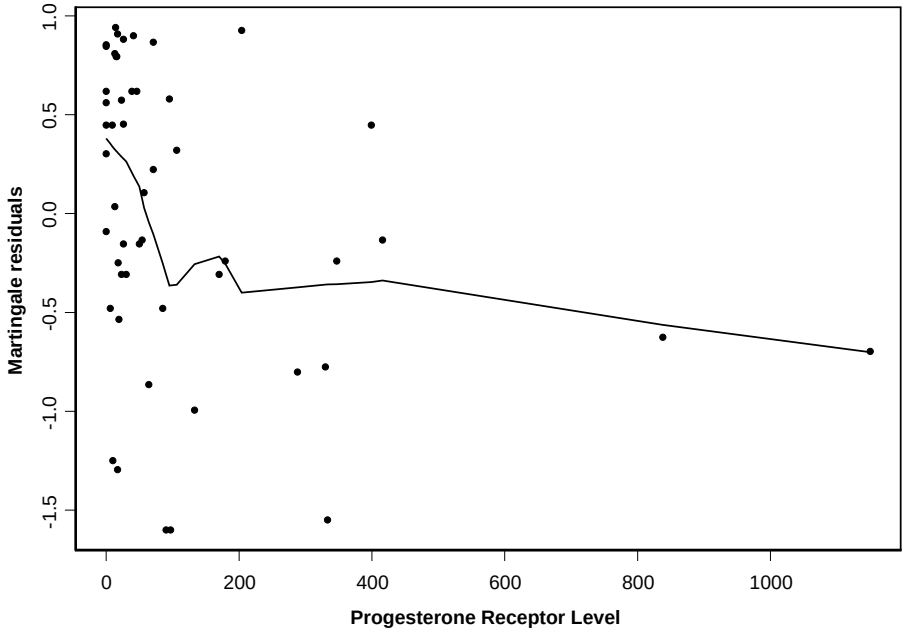

Figure 4.1: Martingale residuals Progesterone receptor covariate

Table 4.3: Covariate coefficients for Cox model with dichotomized progesterone covariate

Covariate	**coef**	**exp(coef)**	**se(coef)**	**Z**	**p**
$T3/T4$	1.057	2.879	0.445	2.38	0.017
P	-0.982	0.375	0.486	-2.02	0.043
$Hist2$	1.236	3.441	0.802	1.54	0.120
$Hist3$	1.713	5.543	0.795	2.15	0.031
$p53$mut	0.990	2.692	0.426	2.32	0.020

Table 4.4: Test of proportional hazards assumption

Covariate	**rho**	**chisq**	**p**
P	-0.0554	0.0813	0.7756
$Hist2$	-0.2193	1.1735	0.2787
$Hist3$	-0.3116	2.3423	0.1259
$p53$mut	-0.1044	0.3206	0.5712
$T3/T4$	-0.3788	3.9570	0.0467
GLOBAL	NA	5.5174	0.3560

The scaled Schoenfeld residual is the difference between the covariate at the failure time and the expected value of the covariate at this time. The residual plot and test investigates departures of the type

$$\beta(t) = \beta + \theta * g(t)$$

for some smooth function g. The function g(t)used here is the S-plus default, $g(t) = 1 - S(t)$.

The Schoenfeld residual is defined as follows:

$$S_{jk} = \delta_j[Z_{jk} - \overline{Z}_k(T_j)] \tag{4.3}$$

$\overline{Z}_k$ is the average covariate value (at time T_j). The residuals are calculated at failure times and the ticks given on the x-axes in the plots are the actual failure times. The results of the test of the proportional hazards assumption is presented in the table 4.4.

The only variable displaying a significant deviation from the proportional hazards assumption is the $T3/T4$ variable, i.e. tumor stage 3 or 4. The scaled Schoenf+eld residuals for the variables are plotted in fig 4.2, together with a smooth. When the proportional hazards assumption holds, a relatively straight horizontal line is expected. Again, the deviation from the proportional hazards assumption is clearly detectable for the $T3/T4$ variable.

A plot of the survival probabilities for the $T3$ and $T4$ shows that the curves are not parallel but diverges, figure 4.3. This is also consistent with the general clinical perception of stage $T4$ and $T3$, $T4$ is a more advanced stage of the disease and the progress of the disease is inevitable. For $T3$ the prognosis is more uncertain and some patients may have somewhat longer life expectancy. The survival probabilities do not start from 1.0 because some patients relapse immediately after surgery. This is most evident for $T4$ patients where the tumor tissue sometimes may not be completely removed

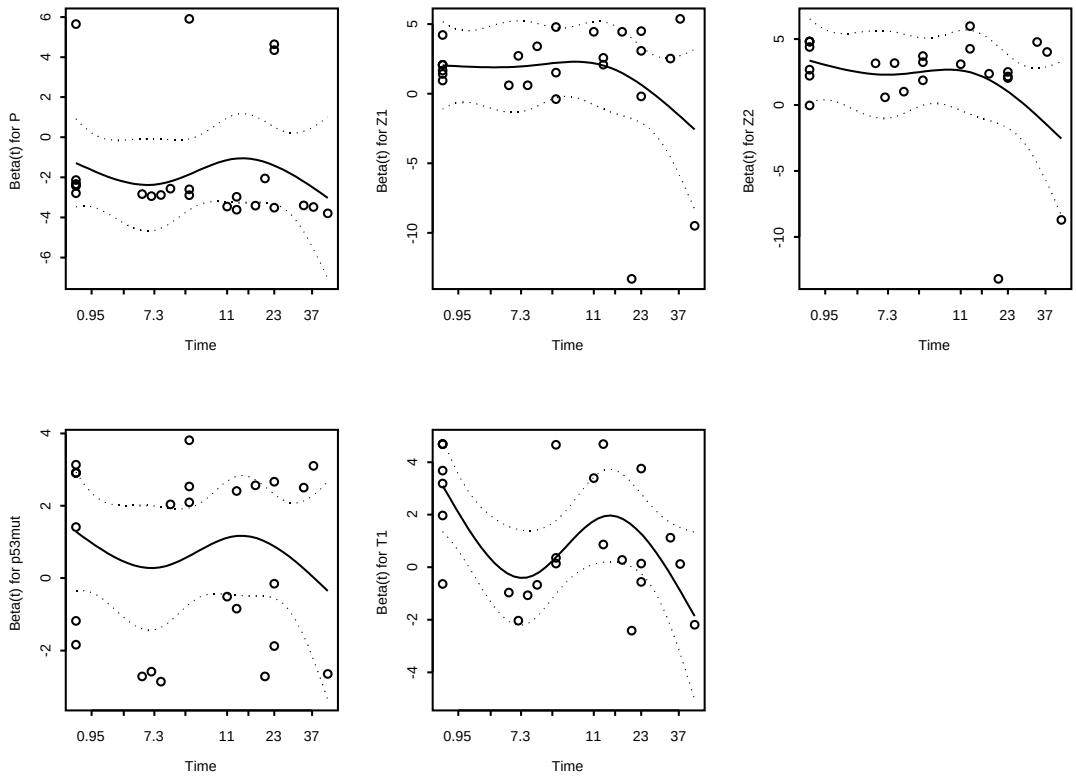

Figure 4.2: Scaled Schoenfeld Residuals

and the tumor continues to progress despite chemotherapy immediately after surgery.

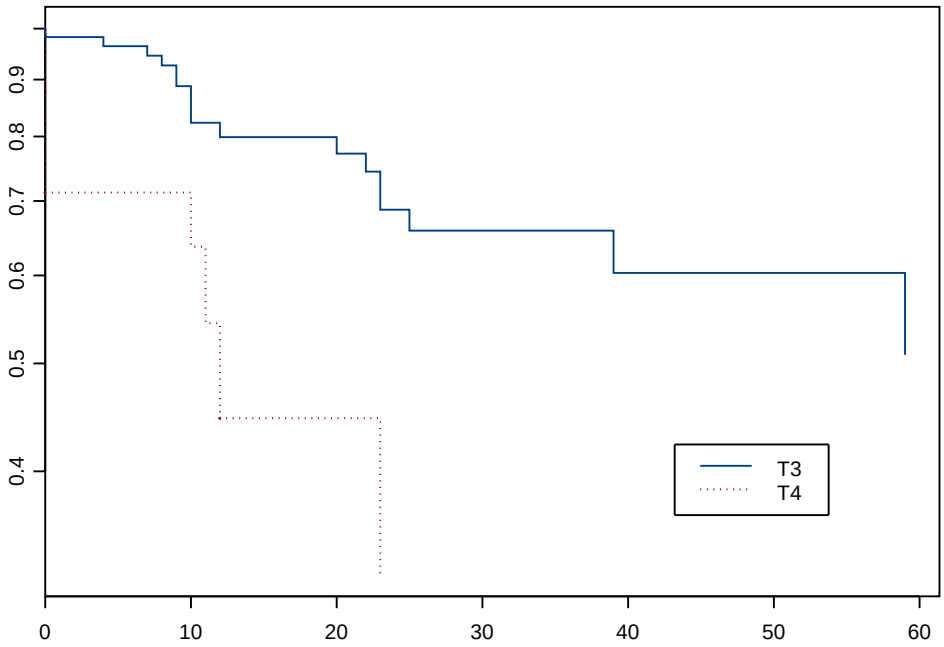

Figure 4.3: Survival probabilities for $T3/T4$ patients

The different baseline hazards for $T3/T4$ patients indicate that a model stratified on tumor stage is suitable. The estimates for the final model, stratified on different tumor stage, is displayed in table 4.5. Stratification on tumor stage means that the covariate coefficients presented are the same for $T3$ and $T4$ patients. The baseline hazard however is different for $T3$ and $T4$ patients which means that the exponentiated $\beta \cdot \mathbf{Z}$ acts multiplicatively on different baseline hazards for T3 and T4 patients. The relative difference between T3 and T4 patients is thus the same for absence/presence of a given covariate but the absolute difference may differ due to different baseline hazards.

Table 4.5: Coefficients for stratified model

Covariate	coef	exp(coef)	se(coef)	z	p
P	-1.001	0.367	0.486	-2.06	0.039
$Hist2$	1.052	2.864	0.786	1.34	0.180
$Hist3$	1.454	4.280	0.774	1.88	0.060
$p53$mut	0.919	2.508	0.425	2.16	0.030

Table 4.6: Individuals with large deviations in score residuals

Patient no	RFS	Histologic grade	Hist2	Hist3	Tumor stage
5	39	1	0	0	$T3$
13	80	1	0	0	$T4$
19	59	1	0	0	$T3$

4.2.4 Influence of individual observations

Influence of individual observations may be studied by the use of score residuals. The optimal way of examining the influence of an individual observation on the estimate b for a given covariate is to estimate β from all data and thereafter estimate $\beta(i)$, i.e β estimated with the i'th observation deleted from the sample. If $\beta - \beta(i)$ is close to zero the individual observation has little influence on the estimate. Below in fig 4.4 score residuals, scaled by the standard error, are displayed for the covariates in the model.

The size and distribution of the residuals seem reasonable at visual inspection. A few noticeable outliers are discussed individually below.

Three patients, no. 5,13 and 19, were outliers for the two indicator variables for histologic grade 2 and histologic grade 3 (Hist2 and Hist3). These patients have the following properties.

The residuals are negative for patients 5 and 19 and positive for patient 13. This indicates that observation 5 and 19, when included give a negative contribution to indicator variables $Hist2$ and $Hist3$. The other way around patient 13 gives a positive contribution to $Hist2$ and $Hist3$. These are reasonable conclusions since patient 5 and 19 had a relatively short Relapse Free Survival (RFS) considering they have the lowest histological grade. The opposite is true for patient 13; here RFS was longer than would be expected from the model.

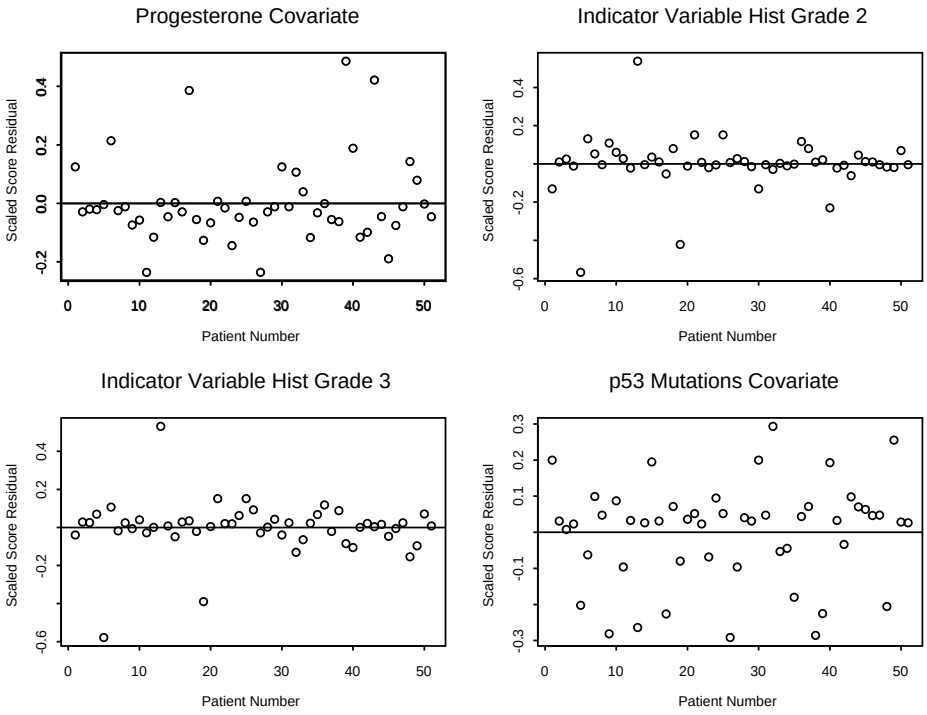

Figure 4.4: Scaled score residuals

Overall the residuals seem reasonable and no subjects will be considered from exclusion from the analysis. An interesting observation is that the progesterone covariate, before the transformation into the (0,1) variable had residuals approaching 2 standard deviations (not shown here). An influence this extensive could indicate that some observations are weighted to heavily. After the transformation the maximum deviation is of the order 0.4-0.45 standard deviations.

Chapter 5

Additive Hazards Regression Model

In the Cox model the covariates are assumed to act multiplicatively on a baseline hazard. The baseline hazard is the hazard for individuals with covariate values equal to zero and this hazard is a function of time. The model is semi-parametric in the sense that constant proportional hazard through the study is assumed. In some cases this assumption of constant proportional hazard may not always be valid.

An alternative to the Cox model, which does not condition on constant proportional hazard, is a linear model, proposed by Odd Aalen, [Aalen 1989]. In this model the covariates are modelled as additive risks to a baseline hazard and allowed to vary freely over time.

5.1 Short description of Aalen's additive regression model

A number of individuals are observed over time to see if a specified event occurs. The individuals are assumed to be independent and any events happening to the individuals are also assumed to be independent between individuals. The life times we observe may be complete or right censored. The basic equation may be formulated as follows

$$h(t|Z) = \alpha_0(t) + \Sigma_{j=1}^{k} \alpha_j(t) Z_j(t) \tag{5.1}$$

The hazard at any time is thus a *sum* of a baseline hazard $\alpha_0(t)$ and a linear combination of the covariate values, Z_j.

The rest of this section will make use of Aalen's terminology. To model the intensity of an event occuring, let $\lambda_i(t)$ denote the intensity (probability) of the event occuring at time t for individual i, given that it has not occured before, n is the number of subjects and k is the number of covariates in the analysis.

The vector of intensities is modelled as follows:

$$\lambda(t) = Y(t)\alpha(t) \tag{5.2}$$

The matrix Y is of size $n \cdot (k+1)$ and is constructed as follows; if the i'th individual is a member of the risk set at time t then the i'th row of Y(t) is the vector $Z^i(t) = (1, Z_1^i(t), Z_2^i(t), ...Z_k^i(t))'$. The vector $Z_j^i(t), j = 1, ...k$ are covariate values, which may be time dependent. If the individual is not at risk at time t, i.e. the event of interest has already occurred or the individual has been censored, the corresponding row in Y(t) contains only zeros. Please not that equations 5.1 and 5.2 are equivalent.

The first element of the vector $\alpha(t)$ is a baseline parameter (intercept) and the remaining elements are called regression functions and estimate the influence of the covariates. These regression functions are the equivalents to the regression parameters (β) in the Cox regression model. In contrast to the Cox model (where the β are constant) the regression functions in the additive model may vary arbitrarily with time. It is unpractical and difficult to estimate the individual regression functions and instead the cumulative regression functions are estimated. The cumulative regression functions are defined as:

$$A_j(t) = \int_0^t \alpha_j(t)ds$$

If $T_1 < T_2 < ...$ are the ordered event times, i.e. the times when an actual event, not censoring, occurs, an estimator of A(t) is given by:

$$A^*(t) = \Sigma_{i=1}^n X(T_i)I_i \tag{5.3}$$

I_i is a vector of zeros except for a one corresponding to the individual experiencing an event at time T_i X(t) is a generalized inverse of Y(t) and a choice based on local least square principles is:

$$X(t) = [Y(t)'Y(t)]^{-1}Y(t)'$$

An estimator of the cumulative intensity is given by:

$$\Lambda^*(t) = A^*(t)'\mathbf{Z} \tag{5.4}$$

Z is the vector of covariate values. The cumulative regression functions are plotted against time and give a description of how the covariates influence the survival over time. It is therefore the change in the cumulative functions, e.g. the slope that is of primary interest.

If the covariate values are standardized, i.e. the mean subtracted before estimation, the baseline regression function (or intercept) gives a description of the cumulative intensity for an individual with average covariate values.

A feature of the additive model which should be noted is that the intensities $\lambda_i(t)$ are not naturally restricted to non-negative numbers. This could lead to the survival function not being monotone over time, but may increase for some values of t. The survival function is given by:

$$P^*(t) = \exp(-\Lambda^*(t))$$

The property of not necessarily being monotonically decreasing follows from equation 5.2 and 5.3 above. For appropriately chosen covariates and data sets of reasonable size this should not be a pronounced problem.

Statistical software for fitting of additive regression models for S-plus /R has been developed by Odd Aalen and Harald Fekjaer, Institute for medical Statistics in Oslo. The choices of generalized inverse, estimation of intensities, choice of weight function etc. are in accordance with the presentation of the model above.

5.1.1 Testing

Test statistics have been developed for the additive model. One primary question may be if a specific covariate has any influence on the distribution of life times. This corresponds to the following null hypothesis:

$$H_j : \alpha_j(t) = 0 \quad \text{for all} \quad t$$

Index (j) corresponds to the jth covariate in the analysis. This null hypothesis may only be tested over the time interval where Y(t) has full rank. A test statistic for H_j is given by the jth element U_j of the vector

$$U = \sum_{T_i} K(T_i)X(T_i)I_i$$

$K(t)$ is an $(r+1)*(r+1)$ diagonal matrix of weight functions. $X(T_k)I_k$ is easily recognized as the cumulative regression function estimator A^* and the test statistic is therefore simply a weighted summation of the cumulative regression function estimator for all event times.

One choice of weight function suggested by Aalen is:

$$K(t) = \{diag[(Y(t)'Y(t))^{-1}]\}^{-1}$$

The weight given to each estimate using this weight function is proportional to the inverse of its variance.This is also the weight function being used in the software for the fittings in this study. The test statistic using this weight function K will be denoted TST. Another alternative mentioned by Aalen is to weigh each estimate according to the number of individual at risk at this time point.This possibility will not be explored further here.

The covariance matrix of U is estimated by the formula

$$V = \sum_{T_i} K(T_i)X(T_i)I_i^D X(T_i)'K(T_i)$$

I_i^D is a diagonal matrix with I_i as elements.

The test statistic is therefore a weighted combination of the individual $\alpha(j)$, i.e. a weighted version of the cumulative regression function.

5.2 Fitting of the additive model

The data described above, with covariate information for oestrogen receptor, progesterone receptor, tumor stage III/IV, indicator for histological grade 2 and 3, $p53$ mutations and number of mutations in $MDR1$ gene, was fitted as an additive model in the the software described above. The results are presented graphically below with time on the x-axis and cumulative regression functions on the y-axis. The dotted lines indicate a 95 percent pointwise confidence interval.

Above are the cumulative regression estimates for estrogen - and progesterone receptor level. Both covariates seem to have protective effects (negative slope of cumulative excess risk curve), more pronounced for the estrogen receptor but with a wider confidence interval. The effect of progesterone receptor level is statistically significant, $p = 0.04$. The estrogen receptor values and the progesterone receptor values are highly correlated and inclusion of both variables in the model would only explain marginally more compared to choosing one of the variables. The confidence interval for the progesterone cumulative regression function is narrower and the estimate appear more precise. The logical choice for inclusion in the model is therefore the progesterone receptor level.

Indicator variable for histological grade 2 and histological grade 3 show an excess risk compared to histological grade 1. The excess risk is slightly more

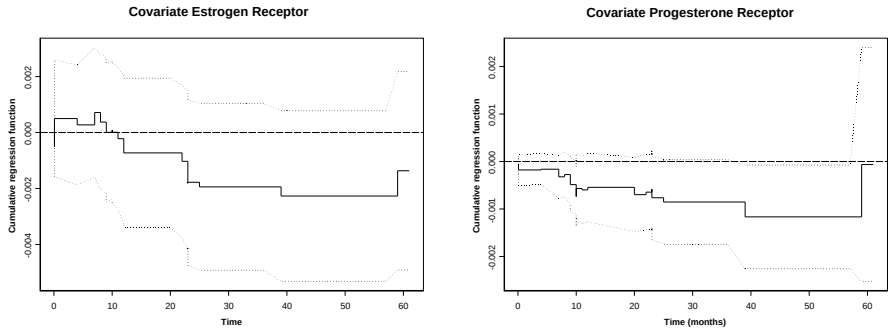

Figure 5.1: Cumulative regression functions for Estrogen receptor and Progesterone receptor value.

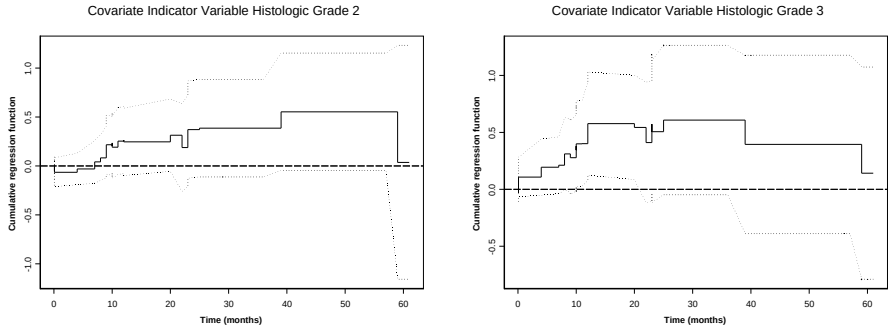

Figure 5.2: Cumulative intensity functions for Histologic grade 2 and Histologic grade 3.

pronounced for histological grade 3, which also would be expected clinically. For both variables the excess risk seem to be more pronounced at earlier time points.

The cumulative intensity functions for the covariates $p53$-mutations and the mdr-mutations are plotted below. Neither is even remotely significant, $p = 0.18$ and $p = 0.60$, and mdr-mutations was the first covariate that was removed from the model.

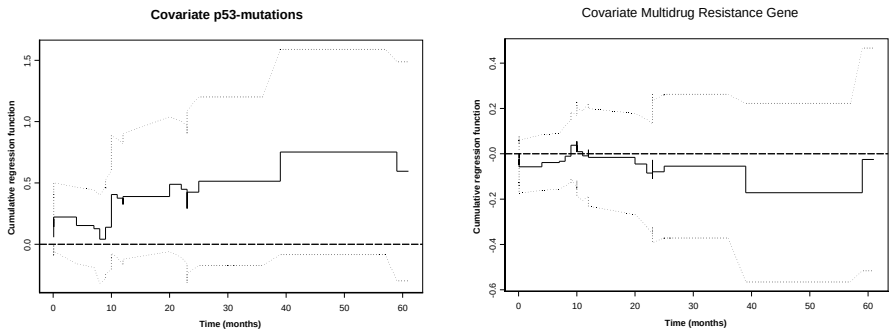

Figure 5.3: Cumulative regression functions for p53-mutations and mutations in Multidrug resistance gene.

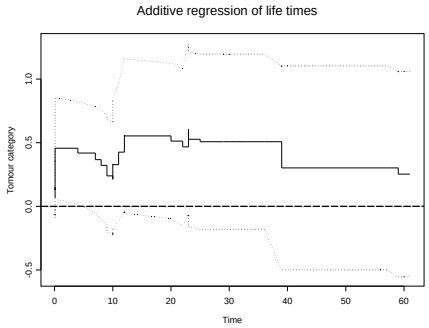

Figure 5.4: Cumulative regression function for tumor stage

The tumor stage covariate clearly showed non-proportional hazard in the Cox proportional hazards model. Tumor stage was therefore included as strata in the model instead of being modelled as a covariate. This non-proportional hazard property is discernable in the cumulative intensity function plot; the excess risk seem to be large at earlier time points but levels

Table 5.1: Value of test statistic TST and p-values for initial fitting of additive model

Covariate	**ER**	**PgR**	**T3/T4**	**Hist2**	**Hist3**	**p53**	**mdrmut**
TST	-1.68	-2.00	1.63	1.59	1.81	1.3	-0.52
P-value	0.09	0.046	0.10	0.11	0.07	0.18	0.60

out at higher time points. This is also consistent with the general clinical perception of $T3$ and $T4$ tumors, where the prognosis is considerably worse for $T4$ tumors. As described under the section "Prognostic factors" $T4$ are tumors of any size, but growing into surrounding tissues, indicating more aggressive biological characteristics. The cumulative regression function indicates an immediate increased risk for relapse for $T4$ patients. Looking at the cumulative regression functions one could conclude that $T4$ patients, after this short term immediate risk increase, does not seem to have substantially higher risk of relapse than $T4$ patients. When fitting a Cox model we discovered this non-proportional hazard behavior by examining residual plots. Residual plots will however not so easily give information on the nature of this non-proportional deviation. When fitting an additive model and plotting the cumulative risk functions, this changing risk profile over time becomes immediately apparent.

Test statistics and p-values for this initial fitting of the additive model is presented in table 5.1.

5.2.1 Power of test statistic TST

Looking at table 5.1 we note that the value of the test statistic TST for histologic grade 3 indicates that the null hypothesis may not be rejected at the 5 percent level, which means that the null hypothesis of $H_j : \alpha_j(t) = 0$ for all t, should be retained. However, looking at figure 5.2 where point wise confidence intervals have been plotted, we can clearly se that the cumulative regression function *is* significantly different from 0 at time points approximately between 12 and 22 months, at least if we do not take multiplicity into account. This would indicate that the null hypothesis is not true. This apparent contradiction is easily comprehended if one keeps the construction of the TST test statistic in mind, and it may be useful to remember that there could be situations where the TST statistic is not so powerful in detecting deviations from the null hypothesis.

5.2.2 Fitting of a Cox model starting from a later time point

To validate the conclusions regarding the nature of the deviation from the non-proportional hazard for the $T4/T3$ covariate presented above, a Cox proportional hazards model was fitted starting at time equal to 4 months after surgery. This means that patients experiencing a relapse ≤ 4 months after surgery were deleted from the analysis. There were 8 patients who relapsed within 4 months of surgery and 6 of these patients were Tumor stage 4 patients. Performing a Cox analysis with these high risk/rapid relapse patients we would expect the covariate coefficient for tumor stage to be much smaller compared to starting from time 0 and no longer significant.

The fitting of a Cox model starting from this later time point did indeed result in the $T3/T4$ variable losing all significance in the Cox analysis, $p = 0.42$. This supports the finding in the additive model: tumor stage 4 does not significantly increase the risk of relapse after the first critical months.

5.2.3 Revised additive model

In the revised model the following variables are included: progesterone receptor level, histological grade 2 and histological grade 3, tumor stage ($T3/T4$) and $p53$ gene mutations. The estrogen receptor covariate was excluded for the same reasons described before.The cumulative regression function curve for the remaining covariates are displayed in fig 5.5.

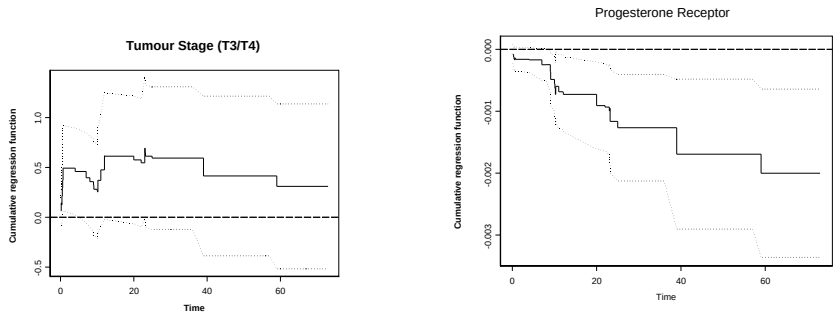

Figure 5.5: Cumulative regression functions for Tumor stage and Progesterone receptor value

The tumor stage variable is still not significant, $p = 0.1605$, and the non-proportional hazard pattern is clearly visible. The cumulative regression function has the same pattern as before.After removal of the estrogen

receptor covariate the cumulative regression function for progesterone now has a higher numerical value and is also clearly much more significant, the 95 percent confidence interval for the cumulative regression function excluding 0 for virtually all time points.

The progesterone receptor level is nicely fitted and the cumulative curve has a relatively constant slope, indicative of proportional hazard throughout the model.

The cumulative excess hazard for indicator variables for histological grade 2 and histological grade 3 are displayed in fig 5.6.

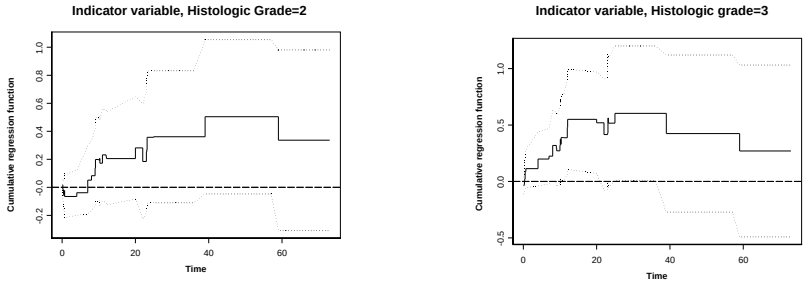

Figure 5.6: Cumulative regression functions for Indicator variables histologic grade 2 and 3

5.3 Model Checking of Additive Model

5.3.1 Martingale residuals

Martingale residuals may be used for model checking both for the Cox model and for the additive model[Aalen 1993]. The martingale residuals may be defined over the time period when estimation is possible, i.e. until the time when the matrix Y(t) loses full rank. That means that all observation beyond this time point are considered to be censored, even if the full life length has been observed at a later time point than this final time point, called R. S_i denotes the observation time and this time is now bounded by R. $N_i(t)$ is a counting process for the i'th individual and $N_i(t)$ is 0 until the event of interest has occurred and 1 afterwards. If the individual is censored $N_i(t)$ will stay at 0.

If we assume that all covariates are time independent (which they are in our example) the cumulative intensity at time (t) for individual (i) will be:

$$\Lambda_i^*(t) = A^*(t \wedge S_i)' Z^i$$

where $t \wedge S_i$ means the minimum of the two times , i.e. $\Lambda_i^*(t)$ is constant after the observation of the individual has been made.

The martingale residual process for the i'th individual is defined as:

$$M_{res,i}(t) = N_i(t) - \Lambda_i^*(t)$$

This means that for each individual the martingale residual process may change value (usually decreasing) up to the point where either an event occurs and 1 is added to the martingale residual or until the individual is censored. After the event has occurred or the individual is censored the martingale residual is constant. The value of the martingale residual process at the final estimation time is called the martingale residual. The advantage of looking at the martingale residual process is that it gives a picture of how accumulated hazard compared to events occurred changes over time. The idea is to compare the martingale residual process for a subgroup within a data set with different covariate values, to see if the model is valid for all subgroups. In our data set there is only one covariate with continuous values; the progesterone receptor value. Fitting the Cox model it was shown that this covariate preferably should be transformed in some way before inclusion in the model. Below is investigated if and how well the untransformed Progesterone covariate fits in the additive model.

Martingale Residual Process

The Martingale residual process was plotted for subgroups of progesterone receptor values in our data set. In our present model all variables except progesterone receptor values are dichotomized and progesterone receptor values is therefore the only variable in this model which may be investigated by plotting the martingale residual process. For a variable which is coded as 1/0 the only possibility for dividing the data in two groups is by value 1/0 of the covariate and these values of the covariate is what the model has been maximized for. Checking the model for a different fit of the two values of these covariates would therefore not be meaningful since we already have decided that this covariate improves the model. Also no transformation of these variables is possible since they already are included in their simplest form.

The data set was divided in three groups depending on their progesterone receptor value, group I: $Pgr < 20$, group II: $20 < Pgr < 100$ and group III: $Pgr > 100$.

Table 5.2: Value of test statistic and p-values for final additive model with dichotomized PgR

Covariate	**intercept**	**P**	**T3/T4**	**Hist2**	**Hist3**	**p53**
TST	4.91	-3.31	1.89	2.07	2.83	1.70
P-value	0	9e-04	0.06	0.04	0.005	0.07

The cut-off points were chosen arbitrarily, but the lowest category was chosen to include fairly low values of progesterone receptor levels, knowing these patients generally have a worse prognosis. Division of PgR values in these ranges should give a good picture of if the model can accommodate a large range of PgR values with sustained fit. If we would see any deviations in the martingale residual process we would expect the group with lowest PgR values to accumulate negative residuals, due to the model not being able to accumulate enough hazard to match the fairly bad prognosis for these patients. Complementary, the residuals for the high PgR value group would be positive, indicating that the model is predicting great protective effects of very high progesterone receptor values.

The martingale residual process for the model and data above is plotted in fig 5.7 below. In this and the following martingale residual plot please note that the process

$$M_{res,i}(t) = \Lambda_i^*(t) - \delta_i(t)$$

has been plotted, which is equivalent to the definition of the process given before, just with opposite signs.

As can be seen we get the kind of deviations in the martingale residual process we could expect from a covariate with the clinically known hazard profile of progesterone receptor value where low values yield an increased hazard but the protective effects do not increase proportionally as the value increases over 100 up to more than 1000. A convenient choice would be to dichotomize the progesterone receptor value as was done in the Cox proportional hazards model. Since we know the optimal cut-off point in the Cox model, $PgR = 85$ we choose the same cut-off here. Dichotomizing the PgR variable and re-run the additive model for the same covariates as in the revised additive model. Refitting of the additive model with a dichotomized PgR covariate yields the test-statistic with respective p-values displayed in table 5.2.

A plot of the martingale residual process for a model with dichotomized progesterone receptor covariate is shown in figure 5.8. Deviations in the

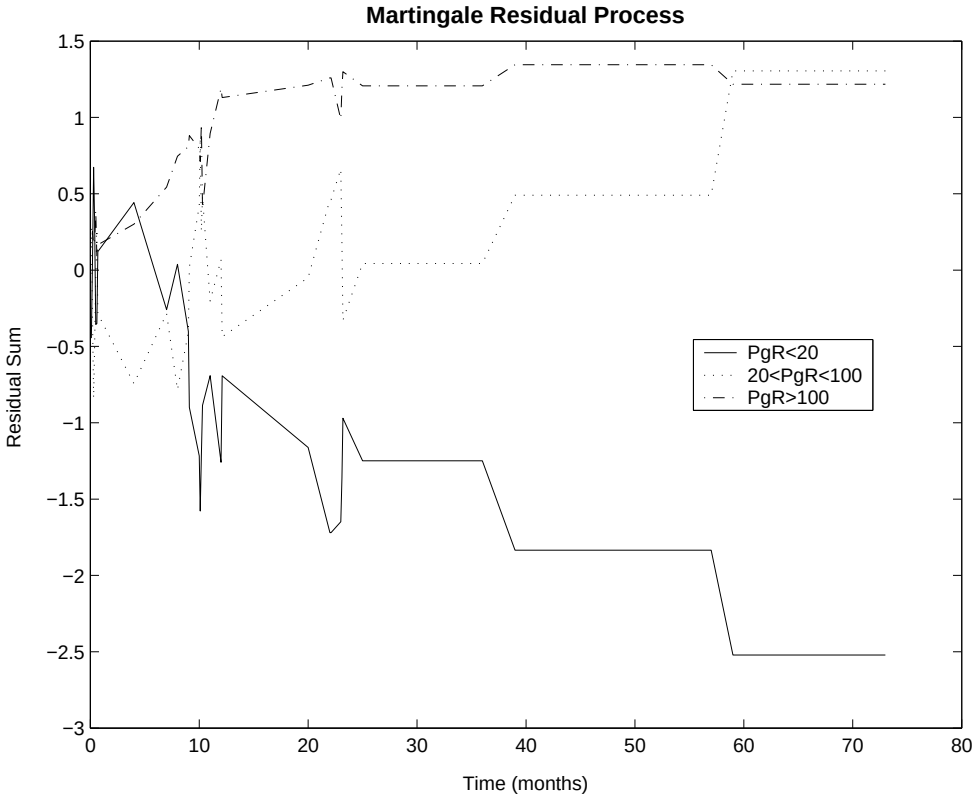

Figure 5.7: Martingale residual process for subgroups of progesterone receptor values

martingale residual process for the dichotomized covariate seem to be less obvious. The receptor value sub-groups have been divided in the same way as before. The maximum deviation is now smaller and the deviations also appear to be more randomly distributed. A model with dichotomized progesterone covariate therefore seems appropriate.

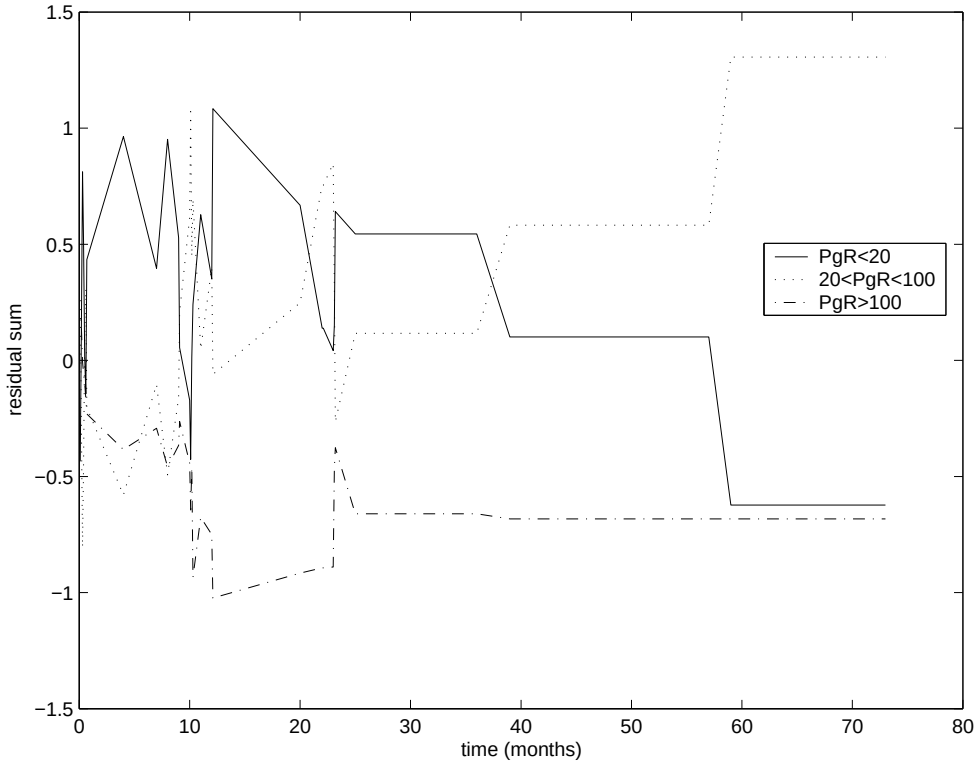

Figure 5.8: Martingale residual process for subgroups of progesterone receptor values, dichotomized covariate

Arja's plot

The concept behind Arja's plot is to plot expected number of failures (abscissa) against actual number of failures (ordinate) in sub-groups with different covariate values. Arjas's plot does not capture the time dependency to the same extent as plotting the martingale residual process. However, to some extent the actual number of failures is a process over time, so the time aspect is indirectly a part also of Arja's plot. The sub-groups chosen were the same sub-groups of progesterone values as in the martingale residual process

plot. These are the sub-groups relevant from a clinical perspective and it also gives an opportunity to compare the model information from these two residual plots. Arjas's plot is not a true residual plot, but deviations from the 45° slope will give essentially the same information. Arja's plot for the three sub-groups of progesterone values is shown in figure 5.9.

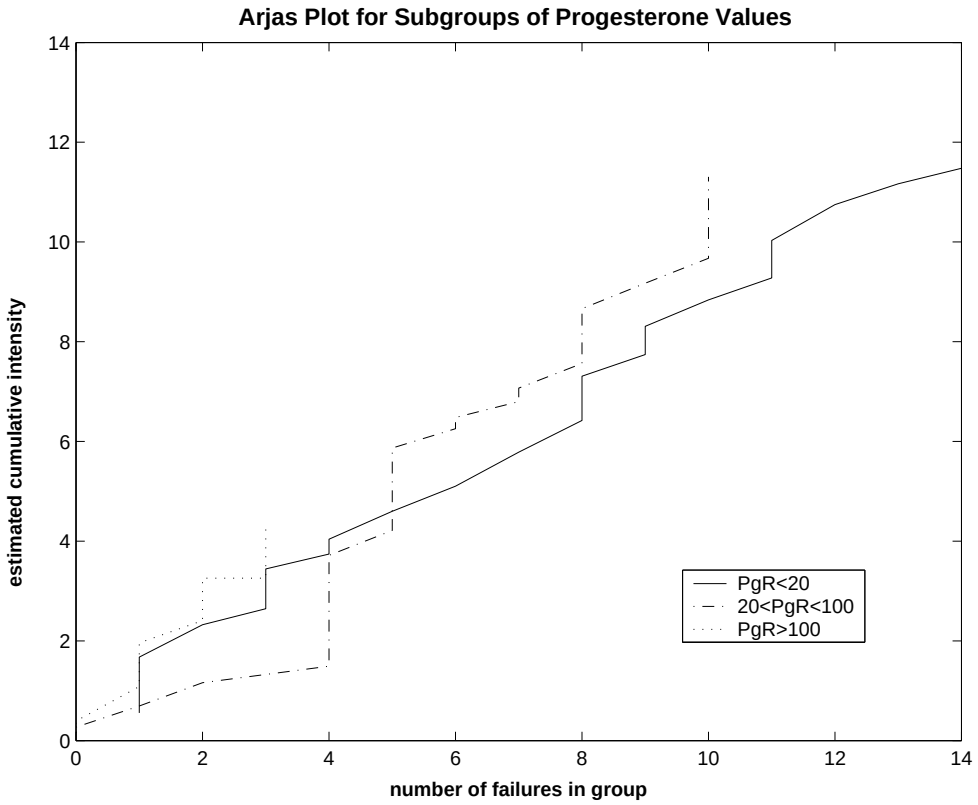

Figure 5.9: Arjas plot for subgroups of progesterone receptor values

Interestingly the deviations from the optimal fit that was clearly visible in the martingale residual plots are not as clearly discernable in Arja's plot for the same sub-groups.

Chapter 6

Concluding remarks

The data set examined in this comparative study is fairly small, 52 patients. The number of initial covariates in the multivariate model is 7. The covariates are correlated to some degree, this is especially true for the progesterone and estrogen receptor levels but also other covariates are correlated. It would therefore not be unexpected if these data would be variable and yield different results depending on model and fitting procedure.

The comparison made here is very informal in nature and information gained from fitting the Cox model has been used in optimizing choices in the additive model.

The Cox model and Aalen's additive model give surprisingly similar results with regard to covariates selected to remain in the model. The actual covariate coefficients estimates are difficult to compare directly since the Cox covariate coefficients act on baseline hazard in a multiplicative way and the additive model gives coefficients for added risks.

The test procedure for the effect of covariates is similar but not equivalent. The null hypothesis for the additive model:

$$H_0 : \alpha_j(t) = 0 \text{ for all t}$$

corresponds to the null hypothesis in the Cox model:

$$H_0 : \beta_j = 0$$

The alternative hypotheses, in favor of which the null hypotheses may be rejected, are however quite dissimilar for the two models. The alternative for the additive model states that $alpha_j(t) \neq 0$for some t, which is a weaker alternative compared to the alternative hypothesis in the Cox model. In the Cox model the alternative is : $H_1 : \beta_j \neq 0$, which is valid for all t. This would imply that to reject the null hypothesis in the Cox model we would

require that the best *overall* estimate of β, which is valid for all time points, is different from 0. To complicate matters however, one must remember that the test statistic in the additive model is designed as a weighted combination of all α_j which means that the null hypothesis may not be so easily rejected even if there are significant deviations from the null hypothesis at a few time points!

No formal comparison of the validity of the tests have been made here, and this is not possible since we do not know the true influence of the covariates. For a formal comparison we would need to use simulated data with known covariate effects. Comparison of likelihoods is not possible since the additive model is not likelihood based. A likelihood based model gives the advantages of likelihood testing, comparison of likelihoods for model decisions and possibly also comparison of nested models. For the additive model more decisions need to be based on the comparison of p-values and residual plots to examine fit.

Both models yielded the same set of covariates, basically by removing the least significant covariates from the models, but the process was of course not unbiased since relevant clinical information was used for selection of covariates and fitting of the Cox model, e.g. transformation of the PgR covariate was applied when fitting the additive model.

Residual plots indicated that neither model could fit the untransformed Progesterone receptor value well. This poor fit of the untransformed PgR value was easily detected for both models.

The tumor stage variable $T3/T4$ displayed clear non-proportional hazard properties and was included as strata in the Cox model. The additive model could easily fit this non-proportional deviation and cumulative regression plots give indications on the nature of this deviation from non-proportional hazard behavior.

A crude way of comparing the models would be comparing p-values for selected covariates. One could argue that size of p-value would be indicative of the power of rejecting the null hypothesis. In table 6.1 above the p-values for the respective models are presented.

At first glance the additive model appear surprisingly powerful since the p-values of the selected covariates are at least as small as for the Cox model, except for $p53$ mutations. Generally a non-parametric model will be less powerful in detecting significant effects. A probable explanation is the different alternative hypotheses, which in this case would indicate that the TST statistic was quite powerful in rejecting the null hypotheses.

Estimation of β and rejecting the null hypothesis for the Cox model immediately gives you a meaningful quantification of the effect, at least relative to individuals having the value 0 of the covariate in question. Estimation

Table 6.1: Comparison of p-value for covariates for Cox model and additive model

Covariate	**p-value Cox**	**p-value Additive**
P	0.039	9e-04
Z1	0.180	0.0384
Z2	0.060	0.0046
$p53mut$	0.03	0.09
T1	Strata	0.0594

of the cumulative regression functions and rejecting the null hypothesis for the additive model doesn't give any immediately meaningful quantification of the effect. If the cumulative regression function can be fitted with a slope, this slope can form a quantitative measure of the covariate effect. In other cases, like for the T3/T4 variable the (close to) rejection of the null hypothesis simply indicates that the effect of the covariates is not 0 at all time points, without giving any quantitative information.

An overall conclusion is that two models give different pieces of information and should not be viewed as alternatives to each other, but as complementary methods that may be used together to give a fuller and more comprehensive understanding of data.

Bibliography

[Klein & Moeschberger 1997] Klein John P., Moeschberger Melvin L.: *Survival Analysis Techniques for Censored and Truncated Data*, Springer-Verlag New York 1997

[Aalen 1989] Aalen Odd O.: *A Linear Regression Model for the Analysis of Life Times*, Statistics in Medicine, 1989 vol 8, 907-925

[Aalen 1993] Aalen Odd O.: *Further Results on the Non-Parametric Linear Regression Model in Survival Analysis*, Statistics in Medicine, 1993 vol 12 (1569-1588)

[Blåboka 2003] Norsk Brystcancer Gruppe: *Blåboka, Brystkreft, Diagnostikk og behandling*, 2003 vol 7 (In Norwegian)

[Venables & Ripley 1999] Venables W.N., Ripley B.D.: *Modern Applied Statistics with S-PLUS*, Springer Verlag 1999 Third edition

[Cox 1972] Cox D.R.: *Regression Models and Life Tables (with discussion)*, B34 Journal of the Royal Statistical Society 1972 (187-220)