Validation and calibration of the ERASL-pre and post risk scores

Results of a Dutch and Japanese investigation

Berend Beumer (381166bb)

July 7, 2019

Abstract

In most medical centers the use of duration models is now part of everyday operation. Often only the relative risk information from a duration models is used to order patients and assign them to distinct risk groups. In contrast, applications from marketing research predominantly use the the richer absolute risk predictions in the form of hazard rates and survival probabilities. Many of these applications can transfer to medicine, provided the predictions are reliability and accurate. Due to the high consequences of medical decisions external validation of proposed models is paramount before implementation. This paper elaborates and applies established external validation methods. Increased attention is given to the lesser known calibration to determine to what extend the absolute risk predictions match with observed data, and calibrate them if need be. The methods are applied on two new validation data sets with the aim to validate the recently published ERASL risk scores. It was found that the model overall orders patients moderate to well for the Okayama data set, and poorly in the Rotterdam data set. Furthermore, was found that the original model systematically over estimates recurrence free survival. This was corrected successfully by embedding the ERASL risk scores in a Weibull calibration model.

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Econometrics and Operational Research [FEB23100-18]

Supervisor: W. Wang

Second assessor: D. Fok

Contents

1 Introduction

How long will it take before the cancer recurs? This question prominently arises for patients just after surgery. Therefore it may come as no surprise that for researchers who investigate treatments and selection criteria, time-to-event data often provide the most insights. Analysis of duration data is however complicated by the way it is collected. Because durations are often measured within a fixed time window, it frequently occurs that at the end of the window the event did not take place. In this case the observation is said to be (right) censored. These censored cases however still provide valuable information, and if not taken into account cause selection bias. To circumvent this issue over the past decades Proportional Hazard (PH) models have been developed, and incorporate the censored observations in their likelyhood function. Once estimated the PH model can be applied in both a relative and absolute manner.

For the aim of simply stratifying patients into distinct risk groups a specification of the linear predictor (LP) suffices, as it contains all relative risk information. The LP, also known as risk score or prognostic index, is the scalar value resulting from the linear combination of explanatory variables and their associated coefficients. The coefficients are most often estimated by maximising the partial likelyhood as first proposed by Cox (1972); Cox (1975). Particularly convenient about this partial likelyhood is that there no need to specify the baseline hazard to obtain the coefficients.

In case the baseline hazard is modeled and estimated an absolute risk description can be obtained in the form of the hazard rate, $h(t)$, or the survival probability, $S(t)$, at a time t. In addition to incorporating the dynamic nature of risk, these also allow for direct comparison between patients, diseases, and treatments.

In the medical literature (surprisingly) little attention is given to evaluating the absolute risk and even less on the validation or practical applications thereof. Interestingly in marketing research this richer expression of the data is more closely investigated and has numerous applications in the field of strategic planning, customer valuation and the timing of promotions (Helsen and Schmittlein, 1993). These decision models could potentially generalize and aid medical decision making as well.

In any application however, the out of sample performance of the PH model is of paramount importance (Steyerberg and Harrell, 2016). Unfortunately their performance is rarely assessed outside their derivation cohorts. Internal validation by means of cross validation or bootstrap is an important first step, though external validation of the model is key in showing a more general pattern rather than mere local results. As described by Royston and Altman (2013) and Rahman et al. (2017) the validation process should ideally be performed on numerous independent data sets and assess: if the model is correctly specified, to what extend the LP can correctly order cases (discrimination), and if the predicted survival probabilities from the survival function match with the observed data (calibration).

According to Moons et al. (2009) if despite correct model specification systematic over- or under- prediction is found, calibrating the model using the new data set should be investigated first before one resorts to re-estimating the model in its entirety. In contrast to the latter, calibrated models incorporates information from both the derivation and validation data sets, hereby improving the stability and generalisability of its predictions. Earlier van Houwelingen (2000) investigated this very matter and used a transformation of time to represent PH model in a Weibull format. He demonstrated that this Weibull model can be used to assess the calibration of the model without the need of risk groups, and provides a natural way to calibrate predictions. Unfortunately did van Houwelingen (2000) not describe the intuition behind the transformation and link the between the Weibull and PH model in much detail. Furthermore is, to the best of my knowledge, the method never applied after its publication in 2000.

Viewing the above it is clear that the application of the survival probabilities is relatively unexploited in medicine. Though only conditional on that the predictions are reliable and unbiased, they can tap into their great potential. Therefore the central aim of this paper is to elucidate the external validation process with increased focus upon the accuracy and calibration of survival probabilities. To this end I will discuss and apply the techniques described by Royston and Altman (2013) and van Houwelingen (2000) to two recently published PH models by Chan et al. (2018). The validation will be performed on two new data sets from The Netherlands and Japan and will hereby also add to the empirical literature.

The remainder of this paper will be organized as follows: in section 2, I discuss the medical context of the Early Recurrence After Surgery for Liver tumour (ERASL) score, and the empirical questions that remain. Section 3 contains a description of the data sets that are used. Next in section 4, I will detail the methods applied to validate and calibrate the model. Whereafter in section 5 and 6 the results and discussion are presented.

2 Medical Context

Liver cancer is a deadly disease affecting millions of people world wide (Ferlay et al., 2015). For patients with sufficient liver reserve a resection of the tumor is indicated to improve survival (Vogel et al., 2018). Although the aim of surgery is curation, in 30-50% of patients the cancer recurs within the first 2-years (Poon et al., 2002; Lise et al., 1998).

Recurrence can originate from incomplete resection, intrahepatic metastasis or multicentric occurence. For each, different risk factors and presentation have been described. The most notable difference is the time until recurrence. Recurrence due to intrahepatic metastasis usually presents within the first two years after surgery, and is often accompanied with diffuse morphology and vascular invasion (Portolani et al., 2006). Multi-centric occurrence is often found more than two years after surgery and usually presents with a few well defined lesions in the cirrhotic remnant liver.

Especially early recurrences form a major challenge. The survival in this group is substantially lower, and the gain from the performed surgery is less clear. Personalized risk prediction of recurrence can aid patients and doctors to decide on whether to perform surgery, the use of adjuvtant chemotherapy, and the intensity of the follow-up.

Both the preoperative (ERASL-pre) and postoperative (ERASL-post) risk scores are developed to predict early recurrence of Hepato Cellular Carcinoma (HCC). Chan et al. (2018) assessed the discriminatory power and calibration of the models in four external validation cohorts from Japan, US, China and Italy. Although several external cohorts are used for validation, the findings are not yet replicated by an independent research group. Also was the calibration only visually assessed and relied heavily on categorization in to risk groups. It was found that in all validation cohorts the ERASL-pre model over estimates recurrence free survival (RFS) for the low and intermediate risk groups. Researchers have not treated this lack of calibration, potential extension or re-calibration of the model in much detail. Furthermore, is the model derived in a hepatitis B prevalent region and uncertainty still exists how well the model generalizes to other areas and will also be investigated here.

3 Data

The data is obtained from the Erasmus Medical Center based in Rotterdam, the Netherlands, and from the Okayama University Hospital in Japan. The data sets contain clinical parameters from patients with HCC who received resection with curative intent.

The Rotterdam Cohort is collected from 2000 until 2017, containing data from 315 surgeries in 308 patients. The recurrence and survival status of patients was last updated in May-2019. In total 175 patients experienced recurrence, of which 126 were found in the first two years after surgery. The maximum follow-up length was 7 years and 5 months. The Japanese cohort is collected between 2007 and 2016, and contains 331 surgeries from 303 patients. Survival parameters where last updated February-2016. In total 161 patients experienced recurrence, of which 121 recurred within the first two years after surgery. The maximum follow-up length was 4 years and 7 months.

In both centers criteria for resection follow the the Barcelona Staging System (Bruix et al., 2001). All patients were discussed at multidisciplinary meeting and had a sufficient performance status, adequate size and function of the remnant liver, and an excision of the tumor with clear margins was deemed possible. After surgery follow-up, including CT and laboratory assessment, was generally performed at 3, 6 and 12 months after discharge and yearly there after for 5 years.

Definitions of the dependent variable and explanatory variables are inline with those described by Chan et al. (2018). The dependent variable, recurrence free survival (RFS), is defined as the time between date of surgery and date of recurrence. For patients who did not experience recurrence or who were lost in

follow up were censored at the date of last radiological examination.

Explanatory variables used in the ERASL scores are: Gender, Albumin (g/l), Bilirubin (μmol/l), serum AFP (μg/l), diameter of largest tumor (cm), and the number of tumors. Microvascular invasion is the only variable added in the postoperative risk score, and is defined as tumor invasion of small vessels only identified upon histological examination. Patients with missing data will be excluded from analysis.

Further is the study exempted by the Medical Research Ethics Committee (MREC). The committee concluded that this study is not subject to the Medical Research Involving Human Subjects Act (WMO) and complies to the declaration of Helsinki (Association et al., 2001).

4 Methodology

This research will focus upon the validation of the ERASL scores. The validation process consists of three stages in which the misspecification, discrimination and calibration will be assessed using the methods and performance measures discussed by Royston and Altman (2013); Rahman et al. (2017); van Houwelingen (2000). First a brief discussion regarding duration models is given, and notation is introduced. Subsequently the misspecification and discrimination statistics used are discussed, whereafter extension of the analysis by means of calibration is explained in more detail.

4.1 Duration model

For estimation a sample is drawn with individuals denoted as $i \in \{1, ..., N\}$. Furthermore we define a random variable $T \in [0, \infty)$ for the duration of a spell, with t the realization thereof. T follows a probability density function $f(t)$ with corresponding cumulative density function $F(t) = P[T \leq t] = \int_0^t f_i(u)du$. The survival function $S(t)$ is defined as $p[T > t]$ and thus holds $S(t) = 1 - F(t)$.

Besides the density functions and survival function the hazard function $\lambda(t)$ is used. The hazard function describes the rate at which spells end at a given time t. More formally $\lambda(t) = \lim_{h \to 0} \frac{P[t \leq T < t+h | T \geq t]}{h}$, or as used more often $\lambda(t) = \frac{f(t)}{S(t)}$. Furthermore is the cumulative hazard function defined as the hazard function integrated from 0 to t, $\Lambda(t) = \int_0^t \lambda(u)du$.

Below is shown that the following general relationships hold. Equations 1 and 2 form intermediate results and form the support of equations 3 and 4. These last two are regularly used to convert between survival and cumulative hazard function.

$$
\begin{aligned}
S(t) &= 1 - F(t) \\
\frac{d}{dt}S(t) &= \frac{d}{dt}(1 - F(t)) \\
\frac{d}{dt}S(t) &= f(t) \\
f(t) &= -\frac{d}{dt}S(t) \quad \square \quad (1)
\end{aligned}
$$

$$
\begin{aligned}
\lambda(t) &= \frac{f(t)}{s(t)} \\
&= -\frac{1}{s(t)} \cdot -f(t) \\
&= -\frac{1}{s(t)} \cdot -(-\frac{d}{dt}S(t)) \\
&= -\frac{d[log(s(t)]}{dt} \quad \square \quad (2)
\end{aligned}
$$

$$
\begin{aligned}
\lambda(t) &= -\frac{d[log(s(t)]}{dt} \\
\int_0^t \lambda(u)du &= \int_0^t -\frac{d}{du}[log(s(u)]du \\
\Lambda(t) &= -log(S(t)) \quad (3) \\
S(t) &= exp(-\Lambda(t)) \quad \square \quad (4)
\end{aligned}
$$

4.1.1 Proportional hazards model

To make the hazard function conditional upon patient characteristics the baseline hazard function λ_0(t), dependent only upon time t, is scaled by $exp(x_i'\beta)$ hereby increasing or decreasing the hazard at all durations. The function $exp(x_i'\beta)$ maps explanatory variables $x_i = (x_{1i}, ..., x_{pi})$ with corresponding coefficients $\beta = (\beta_1 ..., \beta_p)$ onto the domain [0,1]. This leads to the following expression of the conditional hazard function.

$$\lambda(t|x_i) = \lambda_0(t) \exp(x_i'\beta) \tag{5}$$

The β coefficients in equation 5 are often estimated by means of a Cox Proportional Hazards (CPH) model. In this research the CPH model is used in the assessment of the calibration slope, the offset regression and in the forward selection procedure used to examine model adjustments. A brief overview of the CPH model and its estimation is given in appendix section 7.3. The CHP model is also used by Chan et al. (2018) to establish the weights of the ERASL scores. The Linear Predictor (LP) is the scalar value resulting from the linear combination of explanatory variables and their associated weights. The specification of the LP's for the ERASL risk scores are added in Appendix section 7.2.

4.2 Misspecification

As an overall test to assess if the relative risks are correctly specified the *calibration slope* will be computed. The measure is calculated by performing a CPH regression with the LP as the only explanatory variable as shown in equation 6. The estimated coefficient is tested with $H_0 : \alpha_1 = 1$. An α_1 sufficiently close to 1 provides the first evidence that the model is correctly specified (van Houwelingen, 2000).

$$\lambda(t|x_i) = \lambda_0(t) \exp(\alpha_1 * LP) \tag{6}$$

To further investigate to what extend coefficients would differ if they are re-estimated in the validation cohort a CPH model as shown in equation 7 is estimated. Here the LP is used as an offset with its coefficient constrained to 1. The β^* coefficients represent the differences between the derivation and validation cohort. A Likelihood Ratio (LR) test will be used to asses if β^* is jointly significantly different from the null vector.

$$\lambda(t|x_i) = \lambda_0(t) \exp(x_i'\beta^* + 1 * LP) \tag{7}$$

4.3 Discrimination

Discrimination measures assess the ordering of risk scores. A model discriminates well if patients who experience recurrence earlier are assigned higher risk scores compared to those experiencing recurrence later. Discrimination is also known as separation, since the survival curves of models that discriminate well between risk groups lie farther apart. Numerous metrics have been proposed, though most are created around the concordance statistic $C = P[LP_i > LP_j | t_i < t_j]$. C is the probability that individual i with event time before individual j was assigned a higher risk score. This research will evaluate the same metrics as Chan et al. (2018) to aid the comparison, and are briefly mentioned below.

For the calculation of Harrels C-index, random pairs are drawn. The proportion of concordant pairs over all usable pairs is calculated to estimate C (Harrell et al., 1982). The Gönen and Hellers K-statistic is obtained in a similar manner though is based on the reverse definition of concordance $K = P[t < t_j | LP_i > LP_j]$ (Gönen and Heller, 2005).

The Royston and Sauerbrei's D statistic measures the observed difference in survival times between subjects with high and low predicted risk. For the calculation the LPs are ordered, whereafter the order statistics can be written in terms of Expected Standard Normal Order Statistics (rankits) on which a CPH regression is fitted. After scaling, the estimated coefficient can be interpreted as an estimate for the log hazard ratio comparing two-equaly sized prognostic groups defined by dichotomising the distribution of LP's at the median value (Royston and Sauerbrei, 2004). Often the statistic is displayed in its explained variation R^2_D form. The scaling factors σ^2 and κ for duration models are approximately $\pi^2/6$ and $\sqrt{8/\pi}$ respectively.

$$R^2_D = \frac{D^2/\kappa}{\sigma^2 + D^2/\kappa^2} \tag{8}$$

Finally the hazard ratios between risk groups are calculated by running the following regression. Here the ERASL low riskgroup is taken as the reference category.

$$\lambda(t|x_i) = \lambda_0(t) \exp(\alpha_1 * ERASL\_intermediate + \alpha_2 * ERASL\_high)$$

4.4 Calibration

In contrast to the discrimination measures, the calibration assessment looks at the models ability to accurately predict absolute risk levels. Below in subsection 4.4.1 two visualization methods are discussed that display to what extend predictions match the observed data. A weakness however is that these visualizations heavily rely upon risk groups to calculate the Kaplan-Meier survival estimates. The number of risk groups and thresholds are often arbitrarily chosen. Furthermore, though useful for general conclusions and intuition, the visualizations do not suggest how to adjust the model to achieve accurate survival probabilities. Therefore, in subsection 4.4.2 the analysis is extended by rewriting the model in the Weibull format. Subsequently is shown how the Weibull model can be used to to attain a calibrated probabilities.

4.4.1 Visualization

To visualize the accuracy of the survival estimates Chan et al. (2018) compare the predicted survival function to the observed Kaplan-Meier curves for each risk group. The procedure is outlined in Royston and Altman (2013) and will also be followed here.

By means of interpolation a continuous baseline survival function is constructed form the published baseline survival probabilities given at 1, 3, 6, 12, 18 and 24 months after surgery. Thereafter survival functions are calculated for each individual using $S(t, PI_i) = S_0(t)^{exp(PI_i)}$. These survival functions are then averaged per risk group and plotted together with the risk group Kaplan-Meier survival curve. The visualization allows the researcher to observe deviations overtime time and across risk groups.

In a second visualization the same information is displayed though from a different perspective. Here the predicted survival probabilities at fixed points in time (e.g. 1 and 2 years after surgery) are plotted against the Kaplan-Meier estimates and compared against the 45 degree line.

4.4.2 Weibull calibration model

The aim of the Weibull calibration model is not only to assess performance without the need of risk groups, but also improve the local usability of the risk score. The validation data is first used to assess the appropriateness of the baseline hazard and LP, whereafter the model is tuned rather than refit in its entirety.

The model was first proposed by van Houwelingen (2000). The link between the PH and Weibull model was also previously discussed by Jain and Vilcassim (1991) in their appendix. Here a more detailed version is presented. First is discussed how the the hazard function in equation 5 can be rewritten to attain a linear expression in $x'\beta$. Secondly, it is demonstrated that the linearization coincides with a Weibull model. Finally is shown how the Weibull model can be conveniently used to calibrate duration models.

Linearization

Below the survival function is expressed in terms of its baseline survival function using the equation 4

and 5 and the definition of the cumulative hazard.

$$\begin{aligned} S(t|x_i) &= exp(-\Lambda(t|x_i)) \\ &= exp(-\int_0^t \lambda(u|x_i)du) \\ &= exp(-\int_0^t \lambda_0(u)exp(x_i'\beta)du) \\ &= exp(-\int_0^t \lambda_0(u)du \cdot exp(x_i'\beta)) \\ &= [exp(-\Lambda_0(t))]^{exp(x_i'\beta)} \\ &= S_0(t)^{exp(x_i'\beta)} \end{aligned} \tag{9}$$

Next is shown that by using equation 3 and substitution of equation 9 the cumulative hazard function can be written as follows.

$$\begin{aligned} \Lambda(t|x_i) &= -ln(S(t|x_i)) \\ &= -ln(S_0(t)^{exp(x_i'\beta)}) \\ &= -ln(S_0(t) \cdot exp(x_i'\beta) \end{aligned} \tag{10}$$

Deducing that taking the log on both sides again results in an additive relation of $x'\beta$ on the log of the cumulative hazard.

$$\begin{aligned} ln(\Lambda(t|x_i)) &= ln(-ln(S(t|x_i))) \\ &= ln(-ln(S_0(t))) \cdot exp(x_i'\beta) \\ &= ln(-ln(S_0(t))) + ln(exp(x_i'\beta)) \\ &= ln(-ln(S_0(t))) + x_i'\beta \\ &= ln(\Lambda_0(t)) + x_i'\beta \end{aligned} \tag{11}$$

Expression 11 provides the key to the linearization in $x_i\beta$ resulting in equation 12 and completes the derivation.

$$\begin{aligned} ln(\Lambda_0(t)) + x_i'\beta + \epsilon = 0 \\ ln(\Lambda_0(t)) = -x_i'\beta + \epsilon \end{aligned} \tag{12}$$

Weibull model

The linearization shown in equation 12 can also be viewed as the linear representation of a Weibull model, the link is presented next. A start will be made from the Weibull hazard function and build up to the linearization of the survival function that matches equation 12. The Weibull distribution can be parameterized in many ways, here the parameterisation pressented by Therneau and Lumley (2015); Hunter (2011) is followed.

Let θ_1 be the scale parameter and θ_2 be the shape parameter. Then the baseline hazard function of the Weibull function is known to be $\lambda_0(t) = \theta_1^{-\theta_2}\theta_2 t^{\theta_2 - 1}$. Below first the derivation of the baseline cumulative

hazard is shown.

$$\begin{aligned}
\lambda_0(t) &= \theta_1^{-\theta_2}\theta_2 t^{\theta_2-1} \\
\int_0^t \lambda_0(u)du &= \int_0^t \theta_1^{-\theta_2}\theta_2 u^{\theta_2-1}du \\
\Lambda_0(t) &= \frac{\theta_2}{\theta_1^{\theta_2}} \cdot \int_0^t u^{\theta_2-1}du \\
\Lambda_0(t) &= \frac{\theta_2}{\theta_1^{\theta_2}} \cdot \frac{1}{\theta_2} t^{\theta_2} \\
\Lambda_0(t) &= \frac{t^{\theta_2}}{\theta_1^{\theta_2}} = \left(\frac{t}{\theta_1}\right)^{\theta_2}
\end{aligned} \tag{13}$$

The cumulative hazard is formed by scaling the baseline cumulative hazard, resulting in equation 14.

$$\Lambda(t|x_i) = \left(\frac{t}{\theta_1}\right)^{\theta_2} \cdot exp(x_i'\beta) \tag{14}$$

Equation 14 is subsequently substituted in equation 4 leading to the following expression of the survival function.

$$\begin{aligned}
S(t|x) &= exp(-\Lambda(t|x)) \\
&= exp\left(-\left(\frac{t}{\theta_1}\right)^{\theta_2} \cdot exp(x_i'\beta)\right)
\end{aligned} \tag{15}$$

To also attain an additive and linear expression of $x_i'\beta$ a log transformation is used. Let $Y = log(T)$, and substitute $T = e^Y$ in equation 15 to attain the following expression.

$$\begin{aligned}
S(t|x) &= exp\left(-\left(\frac{t}{\theta_1}\right)^{\theta_2} \cdot exp(x_i'\beta)\right) \\
P(Y \geq y|x) &= exp\left(-\left(\frac{e^y}{\theta_1}\right)^{\theta_2} \cdot exp(x_i'\beta)\right) \\
&= exp(-\theta_1^{-\theta_2} \cdot e^{y\theta_2} \cdot e^{x'\beta}) \\
&= exp(-e^{-ln(\theta_1)\theta_2} \cdot e^{y\theta_2} \cdot e^{x'\beta}) \\
&= exp(-exp(-ln(\theta_1)\theta_2 + y\theta_2 + x'\beta))
\end{aligned} \tag{16}$$

To further show that the linearisation of the argument matches with the linearization obtained in equation 12 a change of variables is performed as proposed by Therneau and Lumley (2015); Hunter (2011). Let the scale parameter be $\sigma = \frac{1}{\theta 2}$, the intercept as $\mu = ln(\theta_1)$, and finally let the coefficient parameter be $\gamma = -\beta\sigma$. Substituting this in the equation 16 results in:

$$\begin{aligned}
P(Y \geq y|x) &= exp(-exp(-\frac{1}{\sigma}\mu + \frac{1}{\sigma}y - \frac{1}{\sigma}\gamma x)) \\
&= exp(-exp(\frac{1}{\sigma}(y - \mu - \gamma x)))
\end{aligned} \tag{17}$$

The argument of equation 17 can be linearized by equating to zero and addition of an error term W. The

equation is reorganized and y is back transformed substituting $y = ln(t)$.

$$\begin{aligned}
\frac{1}{\sigma}(y - \mu - \gamma x) &= W \\
\frac{1}{\sigma}(ln(t) - \mu - \gamma x) &= W \\
\frac{1}{\sigma}ln(t) - \frac{1}{\sigma}\mu - \frac{1}{\sigma}\gamma x) &= W \\
\frac{1}{\sigma}ln(t) &= \frac{1}{\sigma}\mu + \frac{1}{\sigma}\gamma x + W \\
ln(t) &= \mu + \gamma x + \sigma W
\end{aligned} \tag{18}$$

As explained by Rodriguez (2010) it can be shown that if error term W follows a extreme value type 1 distribution, then T will follow a Weibull distribution. Further, observing that the linearization of the proportional hazards model equation 12 complies with equation 18 concluding the proof of the link.

Calibration model

Using this link van Houwelingen (2000) proposed the following calibration model using $T* = \Lambda_0(t)$.

$$ln(T*) = \mu + \gamma(LP) + \sigma W \tag{19}$$

The parameters have the following interpretation. The intercept μ describes the correction for the overall risk level, graphically rotating the survival curves around the fixed point (t = 0; S(t) = 1). Where, σ controls the shape of the baseline. Finally γ parameter moderates the impact of the LP, graphically expanding or contracting the distance between the risk-groups survival curves.

Once the parameters of equation 19 are estimated survival probabilities can be retrieved using equation 17 leading to the following expression.

$$S(t|x)_{cal} = P[T > t|x] = exp(-exp(\frac{1}{\sigma}(ln(-ln(S_0(t))) - \mu - \gamma' x))) \tag{20}$$

The improvement will be assessed using the two previously discussed visualizations. Note that in this manner accurate probabilities are attained by using the new data to only estimate three parameters. Apart form the calibration slope, the misspecification and discrimination measures will not change. Since the monotone transformations do not change the ordering of the risk scores or the allocation of patients to one of the risk groups.

4.4.3 Addition of covariates

If major departures are discovered in the earlier discussed offset regression. It might be useful to refit parts of the model using a forward selection process, starting with an empty CPH model with the coefficient of the LP constrained to one. Then in successive rounds the variable that significantly improves the fit the most is added. After each addition the model is re-estimated the model and the is procedure repeated. Note however, that in this manner also the baseline hazard is re-estimated, and that with each addition the model further departs from the original. This procedure should therefore be considered

model building rather than model calibration. Before being implemented this newly found model should go through another round of validation.

5 Results

In the following section first the baseline characteristics and the distribution of the ERASL scores are reported, whereafter the general properties about the risk groups are displayed. In the next three sections the results of the misspecification, discrimination, and calibration analysis will be discussed.

5.1 Baseline characteristics

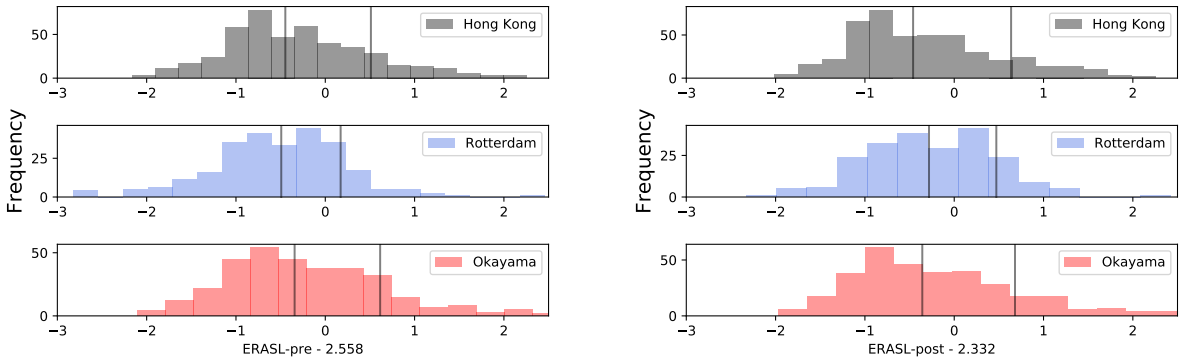
In table 1 the baseline characteristics for the studied cohorts are summarized, information from the Hong Kong derivation cohort is added to aid comparison. Intrestingly a Recurrence Free Survival (RFS) of 21, 31 and 66 months was found for the Rotterdam, Okayama and Hong Kong cohort respectively. Further differences in the proportion of microvascular invasion was found with 58%, 70% for the Rotterdam and Okayama validation cohorts and 27% for Hong Kong derivation data set. Lastly, a notable difference exist with regard to cause of the disease. In the Okayama data set the cause is most often ascribed to Hepatitis C (48%) were in Hong Kong Hepatitis B (84%) is most prominent. In the Rotterdam cohort hepatitis infections occur less often overall with 25% of patients presenting with Hepatitis B and 14% with Hepatitis C.

Table 1: Baseline Characteristics

Variables	Rotterdam	Okayama	Hongkong (derivation)
Patient factors			
n	261	331	451
Male gender, n (%)	181 (69)	269 (81)	387 (86)
Age [years, mean (SD)]	60 (14)	66 (10)	56 (11)
Hepatitis B, n (%)	66 (25)	79 (24)	380 (84)
Hepatitis C, n (%)	37 (14)	159 (48)	18 (4)
Child-Pugh grade, n (%)	n=259	n=296	
A	249 (96)	290 (89)	442 (98)
B	10 (4)	6 (2)	9 (2)
C	0 (0)	0 (0)	0 (0)
ALBI grade, n (%)			
1	210 (80)	215 (65)	329 (73)
2	49 (19)	115 (35)	119 (26)
3	2 (1)	1 (0)	3 (1)
Albumin [g/L, mean (SD)]	42 (5.6)	40 (4.7)	40 (4.4)
Bilirubin [umol/L, median (IQR)]	10 (7, 15)	12 (9, 15)	10 (7, 13)
AFP [ug/L, median (IQR)]	8 (3, 143)	104 (37, 944)	52 (5, 585)
Tumor characteristics			
Tumor size [mm, median (IQR)]	54 (30, 92)	35 (23, 58)	40 (25, 60)
Solitary tumor, n (%)	196 (75)	229 (69)	350 (77)
Microvascular invasion, n (%)	125 (58) n=215	233 (70)	121 (27)
Clinical outcome			
Recurrence within 2 years, n (%)	108 (41)	135 (41)	162 (35.9)
Recurrence-free survival [months (95% CI)]	26,1 (21, 37)	31 (24, 41)	66 (48, 83)

In figure 1 the distributions of the ERASL-pre and ERASL-post scores are displayed. The scores are centered on the published median values of 2.558 and 2.333 for the ERASL-pre and post score respectively. First it can be observed that the median values in the Hong Kong derivation cohort differ from the ones published for the pre and post scores. Further could be noted that for both the pre and post scores the Rotterdam distribution is skewed to the left where the Hong Kong and Okayama cohorts are skewed to the right.

Figure 1: Distribution ERASL LP



Distributions of the ERASL pre and post risk scores in the Hong Kong derivation cohort and the Rotterdam and Okayama validation cohorts. The scores are centered on the median values described in the paper by Chan et al. (2018). In each histogram the left and right black lines represent the 50th and 85th percentile respectively.

5.2 ERASL risk groups

The risk groups for the Rotterdam and Okayama cohort are constructed according to the formulation displayed in appendix 7.2 and follow the paper by Chan et al. (2018). Table 2 and 3 show the risk group sizes, median survival and the relative risk with the low risk group taken as the reference category. It can be observed that for the Rotterdam cohort only 4 (2%) patients are assigned to the high risk group. Furthermore, differences between risk groups in terms of median survival and hazard ratios are overall greater in the Okayama cohort compared to the Rotterdam cohort. Also can be seen that the differences between risk groups increase as information regarding the microvascular invasion is added in the ERASL-post score.

Table 2: ERASL-pre

Cohort	Group	n (%)	Median RFS, months (95%CI)	Hazard Ratio (95%CI)
Rotterdam	Low	202 (77)	27.89 (21.78, 38.60)	1
	Intermediate	55 (21)	17.22 (7.39, 42.80)	1.47 (0.94, 2.30)
	High	4 (2)	4.27 (0.49, not reached)	2.03 (0.50, 8.26)
Okayama	Low	206 (62)	47.08 (36.50, not reached)	1
	Intermediate	96 (29)	14.36 (11.04, 24.00)	2.66 (1.84, 3.85)
	High	28 (8)	4.04 (2.99, 17.70)	5.35 (3.26, 8.77)

Table 3: ERASL-post

Cohort	Group	n (%)	Median RFS, months (95%CI)	Hazard Ratio (95%CI)
Rotterdam	Low	133 (62)	31.77 (24.02, 39.80)	1
	Intermediate	78 (36)	17.08 (10.61, 25.10)	1.89 (1.24, 2.87)
	High	4 (2)	5.65 (0.49, not reached)	6.48 (1.98, 21.19)
Okayama	Low	208 (63)	47.57 (36.99, not reached)	1
	Intermediate	94 (28)	13.93 (10.12, 19.80)	3.01 (2.08, 4.36)
	High	28 (8)	4.04 (3.45, 11.90)	5.85 (3.56, 9.60)

5.3 Misspecification

Calibration slope

Table 4 shows that the calibration slopes or shrinkage factors for the Okayama cohort are closer to one than for the Rotterdam cohort. Furthermore is clear that the factors are larger for the post-operative model. The LR test only indicates that coefficient for the the Rotterdam ERASL-pre score is significantly different from 1.

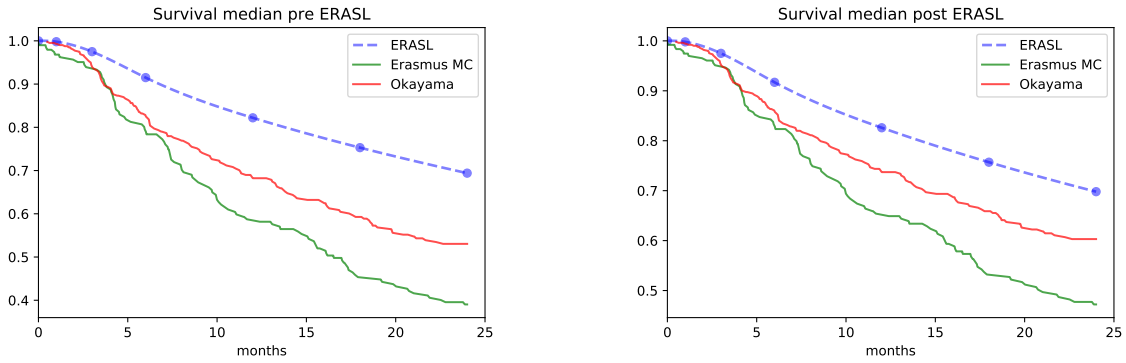
Table 4: Regression on the LP

	ERASL-pre		ERASL-post	
	Rotterdam	Okayama	Rotterdam	Okayama
β (SE)	0.41 (0.15)	0.83 (0.11)	0.82 (0.17)	0.89 (0.10)
p-value	<0.000	0.106	0.286	0.257

Baseline survival

As discussed in the the methodology section the LP in the proportional hazards model scales the baseline hazard. Also the one to one relationship between the the cumulative hazard and the survival function is derived. For ease of interpretation figure 2 shows the baseline survival function. In stead of showing the baseline survival with LP=0, the LP is set to the median ERASL values to match the values published by Chan et al. (2018). As can be observed are the survival functions for both cohorts and risk scores lower than the ones used to obtain the ERASL survival probabilities.

Figure 2: Survival function Median ERASL



Offset regression

The coefficients resulting from the offset regressions represent the difference between the coefficients published by Chan et al. (2018) and coefficients obtained if the model was refit on the validation data. The results are added in tabel 7 and 8 in appendix section 7.4. Except for the ERASL-post risk score in the Rotterdam cohort, all joint tests for coefficients equal to zero were rejected. In the univariate analysis most notable was that less weight was assigned to the gender variable. Furthermore was the difference for the albigrade>1 significantly different from zero in the ERASL-pre offset regression on the Rotterdam cohort.

5.4 Discrimination

In table 5 the discrimination measures are displayed. All discrimination measures are higher in the Japanese cohort, and are higher for the ERASL-post score.

Table 5: Measures of discrimination

Measure of discrimination	ERASL-pre		ERASL-post	
	Rotterdam	Okayama	Rotterdam	Okayama
Harrel C	0.588 (0.032)	0.682 (0.027)	0.650 (0.026)	0.709 (0.022)
Gönen & Heller's K	0.674 (0.007)	0.699 (0.006)	0.677 (0.006)	0.705 (0.006)
Royston-Sauerbrei's Rd^2	0.046 (0.036)	0.239 (0.057)	0.163 (0.059)	0.298 (0.060)

5.5 Calibration

The coefficients estimated for the Weibull model are shown in table 6. Using the earlier discussed change of variables the proportional hazard interpretation can be attained by dividing γ by $-\sigma$. This results in the values 0.40 and 0.81 for the ERASL-pre and post in the Rotterdam cohort and 0.84 and 0.89 for Okayama, and confirm with the estimates attained by regression on the LP. Furthermore the μ estimates are all negative varying between -0.53 and -1.97 and represent the mismatch in the overall risk level.

Furthermore deviations from 1 for the σ parameter are larger for Rotterdam than for Okayama. The estimates for σ are significantly different from one for the of the pre-ERASL scores.

Table 6: Weibull calibration model

	ERASL-pre		ERASL-post	
	Rotterdam	Okayama	Rotterdam	Okayama
μ	-1.97 (0.00)	-0.60 (0.12)	-0.92 (0.06)	-0.53 (0.13)
γ	-0.49 (0.01)	-0.97 (0.00)	-0.93 (0.00)	-1.02 (0.00)
σ	1.23 (0.02)	1.16 (0.05)	1.15 (0.13)	1.14 (0.09)

Figure 3 and 4 show the calibration plots for the Rotterdam and Okayama cohorts. Since a kaplan meier curve can only be estimated for a group, the patient level risk functions attained from the model are averaged per risk group. The smooth solid lines represent the original model, and the dashed curves results after the calibration described above. Calibrated survival probabilities where obtained with the estimates in table 6 and equation 19.

For the Rotterdam cohort the high risk group is not displayed as there were deemed to too few data points to provide additional insight. In both the pre-operative and post-operative setting the origional ERASL models systematically overestimate the RFS for the low and intermediate risk groups, and is seen in both the Rotterdam and Okayama cohorts. Further can also be observed that the lines predicted by the origional ERASL model lie further apart than the kaplan meier curves. The calibrated model follows the kaplan meier curves much closer. Though now under estimation of the calibrated model is visible for the high risk group.

Figure 3: Calibration plot Rotterdam

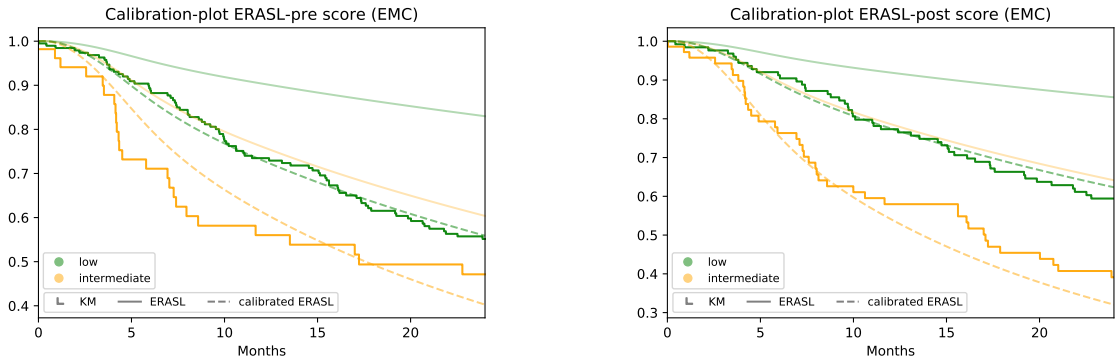
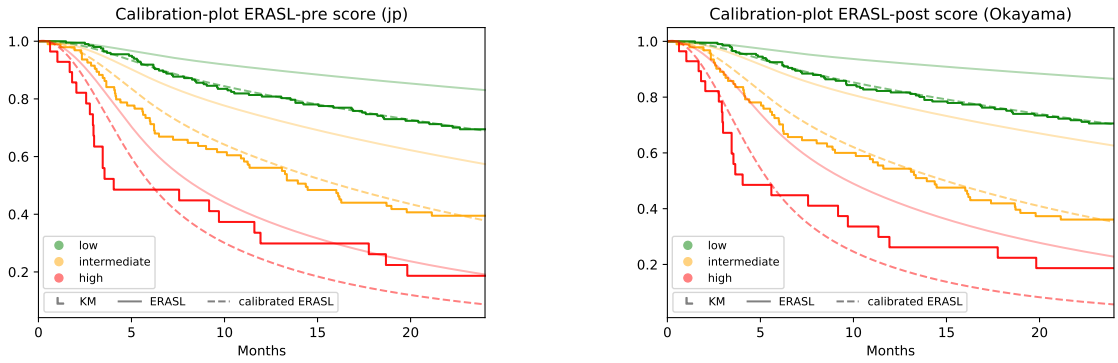


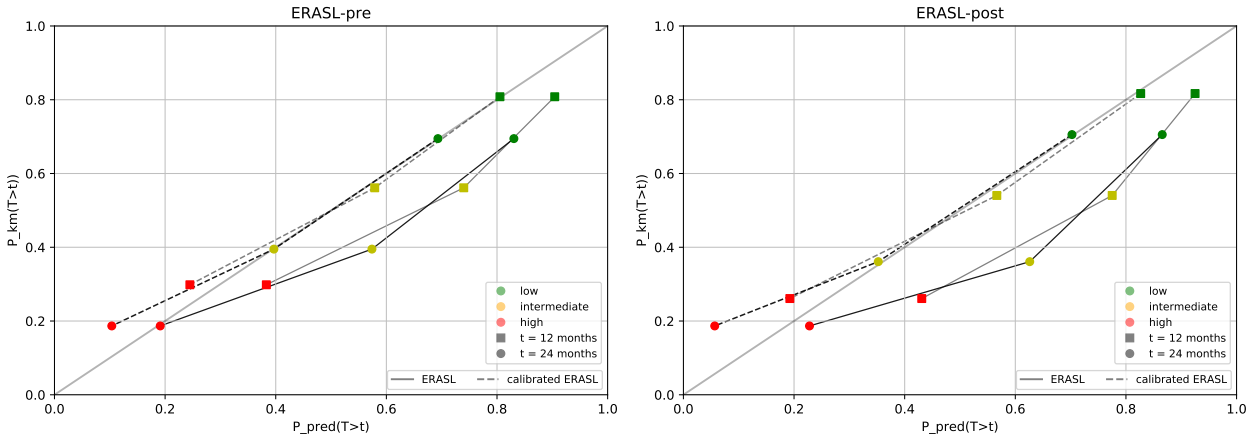
Figure 4: Calibration plot Okayama



An alternative visualization that shows to what extend predictions match the observed data, is displayed in figure 5. The predicted survival probabilities are plotted against the observed Kaplan Meier estimate at 12 and 24 months after surgery. The colors green, yellow, and red represent the risk groups. The original and calibrated models are distinguished by full and dashed lines respectively.

What stands out from figure 5 is that the survival probabilities of the original ERASL model are too optimistic and depart from the ideal 45 degree line. Similar to what is observed in figure 4, the calibrated version slightly underestimates the survival probabilities of the high risk group.

Figure 5: Calibration plot Okayama



After the calibration the *calibration slopes* for the calibrated pre and post models are 0.97 (0.35), 0.98 (0.20), and 0.95 (0.12), 0.96 (0.20) for the Rotterdam and Okayama cohort respectively. Also the mismatch in baseline hazard is resolved for the pre-operative model, and strongly reduced for the post-operative model (not shown).

5.6 Variable addition or partial refit

To investigate the impact of the Hepatitis B and C infections. These are added one by one to the model with the LP constraint to 1. The coefficients for these variables did not achieve significance in both the

pre- and post operative models for both cohorts.

Modification of coefficients from variables already included in the risk score was also investigated. Starting with the LP of the original models constraint to one, variables were added in a stepwise forward selection manner. For the Rotterdam cohort in the pre-operative model, non of the variables achieved significance. In the post operative setting only micro-vascular invasion was significantly different from zero with coefficient 0.638 (p=0.03). For the Japanese cohort the only variable achieving significance in the pre- and post operative models was gender with coefficients -0.752 (p=0.0015) and -0.644 (p= 0.0052).

6 Discussion

As seen in table 1 the overall patient populations appear similar. Yet a remarkable difference in the median RFS of almost three years is seen between the derivation and our validation cohorts. This difference is also observed in all other validation cohorts published by Chan et al. (2018), and raises questions about the patient selection in the Hong Kong derivation cohort. The authors have failed to mention this result or investigated its origin, though the impact on the predicted survival probabilities might be profound. Even though the survival data is censored at 24 months, it is likely that the excellent long term survival translates to the baseline survival function. Evidence of the mismatch between baseline survival functions for our cohorts is shown in figure 2. As the baseline survival function is a key part in forming the predictions it is likely to affect the accuracy of the risk score.

Another point of interest is that the published 50th and 85th quantiles on which the risk score thresholds are based, do not match the quantiles of the derivation cohort as can be seen in figure 1. This causes the proportion of cases in the intermediate and high risk group to be smaller than the intended 35% and 15%, as shown in in tables 2 and 3. This result can also be observed in the other validation cohorts published by Chan et al. (2018). Although the categorization does not affect the relative risks or the individual survival estimates, summary statistics describing the high risk group are less stable and warrant a different interpretation.

Further, regarding the distribution of the ERASL risk scores can be stated that, overall the spread and center of mass in the validation cohorts are similar to the derivation cohort. However, as can be seen in figure 1 are the ERASL scores in the Rotterdam cohort skewed towards the left, potentially reflecting the conservative view towards treatment. This causes there to be a relative absence of high risk patients in comparison to the Okayama and Hong Kong datasets.

Arguably the most important metrics of a prediction model are the discriminatory measures. These measures reported in table 5 can best be viewed in relation with those published by Chan et al. (2018). In the Rotterdam cohort the model performs similar to the Italian validation cohort and substantially

lower than the Hong Kong derivation cohort. In these European cohorts the discriminatory ability of the risk score can be seen as moderate to low. On the contrary in the Okayama cohort the models almost achieve the same level as attained in the derivation data set, and thus hint at a missing variable to describe the difference between the two regions.

A potential candidate risk factor that might explain this difference is the presence of Hepatitis B or C. As seen in table 1 strong differences between cohorts exist with respect to the fraction of patients presenting with Hepatitis B or C. In the stepwise forward selection procedure both variables where never found to be significant. It appears that although HBV and HCV are important factors for diagnosis and treatment, they do not accurately reflect the severity of the disease after having accounted for the other ERASL score variables.

Apart from the models ability to assign the right risk group to each patient, accurate survival probabilities are paramount. Figures 3 and 4 show that the original models are poorly calibrated for the low and intermediate risk groups. The high risk group appears to fit better, though should be adopted with caution as the number of cases supporting the Kaplan Meier curve are minimal. Further can be seen that the original ERSAL models exaggerate the difference in survival between risk groups, as they lie too far apart. Evidence of this can also be seen in table 4 as all slopes are below one.

Turning now to the Weibull calibration model, where for each cohort and model just three parameters are estimated. On the level of averaged predictions, the lack of calibration seems to be severely reduced or even eliminated. It should however be noted that moderate lack of fit is still observed in the Rotterdam intermediate risk group for both pre and post-operative models. Furthermore are the estimates in the high risk group in japan is sill unstable due to the small sample sizes in this risk group.

To explore if the calibration can further be improved or to what extend the misspecification impacts the model, the re-estimation of already incorporated variables is inspected. In both the offset regression as the forward selection procedure the coefficient for the variable *gender* was negative and strongly significant for the Okayama cohort.

For the stepwise selection in the Rotterdam cohort only micro-vascular invasion was found to be significant. Looking at the baseline characteristics in table 1 the proportion of patients with micro-vascular invasion also differs widely. A potential explanation for this might be that, since the presence of micro-vascular invasion is done by subjective assessment of the pathologist, slight differences in definition cause the high variety in observed proportions and estimated coefficients.

After careful consideration, the LP was not redefined by re-estimation the coefficients for gender or microvascular invasion. A general pattern was lacking, and in this application a slight bias is preferential over an unreliable estimates.

Further improvements can certainly be made to expand on the above, and address the following limitations. My analysis was performed on validation cohorts with limited sample sizes. Especially conclusions for the high risk group, clinically most relevant, might be unstable. Repeating the analysis on a larger cohort is advised. Also should more extensively be looked at model expansion as the difference in discriminatory power between cohorts remain unexplained.

Furthermore, in this research the adequacy of the non-parametric baseline hazard was not established. Potentially parametric baseline hazard functions improve the efficiency of the model. Future research could use the flexible Box-Cox formulation of the baseline hazard used by Jain and Vilcassim (1991). The formulation encapsulates, among others, the commonly used Exponential, Weibull, and Gompertz distributions. Adequacy of these could be tested using the Wald test on appropriate parameter restrictions. To aid future research, appendix section 8 is added to provide a sketch how the Box-Cox proportional hazard model can be estimated using the Newton Ralphson method. In addition to alternate specifications of the baseline hazard, stratification thereof or the addition of time dependent variables could be investigated. These modeling techniques are all absent in the ERASL risk scores and might prove fertile ground to improve the model. This however should be considered model building rather than validation, and lied outside of the scope of this research.

Lastly future research should focus upon the implementation of the model into clinical decision making, as abstract survival probabilities might prove hard for patients and doctors to intuitively incorporate in their decisions. Helsen and Schmittlein (1993) outline numerous applications how the hazard function can aid decision making. An overview of duration model specifications used to this end can can be found in the paper of Seetharaman and Chintagunta (2003). One example of a marketing application is in the determination of customer life time value. Accurate forecasts allows a firm to know if, what type of marketing intervention and how much to invest in customer acquisition or retention. Parallels with with medicine can easily be drawn with quality of life as currency equivalent. For a detailed discussion regarding customer life time value the reader is referred to Gupta et al. (2006); Rosset et al. (2002). Another interesting marketing application is found in modeling the purchase timing to help decide on the timing of direct marketing, sales calls. Also here parallels could be drawn for instance in determining when to best plan the follow-up visit to perform the ct-scan or test the lab values.

In summary, this research aimed to validate two duration models for early recurrence after surgical resection of hepatocellular carcinoma. Concluding that the models discriminatory power varies between the Dutch Rotterdam and Japanese Okayama cohort. With a relatively low level of discrimination in the first and moderate to high in the latter. Furthermore, concluding that the original model systematically over estimated survival probabilities, and that with the Weibull model only three parameters need to be estimated calibrate these.

7 Appendix A

7.1 Code

All used code is made available in the following public Dropbox folder.

https://www.dropbox.com/sh/51glt0iaqimabsn/AAC7LdyDoeiQRjh_CNDeNTXVa?dl=0

7.2 Specification ERASL scores

$$\begin{aligned}\text{ERASL-pre score} = \quad & 0.818 * \text{Gender (0: Female, 1: Male)} \\ & + 0.447 * \text{ALBI grade (0: Grade 1; 1: Grade 2 or 3)} \\ & + 0.100 * ln(\text{Serum AFP in lg/L}) \\ & + 0.580 * ln(\text{Tumour size in cm}) \\ & + 0.492 * \text{Tumour number (0: Single; 1: Two or three; 2: Four or more)}\end{aligned}$$

Risk groups were assigned based on the following cut-offs: $\leq$2.558 (low), >2.558 to $\leq$3.521 (intermediate), and >3.521 (high).

$$\begin{aligned}\text{ERASL-post score} = \quad & 0.677 * \text{Gender (0: Female, 1: Male)} \\ & + 0.458 * \text{ALBI grade (0: Grade 1; 1: Grade 2 or 3)} \\ & + 0.082 * ln(\text{Serum AFP in lg/L}) \\ & + 0.451 * ln(\text{Tumour size in cm}) \\ & + 0.379 * \text{Tumour number (0: Single; 1: Two or three; 2: Four or more)} \\ & + 0.661 * \text{Microvascular invasion (0: no, 1: yes)}\end{aligned}$$

Risk groups were assigned based on the following cut-offs: $\leq$2.332 (low), >2.332 to $\leq$3.445 (intermediate), and >3.445 (high).

ALBI grade is calculated as described below and follows the discussion of Johnson et al. (2015); Chan et al. (2018)

$$ALBIscore = -0.085 * (\text{albumin } g/l) + 0.66 * log_{10}(\text{bilirubin } \mu mol/l)$$

$$ALBIgrade = \begin{cases} 1 & \text{if } ALBIscore \leq -2.60 \\ 2 & \text{if } ALBIscore > -2.60, \leq -1.39 \\ 3 & \text{if } ALBIscore > -1.39 \end{cases}$$

7.3 Cox Proportional Hazard Model

The CPH model is currently the most widely used duration model due to its ability to estimate relative risks without the need to specify the baseline hazard. The following subsection will shortly derive the interpretation and estimation of the coefficients from the Cox Proportional Hazards (CPH) model. Further more is shown how the non-parametric survival function can be obtained. Notation is followed as explained by (Cameron and Trivedi, 2005).

Interpretation

As seen earlier in equation 5 the the conditional hazard function can be written as $\lambda(t|x_i) = \lambda_0(t)\exp(x_i'\beta)$. By taking the logarithm and differentiating with regard to explanatory variable x_{1i}, it can be observed that the β_1 can be interpreted as the proportional effect on the log hazard rate. $\frac{\partial ln(\lambda(t|x_i))}{\partial x_{1i}} = \beta_{1i}$

Estimation

The main reason for the wide spread use of the CPH model is that the coefficients can be estimated without specification of the baseline hazard. To achieve this place the event times of sample ascending order ($t_1 < t_2 < ... < t_j < ... < t_k$), and define $R(t_j)$ as the set of cases with an event time greater or equal than event time t_j with $j \in 1, ..., N$. Furthermore, define set $D(t_j)$ to contain all cases with an event time equal to t_j and finally d_j as the number of cases with event time equal to t_j.

$R(t_j) = \{l : t_l \geq t_j\}$

$D(t_j) = \{l : t_l = t_j\}$

$d(t_j) = \sum_l 1(t_l = t_j)$

To perform Maximum Likelihood estimation an expression for the probability function is needed and is presented below. After the hazard function is expanded the baseline hazard can be canceled out.

$$
\begin{aligned}
P[T_j = t_j | R(t_j)] &= \frac{P[T_j = t_j | T_j \geq t_j]}{\sum_{l \in R(t_j)} P[T_l = t_j | T_l \geq t_j]} \\
&= \frac{\lambda_j(t_j|x_j, \beta)}{\sum_{l \in R(t_j)} \lambda_l(t_j|x_l, \beta)} \\
&= \frac{\lambda_0(t_j)\phi(x_j, \beta)}{\sum_{l \in R(t_j)} \lambda_0(t_j)\phi(x_l, \beta))} \\
&= \frac{\cancel{\lambda_0(t_j)}\phi(x_j, \beta)}{\cancel{\lambda_0(t_j)} \sum_{l \in R(t_j)} \phi(x_l, \beta))} \\
&= \frac{\phi(x_j, \beta)}{\sum_{l \in R(t_j)} \phi(x_l, \beta))}
\end{aligned}
$$

The discrete way in which event times are observed cause ties between event times due to grouping. The formal likelyhood contribution of two events j_1 and j_2 is written in equation 21.

$$\frac{\phi(x_j 1, \beta)}{\sum\limits_{l \in R(t_j)} \phi(x_l, \beta))} + \frac{\phi(x_{j2}, \beta)}{\sum\limits_{l \in R(t_j) j_1} \phi(x_l, \beta))} + \frac{\phi(x_{j2}, \beta)}{\sum\limits_{l \in R(t_j)} \phi(x_l, \beta))} + \frac{\phi(x_{j1}, \beta)}{\sum\limits_{l \in R(t_j) j_2} \phi(x_l, \beta))} \tag{21}$$

To speed up the computation often the Breslow & Peto approximation is used to resolve ties.

$$P[T_j = t_j | R(t_j)] \simeq \frac{\prod\limits_{m \in D(t_j)} \phi(x_m, \beta)}{[\sum\limits_{l \in R(t_j)} \phi(x_l, \beta)]^{d_j}} \tag{22}$$

With the Breslow estimator the partial likelyhood function is constructed and used to find the model parameters.

$$L_p(\beta) = \prod_{j=1}^{N} \frac{\prod\limits_{m \in D(t_j)} \phi(x_m, \beta)}{[\sum\limits_{l \in R(t_j)} \phi(x_l, \beta)]^{d_j}} \tag{23}$$

Non-parametric baseline survival function

In addition to the coefficients, also the baseline survival function from the cox proportional hazard model can be obtained. This baseline survival function has a similar form as the Kaplan Meier estimates, with piece wise instantaneous conditional survival probabilities. The following discussion follows Cameron and Trivedi (2005) and was earlier presented by Kalbfleisch and Prentice (2002).

Let α_j be the instantaneous conditional survival probability. Further define $S_0(t_{j+1}) = \prod\limits_{l=1}^{j} \alpha_l = \alpha_j S(t_j)$. An individual with duration t_j can contribute to the likelyhood function in either one of two ways. It can either have the event or can be censored. The contributions for the two cases are separately discussed.

Case1: Event at t_j

$$\begin{aligned}
&P[T = t_{j-1}] - P[T = t_j] \\
&= S(t_j | X\beta) - S(t_{j+1} | X\beta) \\
&= S_0(t_j)^{exp(X\beta)} - S_0(t_{j+1})^{exp(X\beta)} \\
&= (\alpha_j^{-1} - S_0(t_j + 1)^{exp(X\beta)} - S_0(t_{j+1})^{exp(X\beta)} \\
&= (\alpha_j^{-exp(X\beta)} - 1) * S_0(t_{j+1})^{exp(X\beta)}
\end{aligned}$$

Case2: Censored at t_j

$$\begin{aligned}
&P[T > t_j] \\
&= S_0(t_{j+1})^{exp(X\beta)} \\
&= \prod_{l=1}^{j} \alpha_l^{exp(X\beta)}
\end{aligned}$$

Full likelyhood then becomes:

$$L(\alpha, \beta) = \prod_{j=1}^{k} \left[\prod_{l \in D(t_j)} (\alpha_j^{-exp(X_l \beta)} - 1) * \prod_{m \in R(t_j)} \alpha_j^{-exp(X_m \beta)} \right]$$

With log likelyhood:

$$l(\alpha, \beta) = \sum_{j=1}^{k} \left[\sum_{l \in D(t_j)} (ln(\alpha_j^{-exp(X_l \beta)} - 1) + \sum_{m \in R(t_j)} -exp(X_m \beta) * ln(\alpha_j) \right]$$

Using maximisation of the log likelyhood the α_j can be retrieved for every duration j. Then with $S_0(t_{j+1}) = \prod\limits_{l=1}^{j} \alpha_l$ the baseline survival function can be retrieved.

7.4 Results offset regression

Table 7: Offset regression ERASL-pre

	Rotterdam				Okayama			
Variable	Coef	exp(Coef)	se(Coef)	p-value	Coef	exp(Coef)	se(Coef)	p-value
Gender	-0.75	0.473	0.213	<0.001	-0.86	0.423	0.227	<0.001
ALBI grade >1	-0.56	0.570	0.254	0.027	0.12	1.132	0.179	0.488
ln(AFP)	-0.02	0.977	0.029	0.415	-0.03	0.968	0.030	0.280
ln(Tumor size)	-0.25	0.783	0.141	0.082	-0.03	0.973	0.141	0.846
Tumor number	-0.38	0.686	0.197	0.056	0.01	1.014	0.115	0.905
LR test with 5 df		*22.81 (p=4e-04)*				*13.45 (p=0.02)*		

Table 8: Offset regression ERASL-post

	Rotterdam				Okayama			
Variable	Coef	exp(Coef)	se(Coef)	p-value	Coef	exp(Coef)	se(Coef)	p-value
Gender	-0.58	0.557	0.246	0.017	-0.69	0.500	0.228	0.002
ALBI grade >1	-0.52	0.597	0.283	0.069	0.20	1.217	0.181	0.276
ln(AFP)	0.01	1.007	0.032	0.820	-0.03	0.966	0.030	0.254
ln(Tumor size)	-0.12	0.886	0.158	0.446	-0.27	0.764	0.164	0.100
Tumor number	-0.32	0.729	0.218	0.147	0.09	1.097	0.115	0.419
microvascular	0.30	1.355	0.243	0.212	0.31	1.357	0.217	0.158
LR test with 6 df		*11.98 (p=0.06)*				*14.77 (p=0.02)*		

8 Appendix B

Below a detailed sketch is shown of how a proportional hazards model with a flexible Box-Cox baseline hazard function could be estimated using the Newton Ralphson method. First the log likelyhood is constructed, whereafter the gradient and hessian are derived.

Box-Cox formulation — flexible baseline hazard function

$$h_0(t) = \exp\left[\gamma_0 + \sum_{k=1}^{K} \gamma_k \left(\frac{t^{\lambda_k} - 1}{\lambda_k}\right)\right]$$

Restrict to $K=3$, $\lambda_1 = 1$, $\lambda_2 \to 0$, $\lambda_3 = 2$

Then $h_0(t) = \exp(\gamma_0' + \gamma_1 t + \gamma_2 \ln(t) + \gamma_3' t^2)$ met $\gamma_3' = \gamma_3/2$

$h(t_i|x_i) = \exp(\gamma_0' + \gamma_1 t + \gamma_2 \ln(t) + \gamma_3' t^2 + x_{1i}\beta_1 + \cdots + x_{pi}\beta_p)$ $\gamma_0' = \gamma_0 - \gamma_1 - \gamma_3/2$

Likelyhood Proportional hazard model:

$$L_i(\theta|x) = f(t_i|x_i)^{\delta_i} \cdot S(t_i|x_i)^{(1-\delta_i)}$$

$$Ln(L_i(\theta|x)) = \ln\left(f(t_i|x_i)^{\delta_i} S(t_i|x_i)^{(1-\delta_i)}\right)$$

$$= \ln\left(f(t_i|x_i)^{\delta_i}\right) + \ln\left(S(t_i|x_i)^{(1-\delta_i)}\right)$$

$$= \delta_i \underbrace{\ln(f(t_i|x_i))}_{①} + (1-\delta_i)\underbrace{\ln(S(t_i|x_i))}_{②}$$

in general — as seen before

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

$f(t) = h(t)S(t)$

$f(t) = h(t)\exp\left(-\int_0^t h(u)du\right)$

met $S(t) = \exp\left(-\int_0^t h(u)du\right)$

$$= \delta_i\left(\ln(h(t_i)) - \int_0^{t_i} h(u)du\right) + (1-\delta_i)\left(-\int_0^{t_i} h(u)du\right)$$

$$= \delta_i \ln(h(t_i)) - \delta_i \int_0^{t_i} h(u)du - \int_0^{t_i} h(u)du + \delta_i \int_0^{t_i} h(u)du$$

$$= \delta_i \ln(h(t_i)) - \int_0^{t_i} h(u)du$$

$$= \delta_i \ln(h(t_i)) - H(t_i) \quad \text{met } H(t_i) = \int_0^{t_i} h(u)du$$

① $\ln(f(t_i|x_i))$

$= \ln\left(h(t_i)\exp\left(-\int_0^{t_i} h(u)du\right)\right)$

$= \ln(h(t_i)) + \ln\left(\exp\left(-\int_0^{t_i} h(u)du\right)\right)$

$= \ln(h(t_i)) - \int_0^{t_i} h(u)du$ ∎

$$\ell = \sum_{i=1}^{N} \delta_i \ln(h(t_i)) - H(t_i)$$

② $\ln(S(t_i|x_i))$

$= \ln\left(\exp\left(-\int_0^{t_i} h(u)du\right)\right)$

$= -\int_0^{t_i} h(u)du$ ∎

Likelyhood function for Box-Cox formulation

substitute $h(t) = h_0(t)\exp(x_i\beta)$ met $h_0(t) = \exp(\gamma_0' + \gamma_1 t + \gamma_2 \ln(t) + \gamma_3' t^2)$

$$\ell = \sum_{i=1}^{N} \delta_i x_i\beta + \delta \ln(h_0(t_i)) - \exp(x\beta)\boxed{H_0(t)}$$ ③

③ $H_0(t) = \int_0^t h_0(u)du$

$= \int_0^t (\gamma_0' + \gamma_1 u + \gamma_2 \ln(u) + \gamma_3' u^2)du$

$= \gamma_0' \int_0^t du + \gamma_1 \int_0^t u\,du + \gamma_2 \int_0^t \ln(u)du + \gamma_3' \int_0^t u^2 du$

$= \gamma_0' t + \frac{1}{2}\gamma_1 t^2 + \gamma_2(t\ln(t) - t) + \frac{1}{3}\gamma_3' t^3$ ∎

for optimization of the likelyhood via Newton Raphson method holds:

$$\theta_n = \theta_{n-1} - \text{Hessian}(\theta_{n-1})^{-1}\,\text{Gradient}(\theta_{n-1})$$

let $\alpha = [\gamma_0', \gamma_1, \gamma_2, \gamma_3']'$ ← parameters for baseline hazard

$$\text{gradient}(\theta) = \begin{bmatrix} \frac{\partial \ell(\theta)}{\partial \beta} \\ \frac{\partial \ell(\theta)}{\partial \alpha} \end{bmatrix}$$

dim (9×1)

$$\text{hessian}(\theta) = \begin{bmatrix} \frac{\partial^2 \ell(\theta)}{\partial \beta \partial \beta'} & \frac{\partial^2 \ell(\theta)}{\partial \alpha \partial \beta} \\ \left(\frac{\partial^2 \ell(\theta)}{\partial \alpha \partial \beta}\right)' & \frac{\partial^2 \ell(\theta)}{\partial \alpha \partial \alpha'} \end{bmatrix}$$

dim (9×9)

Gradient log likelyhood function Box-cox. formulation

for each $\beta_j \in \beta$, $j = 1, \dots, p$

$$\frac{\partial \ell(\theta)}{\partial \beta} = \frac{\partial}{\partial \beta}\left[\sum_{i=1}^{N} \delta_i X_i'\beta + \delta_i \ln(h_0(t_i)) - \exp(X_i'\beta) H_0(t_i)\right]$$

$$= \sum_{i=1}^{N} \delta_i X_i - \exp(X_i'\beta) H(t_i) X_i' \xrightarrow{\text{IE}} \frac{\partial \ell(\theta)}{\partial \beta_1} = \sum_{i=1}^{N} \delta_i X_{1i} - \exp(X_{1i}\beta_1) H_0(t_i) X_{1i}$$

for $$\frac{\partial \ell(\theta)}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left[\sum_{i=1}^{N} \delta_i X_i'\beta + \delta_i \ln(h_0(t_i)) - \exp(X_i'\beta) H(t_i)\right]$$

$$= \sum_{i=1}^{N} \frac{\delta_i}{h_0(t_i)} \cdot \underbrace{\frac{\partial h_0(t)}{\partial \alpha}}_{1.1} - \exp(X_i'\beta) \underbrace{\frac{\partial H_0(t_i)}{\partial \alpha}}_{1.2}$$

* no general "rule" exists therefore the partial derivatives are calculated for each separately below.

$$\frac{\partial \ell(\theta)}{\partial \alpha} = \left[\underbrace{\frac{\partial \ell(\theta)}{\partial \gamma_0}}_{1}, \underbrace{\frac{\partial \ell(\theta)}{\partial \gamma_1}}_{2}, \underbrace{\frac{\partial \ell(\theta)}{\partial \gamma_2}}_{3}, \underbrace{\frac{\partial \ell(\theta)}{\partial \gamma_3}}_{4}\right]$$

1. $\frac{\partial \ell(\theta)}{\partial \gamma_0} = \sum_{i=1}^{N} \frac{\delta_i}{h_0(t_i)} \cdot 1 - \exp(X_i'\beta) \cdot t$
2. $\frac{\partial \ell(\theta)}{\partial \gamma_1} = \sum_{i=1}^{N} \frac{\delta_i}{h_0(t_i)} \cdot t - \exp(X_i'\beta) \cdot \frac{1}{2} t^2$
3. $\frac{\partial \ell(\theta)}{\partial \gamma_2} = \sum_{i=1}^{N} \frac{\delta_i}{h_0(t_i)} \cdot \ln(t) - \exp(X_i'\beta) \cdot (t\ln(t) - t)$
4. $\frac{\partial \ell(\theta)}{\partial \gamma_3} = \sum_{i=1}^{N} \frac{\delta_i}{h_0(t_i)} \cdot t^2 - \exp(X'\beta) \cdot \frac{1}{3} t^3$

1.1 Partial derivatives $\frac{\partial h_0(t)}{\partial \alpha}$

$$\frac{\partial h_0(t)}{\partial \gamma_0} = \frac{\partial}{\partial \gamma_0}\left[\gamma_0 + \gamma_1 t + \gamma_2 \ln(t) + \gamma_3 t^2\right] = 1$$

$$\frac{\partial h_0(t)}{\partial \gamma_1} = t$$

$$\frac{\partial h_0(t)}{\partial \gamma_2} = \ln(t)$$

$$\frac{\partial h_0(t)}{\partial \gamma_3} = t^2$$

1.2 Partial derivatives $\frac{\partial H_0(t)}{\partial \alpha}$

$$\frac{\partial H_0(t)}{\partial \gamma_0} = \frac{\partial}{\partial \gamma_0}\left[\gamma_0 t + \frac{1}{2}\gamma_1 t^2 + \gamma_2(t\ln(t) - t) + \frac{1}{3}\gamma_3 t^3\right] = t$$

$$\frac{\partial H_0(t)}{\partial \gamma_1} = \frac{1}{2} t^2$$

$$\frac{\partial H_0(t)}{\partial \gamma_2} = (t\ln(t) - t)$$

$$\frac{\partial H_0(t)}{\partial \gamma_3} = \frac{1}{3} t^3$$

In summary is the gradient:

$$\text{gradient}(\theta) = \begin{bmatrix} \sum_{i=1}^{N} \delta_i X_{1i} - \exp(X_{1i}\beta_1) H_0(t_i) X_{1i} \\ \sum_{i=1}^{N} \delta_i X_{2i} - \exp(X_{2i}\beta_2) H_0(t_i) X_{2i} \\ \vdots \\ \sum_{i=1}^{N} \delta_i X_{5i} - \exp(X_{5i}\beta) H_0(t_i) X_{5i} \\ \sum_{i=1}^{N} \frac{\delta_i}{h_0(t_i)} - \exp(X_i'\beta) \cdot t \\ \sum_{i=1}^{N} \frac{\delta_i}{h_0(t_i)} \cdot t - \exp(X_i\beta) \cdot \frac{1}{2} t^2 \\ \sum_{i=1}^{N} \frac{\delta_i}{h_0(t_i)} \cdot \ln(t) - \exp(X_i\beta)(t\ln(t) - t) \\ \sum_{i=1}^{N} \frac{\delta_i}{h_0(t_i)} \cdot t^2 - \exp(X_i'\beta) \frac{1}{3} t^3 \end{bmatrix}$$

$\dim(9 \times 1)$

Hessian log likelyhood function Box-cox formulation

$$\frac{\partial^2 \ell(\theta)}{\partial \beta \partial \beta'} = \frac{\partial}{\partial \beta}\left[\frac{\partial \ell(\theta)}{\partial \beta}\right]$$

$$5\times 5 \quad = \frac{\partial}{\partial \beta}\left[\frac{\partial \ell(\theta)}{\partial \beta_1} \quad \frac{\partial \ell(\theta)}{\partial \beta_2} \quad \cdots \quad \frac{\partial \ell(\theta)}{\partial \beta_5}\right] = \begin{bmatrix} \frac{\partial \ell(\theta)}{\partial \beta_1 \partial \beta_1} & \frac{\partial \ell(\theta)}{\partial \beta_1 \partial \beta_2} & \cdots \\ \frac{\partial \ell(\theta)}{\partial \beta_2 \partial \beta_1} & \frac{\partial \ell(\theta)}{\partial \beta_2 \partial \beta_2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

concluding that:

all OFF diagonal elements are 0

all on diagonal elements follow $\sum_{i=1}^{N} -\exp(X_i\beta)\, H_0(t_i)\, X_i^2$

dd 1 1

$$\frac{\partial}{\partial \beta_1}\left[\frac{\partial \ell(\theta)}{\partial \beta_1}\right] = \frac{\partial}{\partial \beta_1}\left[\sum_{i=1}^{N} \delta_i X_{1i} - \exp(X_{1i}\beta_1)\, H_0(t_i)\, X_{1i}\right]$$

$$= \sum_{i=1}^{N} -\exp(X_{1i}\beta_1)\, H_0(t_i)\, X_{1i}^2$$

$$\frac{\partial}{\partial \beta_1}\left[\frac{\partial \ell(\theta)}{\partial \beta_2}\right] = \frac{\partial}{\partial \beta_1}\left[\sum_{i=1}^{N} \delta_i X_{2i} - \exp(X_{2i}\beta_2)\, H_0(t_i)\, X_{2i}\right]$$

$$= 0$$

$$= \frac{\partial^2 \ell(\theta)}{\partial \alpha \partial \alpha'} = \begin{bmatrix} \frac{\partial^2 \ell(\theta)}{\partial \gamma_0 \partial \gamma_0} & \frac{\partial^2 \ell(\theta)}{\partial \gamma_0 \partial \gamma_1} & \frac{\partial^2 \ell(\theta)}{\partial \gamma_0 \partial \gamma_2} & \frac{\partial^2 \ell(\theta)}{\partial \gamma_0 \partial \gamma_3} \\ \frac{\partial^2 \ell(\theta)}{\partial \gamma_1 \partial \gamma_0} & \frac{\partial^2 \ell(\theta)}{\partial \gamma_1 \partial \gamma_1} & \cdots & \cdots \\ \vdots & & \ddots & \\ \frac{\partial^2 \ell(\theta)}{\partial \gamma_3 \partial \gamma_0} & \cdots & \cdots & \frac{\partial^2 \ell(\theta)}{\partial \gamma_3 \partial \gamma_3} \end{bmatrix}$$

4×4

dd 1 2

$$\frac{\partial^2 \ell(\theta)}{\partial \gamma_0 \partial \gamma_0} = \frac{\partial}{\partial \gamma_0}\left[\frac{\partial \ell(\theta)}{\partial \gamma_0}\right]$$

$$= \frac{\partial}{\partial \gamma_0}\left[\sum_{i=1}^{N} \frac{\delta_i}{h_0(t)} \cdot 1 - \exp(X_i\beta) \cdot t\right]$$

$$= \frac{\partial}{\partial \gamma_0} \sum_{i=1}^{N} \delta_i \frac{1}{(\gamma_0 + \gamma_1 t + \gamma_2 \ln(t) + \gamma_3 t^2)}$$

$$= \sum_{i=1}^{N} \frac{\delta_i}{h_0(t)^2}$$

$$\frac{\partial^2 \ell(\theta)}{\partial \gamma_0 \partial \gamma_1} = \frac{\partial}{\partial \gamma_0}\left[\frac{\partial \ell(\theta)}{\partial \gamma_1}\right]$$

$$= \frac{\partial}{\partial \gamma_0}\left[\sum_{i=1}^{N} \frac{\delta_i}{h_0(t)} \cdot t_i - \exp(X_i\beta) \cdot \tfrac{1}{2} t^2\right]$$

$$= \frac{\partial}{\partial \gamma_0}\left[\sum_{i=1}^{N} \delta_i t_i \cdot \frac{1}{\gamma_0 + \gamma_1 t + \gamma_2 \ln(t) + \gamma_3 t^2}\right]$$

$$= \sum_{i=1}^{N} \frac{\delta_i t}{h_0(t)^2}$$

$$\frac{\partial^2 \ell(\theta)}{\partial \gamma_0 \partial \gamma_2} = \sum_{i=1}^{N} \frac{\delta_i \ln(t)}{h_0(t)^2}$$

$$\frac{\partial^2 \ell(\theta)}{\partial \gamma_1 \partial \gamma_0} = \frac{\partial}{\partial \gamma_1}\left[\frac{\partial \ell(\theta)}{\partial \gamma_0}\right]$$

$$= \sum_{i=1}^{N} \frac{\delta_i t}{h_0(t)^2}$$

$$\frac{\partial^2 \ell(\theta)}{\partial \gamma_1 \partial \gamma_1} = \sum_{i=1}^{N} \frac{\delta_i t^2}{h_0(t)^2}$$

notice matrix algebra pattern

$$\frac{\partial^2 \ell(\theta)}{\partial \alpha \partial \alpha'} = \begin{bmatrix} \sum \frac{\delta_i}{h_0(t)^2} & \sum \frac{\delta_i}{h_0(t)^2} t & \sum \frac{\delta_i}{h_0(t)^2} \ln(t) & \sum \frac{\delta_i}{h_0(t)^2} t^2 \\ \sum \frac{\delta_i}{h_0(t)^2} t & \sum \frac{\delta_i}{h_0(t)^2} t^2 & \sum \frac{\delta_i}{h_0(t)^2} t\ln(t) & \sum \frac{\delta_i}{h_0(t)^2} t^3 \\ \sum \frac{\delta_i}{h_0(t)^2} \ln(t) & \sum \frac{\delta_i}{h_0(t)^2} t\ln(t) & \sum \frac{\delta_i}{h_0(t)^2} \ln(t)^2 & \sum \frac{\delta_i}{h_0(t)^2} t^2\ln(t) \\ \sum \frac{\delta_i}{h_0(t)^2} t^2 & \sum \frac{\delta_i}{h_0(t)^2} t^3 & \sum \frac{\delta_i}{h_0(t)^2} t^2\ln(t) & \sum \frac{\delta_i}{h_0(t)^2} t^4 \end{bmatrix}$$

$$= \sum_{i=1}^{N} \frac{\delta_i}{h_0(t)^2} \cdot \begin{bmatrix} 1 \\ t \\ \ln(t) \\ t^2 \end{bmatrix} \begin{bmatrix} 1, t, \ln(t), t^2 \end{bmatrix}$$

4×4

$4 \times 1 \quad 1 \times 4$

Hessian log likelyhood function Box-cox formulation (continiuned)

$$\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\beta} = \frac{\partial}{\partial\alpha}\left[\frac{\partial}{\partial\beta}\ell(\theta)\right]$$

5x4

$$= \frac{\partial}{\partial\alpha}\left[\sum_{i=1}^{N}\delta_i X_i - \exp(X_i\beta)H(t_i)X_i'\right]$$

$$= \frac{\partial}{\partial\alpha}\left[\sum_{i=1}^{N}\delta_i X_i - \exp(X_i\beta)\left(\gamma_0' t + \tfrac{1}{2}\gamma_1 t^2 + \gamma_2(t\ln(t)-t) + \tfrac{1}{3}\gamma_3' t^3\right)X_i\right]$$

further expended

- γ_0': $\frac{\partial}{\partial\gamma_0'}\left[\frac{\partial}{\partial\beta}\ell(\theta)\right] = \sum_{i=1}^{N} -\exp(X_i\beta)\cdot t_i X_i = \begin{bmatrix}\sum_{i=1}^{N} -\exp(X_{1i}\beta_1)\cdot t_i X_{1i}\\ \sum_{i=1}^{N} -\exp(X_{2i}\beta_2)\cdot t_i X_{2i}\\ \vdots \\ \sum_{i=1}^{N} -\exp(X_{5i}\beta_5)\cdot t_i X_{5i}\end{bmatrix}$ (5x1)

- $\frac{\partial}{\partial\gamma_1}\left[\frac{\partial}{\partial\beta}\ell(\theta)\right] = \sum_{i=1}^{N} -\exp(X_i\beta)\tfrac{1}{2}t^2 X_i$

- $\frac{\partial}{\partial\gamma_2}\left[\frac{\partial}{\partial\beta}\ell(\theta)\right] = \sum_{i=1}^{N} -\exp(X_i\beta)(t\ln(t)-t)X_i$

- $\frac{\partial}{\partial\gamma_3'}\left[\frac{\partial}{\partial\beta}\ell(\theta)\right] = \sum_{i=1}^{N} -\exp(X_i\beta)\tfrac{1}{3}t^3 X_i$

noticing the algebra trick:

$$\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\beta} = \begin{bmatrix}\Sigma -\exp(x_{1i}\beta_1)x_{1i}\\ \Sigma -\exp(x_{2i}\beta_2)x_{2i}\\ \Sigma -\exp(x_{3i}\beta_3)x_{3i}\\ \Sigma -\exp(x_{4i}\beta_4)x_{4i}\\ \Sigma -\exp(x_{5i}\beta_5)x_{5i}\end{bmatrix}\left[t_i,\ \tfrac{1}{2}t^2,\ (t\ln(t)-t),\ \tfrac{1}{3}t^3\right]$$

5x4 = (5x1) (1x4)

combining these 3 parts yields the hessian.

hereafter the coefficients and their associated standard errors can be found in the following way using the Newton Raphson method.

$\theta_h = \theta_{h-1} - H^{-1}(\theta_{h-1})\,G(\theta_{h-1})$ met H := hessian, G = the gradient

$Var(\theta) = [-E[H(\theta)]]^{-1}$ SE(θ) zijn de $\sqrt{}$ v.d. diagonaal elementen.

References

Association, W. M. et al. (2001). World medical association declaration of helsinki. ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, 79(4):373.

Bruix, J., Sherman, M., Llovet, J. M., Beaugrand, M., Lencioni, R., Burroughs, A. K., Christensen, E., Pagliaro, L., Colombo, M., and Rodés, J. (2001). Clinical management of hepatocellular carcinoma. conclusions of the barcelona-2000 easl conference. *Journal of hepatology*, 35(3):421–430.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Chan, A. W., Zhong, J., Berhane, S., Toyoda, H., Cucchetti, A., Shi, K., Tada, T., Chong, C. C., Xiang, B.-D., Li, L.-Q., et al. (2018). Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. *Journal of hepatology*, 69(6):1284–1293.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *International journal of cancer*, 136(5):E359–E386.

Gönen, M. and Heller, G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92(4):965–970.

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., and Sriram, S. (2006). Modeling customer lifetime value. *Journal of service research*, 9(2):139–155.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., and Rosati, R. A. (1982). Evaluating the yield of medical tests. *Jama*, 247(18):2543–2546.

Helsen, K. and Schmittlein, D. C. (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Science*, 12(4):395–414.

Hunter, D. (2011). Stat 525 notes on the weibull hazard and survreg in r. (accessed June 6, 2019) - http://personal.psu.edu/drh20/525/weekly/weibull.pdf.

Jain, D. C. and Vilcassim, N. J. (1991). Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Science*, 10(1):1–23.

Johnson, P. J., Berhane, S., Kagebayashi, C., Satomura, S., Teng, M., Reeves, H. L., O'Beirne, J., Fox, R., Skowronska, A., Palmer, D., et al. (2015). Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach—the albi grade. *Journal of Clinical Oncology*, 33(6):550.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Lise, M., Bacchetti, S., Pian, P. D., Nitti, D., Pilati, P. L., and Pigato, P. (1998). Prognostic factors affecting long term outcome after liver resection for hepatocellular carcinoma: results in a series of 100 italian patients. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 82(6):1028–1036.

Moons, K. G., Altman, D. G., Vergouwe, Y., and Royston, P. (2009). Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *Bmj*, 338:b606.

Poon, R. T.-P., Fan, S. T., Lo, C. M., Liu, C. L., and Wong, J. (2002). Long-term survival and pattern of recurrence after resection of small hepatocellular carcinoma in patients with preserved liver function: implications for a strategy of salvage transplantation. *Annals of surgery*, 235(3):373.

Portolani, N., Coniglio, A., Ghidoni, S., Giovanelli, M., Benetti, A., Tiberio, G. A. M., and Giulini, S. M. (2006). Early and late recurrence after liver resection for hepatocellular carcinoma: prognostic and therapeutic implications. *Annals of surgery*, 243(2):229.

Rahman, M. S., Ambler, G., Choodari-Oskooei, B., and Omar, R. Z. (2017). Review and evaluation of performance measures for survival prediction models in external validation settings. *BMC medical research methodology*, 17(1):60.

Rodriguez, G. (2010). Parametric survival models. (accessed June 6, 2019) - https://data.princeton.edu/pop509/ParametricSurvival.pdf.

Rosset, S., Neumann, E., Eick, U., Vatnik, N., and Idan, Y. (2002). Customer lifetime value modeling and its use for customer retention planning. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 332–340. ACM.

Royston, P. and Altman, D. G. (2013). External validation of a cox prognostic model: principles and methods. *BMC medical research methodology*, 13(1):33.

Royston, P. and Sauerbrei, W. (2004). A new measure of prognostic separation in survival data. *Statistics in medicine*, 23(5):723–748.

Seetharaman, P. and Chintagunta, P. K. (2003). The proportional hazard model for purchase timing: A comparison of alternative specifications. *Journal of Business & Economic Statistics*, 21(3):368–382.

Steyerberg, E. W. and Harrell, F. E. (2016). Prediction models need appropriate internal, internal–external, and external validation. *Journal of clinical epidemiology*, 69:245–247.

Therneau, T. M. and Lumley, T. (2015). Package 'survival'.

van Houwelingen, H. C. (2000). Validation, calibration, revision and combination of prognostic survival models. *Statistics in medicine*, 19(24):3401–3415.

Vogel, A., Cervantes, A., Chau, I., Daniele, B., Llovet, J. M., Meyer, T., Nault, J. C., Neumann, U., Ricke, J., Sangro, B., Schirmacher, P., Verslype, C., Zech, C. J., Arnold, D., Martinelli, E., and Committee, E. G. (2018). Hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology*, 29(Supplement$_4$) : $iv238 - -iv255$.