# A Survival Analysis of Fixation Times in Reading

**Mattias Nilsson**
Department of Linguistics and Philology
Uppsala University
mattias.nilsson@lingfil.uu.se

**Joakim Nivre**
Department of Linguistics and Philology
Uppsala University
joakim.nivre@lingfil.uu.se

## Abstract

Survival analysis is often used in medical and biological studies to examine the time until some specified event occurs, such as the time until death of terminally ill patients. In this paper, however, we apply survival analysis to eye movement data in order to model the survival function of fixation time distributions in reading. Semiparametric regression modeling and novel evaluation methods for probabilistic models of eye movements are presented. Survival models adjusting for the influence of linguistic and cognitive effects are shown to reduce prediction error within a critical time period, roughly between 150 and 250 ms following fixation onset.

## 1 Introduction

During reading, the eyes move on average four times per second with substantial variation in individual fixation times, reflecting, at least in part, momentary changes in on-line language processing demands. In psycholinguistics, it is commonly assumed that derivative measures of fixation times, such as first fixation duration and gaze duration, reflect cognitive processes during reading. It is less clear, however, how the distribution of *individual* fixation times in reading is affected by on-line processing activities. In eye movement oriented research, models that attempt to model the distribution of individual fixation times in reading typically assume that saccadic movements are executed relatively randomly in time, with cognition only occasionally influencing the timing of saccades (Feng, 2006; McConkie et al., 1994; Yang and McConkie, 2001; Yang, 2006). In the model by Yang and McConkie (2001), for example, it is assumed that cognitive control can have a direct influence over the timing of saccades only with very long fixations, after the normal saccade has been canceled due to processing difficulty. Distributional models have often made use of the hazard function in order to analyze fixation times in reading (Feng, 2006; Feng, 2009; Yang and McConkie, 2001). The hazard function, in general terms, is a function of time representing the instantaneous risk that an event (e.g., a saccade) will occur at some specified time $t$ given that it has not occurred prior to time $t$.

In this paper, we model the distribution of fixation times in terms of a different but related quantity, namely the survival function, which defines the probability of being alive, i.e., the probability of the event not having occurred, at some specified time $t$. We use semiparametric regression for modeling the influence of linguistic and cognitive effects on the survival function, and we assess the results using survival-based time-dependent evaluation metrics. More specifically, our objectives are as follows. We first estimate the survival functions for ten different readers using the Kaplan-Meier method (Kaplan and Meier, 1958) in order to establish the general shape of the survival function for reading time data. Then, we estimate adjusted survival functions using Cox proportional hazards model (Cox, 1972) in order to examine the influence of stimulus variables on survival. Finally, we assess the adjusted survival models both with respect to the estimated effects of covariates and with respect to the predictive perfor-



*Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pages 107–115,
Portland, Oregon, June 2011. 

mance on held out data. The experiments we report in this paper are based on first fixation data (multiple refixations discarded) from the English section of the Dundee Corpus of eye movements in reading (Kennedy and Pynte, 2005).

The remainder of this paper is organized as follows. Section 2 introduces survival analysis and further motivates its use for modeling fixation durations in reading. Section 3 introduces and applies the Kaplan-Meier estimate, to compare the survival functions for the different readers in the corpus. Section 4 introduces the Cox proportional hazards model and section 5 outlines two methods for assessing the performance of survival models on new data. Section 6 presents the experimental evaluation of using Cox proportional hazards to model the survival function and summarize and discuss the results. Section 7, finally, concludes this paper.

## 2 Background

Survival analysis is the study and modeling of the time it takes for *events* to occur. Because methods for survival analysis originally were developed for studying the lifetime distributions of humans in an epidemiological context, the prototypical event in these studies is death and the primary variable of interest thus time until death occurs. The use of survival analysis, however, reaches beyond the clinical and medical sciences and survival methods apply to any study with a naturally identifiable starting point and a well-defined event of interest as end point. In non-medical contexts, survival analysis often goes by other names, such as *failure time analysis* or *reliability analysis* in engineering applications, *event history analysis* in sociology, or simply *duration analysis* in yet other contexts.

A defining characteristic of survival analysis is the ability to deal with censoring in a principled manner. Censoring is said to occur when only partial information about the survival time of an individual (human or other) is available. The most common type of censoring is referred to as right-censoring, which occurs when an individual is not subject to the event of interest during the course of the observation period. In this case, it is only known that the individual did not experience the event prior to the end of the study, but may perhaps do so at a later point in time and this piece of partial information about the censored survival time is included in the analysis.

There are, however, potentially good reasons for using survival analysis even in time-to-event studies that do not necessarily involve censored data, such as when measuring the brief periods of time elapsing between a stimulus appearance and a button-press in response-time studies, or when measuring the time between one saccade and the next during reading using eye-tracking. Such data is usually not normally distributed and even in the absence of censoring one may take advantage of the fact that survival data is almost never assumed to be normally distributed and the methods of survival analysis are designed to reflect this. Furthermore, if the correct parametric model for the data is not known, or one is not confident enough that a given parametric model is appropriate, the Cox proportional hazards model provides a robust[1] and widely used semiparametric regression method for time-to-event data. With respect to eye movement data, the Cox model appears appealing because, as pointed out by Feng (2006, 2009), several different types of distributions have been proposed as models of fixation times in reading at one time or another, suggesting there is indeed little agreement with respect to the correct parametric model.

### 2.1 Survival and Hazard

Survival data is commonly analyzed and modeled in terms of the survival and the hazard function.

The survival function describes the probabilistic relationship between survival and time. Let $T$ be a random variable denoting an individuals' survival time ($T \geq 0$). The survival function, $S(t)$, defines the probability that the individual survives longer than some specified time $t$:

$$S(t) = P(T > t) \tag{1}$$

The survival function is a monotonically decreasing function heading downward as $t$ increases and has the following theoretical properties: $S(0) = 1$, the probability of surviving past time 0 is 1; and $S(\infty) = 0$, eventually nobody survives and $S(t)$

[1] Cox proportional hazards model is "robust" in the sense that the results will be reasonably close to those obtained using the correct parametric model.

falls to zero as $t$ tends to infinity. Notice also that if $F(t)$ is the cumulative distribution function for $T$, the survival function, $S(t)$, is $1 - F(t)$.

In the present study, we let the event of interest be the occurrence of a saccade following a fixation period, and the most reasonable starting point for our measurements, at least in practice, appears to be the beginning, or the onset, of the fixation period. We will refer to the period onset-to-saccade interchangeably as the fixation time or the survival time. Thus, in this context, the survival function $S(t)$ simply expresses the probability that a given fixation lasts, or survives, longer than some specified time $t$.

The hazard function, $h(t)$, gives the instantaneous potential, per unit time, for an event to occur in some small time interval after $t$, given survival up to time $t$:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \quad (2)$$

The conditional probability in the formula for the hazard function expresses the probability that the survival time, $T$, will lie in the time interval between $t$ and $t + \Delta t$, given that the survival time is greater than or equal to $t$, where $\Delta t$ denotes an infinitesimally small interval of time. As already suggested, in this study the hazard function represents the instantaneous risk, or hazard, of a saccade occurring following a fixation at some specified time $t$, given that it has not yet occurred.

## 3 Kaplan-Meier Survival Estimate

The survival function for time-to-event data can be estimated from a sample of survival times, both censored and uncensored, using the Kaplan-Meier (aka Product-Limit) method. This is a non-parametric estimate of the survival function which orders the survival times, from the shortest to the longest, and adjusts, for each of the event times, the number of cases still alive according to the number of cases that were either subject to the event or censored in the previous time period.

Let $d_j$ be the number of saccades that occur at time $t_j$, and let $n_j$ be the number of fixations for which no saccade has yet occurred at time $t_j$. The Kaplan-Meier estimate of the survival function $S(t)$ is then given by:

$$\hat{S}(t) = \prod_{t(j) \leq t} (1 - \frac{d_j}{n_j}) \quad (3)$$

In the absence of censored observations, the Kaplan-Meier estimate is equivalent to the empirical distribution, and the cumulative survival probability at time $t_j$ reduces to the number of surviving fixations at time $t_j$ divided by the total number of fixations in the sample. The value of $\hat{S}(t)$ is constant between event times and the estimated function is therefore a step function that changes value only at times when one or more saccades occur.

### 3.1 Kaplan-Meier Survival of Reading Data

Feng (2009) estimated the hazard function for the distribution of fixation times for the readers of the Dundee corpus. Here, we give a complementary account by estimating the corresponding survival function for these readers using the Kaplan-Meier method. Figure 1 shows the survival functions for each reader plotted against time. Individual differences in the survival function emerge soon after 50 ms and at 100 ms we can spot different tendencies with respect to how fast or slow the curves decline. Overall, however, the behavior of the survival function appears similar across readers. Typically, the survival function begins with a slow decline up until about 150 ms and is then followed by a very rapid decline during the next 100 ms. Thus, we can see in figure 1 that the overall survival rates drop from about 80% to 20% in the time interval 150-250 ms. Thereafter, the function flattens again and at about 400 ms it appears to be converging between the readers. It is worth noting, however, that the reliability of the estimate decreases with time since the number of surviving fixations becomes fewer and fewer.

Median survival time is the point in time when 50% of the total number of fixations has been terminated by a saccade. It is thus read off the plot as the time where the probability of survival is 0.5. Median survival time ranges from 168 ms (reader *g*) to 220 ms (reader *b*). Mean median survival time across all ten readers is 191.4 ms with a standard deviation of 14.9 ms.

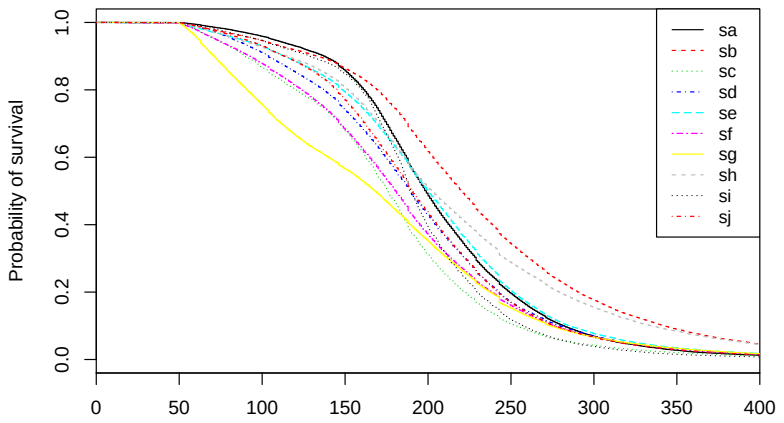


Figure 1: Kaplan-Meier curves for fixation durations showing the cumulative survival probability, following fixation onset, grouped by individual reader (subject a-j).

## 4 Cox Proportional Hazards Model

This section introduces the Cox proportional hazards model. We will later apply this model in the experimental part of the paper to obtain adjusted estimates of the survival function for the readers in the Dundee corpus.

The Cox proportional hazards model is a semi-parametric regression model for survival data relating survival time to one or more predictors or covariates. More precisely, the Cox model regresses the hazard function on a set of predictors, providing estimates of their effects in terms of hazard ratios. The Cox proportional hazards model has the following form:

$$h(t) = h_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n\} \quad (4)$$

where $h_0(t)$ is the baseline hazard function at time $t$, $x_1 \ldots x_n$ are the set of covariates or predictor variables, and $\beta_1 \ldots \beta_n$ are the corresponding coefficients to be estimated[2]. Thus, this model gives an expression for the hazard at time $t$ for a particular individual with a given set of covariates.

The baseline hazard, $h_0(t)$, represents the value of the hazard function when all covariates are equal to zero, and in the Cox model this baseline hazard is left unspecified and varies as a function of time. Since no assumptions are made with respect to the form or distribution of the baseline hazard, this can be regarded as the nonparametric part of the Cox proportional hazards model. However, the Cox model assumes a parametric form with respect to the effect of the predictors on the hazard. In particular, as seen in equation 4, the predictors are assumed to multiply hazard at any point in time. This is an important assumption of the Cox model referred to as the assumption of proportional hazards. It means that the hazard functions for any two individuals at any point in time should be proportional. In other words, if a certain individual has a risk of the event at some initial point in time that is twice as high as that of another individual, then, under the proportional hazards assumption the risk remains twice as high also at all later times. There are a variety of different graphical and goodness-of-fit based procedures that can be used to evaluate the proportional hazards assumption for survival data (see Kleinbaum and Klein (2005) for an overview.).

The parameter estimates in a fitted Cox model are commonly interpreted in terms of their hazard ratios. If $b_i$ is the value of the coefficient for predictor $x_i$, the exponentiated coefficient, $e^{b_i}$, gives the estimated hazard ratio for $x_i$. For continuous variables, the hazard ratio refers to the risk change associated with one unit increase in $x_i$, controlling for the effect of the other variables. A hazard ratio above one indicates a raised risk of the event occurring and the predictor is in this case thus negatively associated with survival. Correspondingly, a value below one indi-

[2]Parameter estimates in the Cox model are obtained by maximizing the "partial" likelihood, as opposed to the (full) likelihood. Details of procedures for parameter estimation can be found, for example, in Kalbfleisch and Prentice (1980).

cates a decreased risk and the predictor is thus positively associated with survival. Lastly, if the hazard ratio is equal to one, there is no indication of any associated risk change.

## 5 Assessment of Survival Models

Accurate prognoses are of critical importance in many areas where survival analysis apply, for instance in medical contexts where doctors have to estimate the expected remaining life time for terminally ill patients. Survival models are thus often assessed with respect to their predictive performance on novel data, in addition to the statistical significance of model covariates. We now briefly review two of the most commonly used measures for assessing the quality of survival models on independent data sets.

### 5.1 Prediction Error Curves

The prediction error for survival data is defined as a function of time and can be measured by the Brier score (Brier, 1950). Intuitively, if an individual is alive at time $t$, the predicted survival probability should be close to 1, and otherwise close to 0. The prediction error, or Brier score, at time point $t$ is defined as the mean squared error between the observed survival status $Y_i(t)$ for the individual $i$ at time $t$, which is equal to 1 if the individual is alive at $t$, and 0 otherwise, and the predicted survival probability for $i$ at time $t$:

$$\hat{BS}(t) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(t) - S_i(t)\}^2 \qquad (5)$$

The lower the Brier score, the lower the prediction error. Various benchmark values for the Brier score at time $t$ exists. The values 0.25 and 0.33, for example, correspond to a constant predicted survival probability of 50% and to a randomly predicted value between 0 and 1, respectively. Often, however, the Kaplan-Meier estimate of the survival function over the training sample is used. In this case, the benchmark prediction at time point $t$ corresponds to the proportion of individuals surviving past $t$, thus ignoring all available covariate information. By tracking the prediction error over time we get the prediction error curve (Graf et al., 1999) and a summary measure of the error for the whole observation period can be obtained by integrating over time (the integrated Brier score).

### 5.2 Concordance Index

The concordance index (Harrell et al., 1982), or *C*-index, estimates the probability that a given prediction agrees, or concurs, with the observed outcome. For uncensored data, the concordance index is given by the relative frequency of *concordant pairs* among all pairs of individuals. A pair is said to be concordant if the individual with the shorter survival time is also predicted by the model to have the highest risk of the two. Useful reference values for the concordance index are 0.5 which indicates that the predictions are no better than chance, and 1 which indicates that the model discriminates the pairs perfectly.

## 6 Experimental Evaluation

In order to study the influence of cognitive and linguistic effects on the survival function, the following experiment is performed. First, the Cox proportional hazards model is used to regress fixation times on five different stimulus variables associated with the current fixation, thus providing estimates of the hazard ratios for the effects of each variable adjusted for the other variables in the model. Second, we obtain adjusted survival functions, i.e. survival curves that adjust for the stimulus variables used as covariates, and we assess these curves with respect to the generalization error on held-out corpus data.

It is worth pointing out that regression studies on the Dundee Corpus of eye movements have been carried out before (e.g., Demberg and Keller, 2008; Pynte and Kennedy, 2006). Our experiment, however, differs from previous studies in at least three ways: (1) our goal is to model the survival function of fixation time distributions in reading, which means that we use the survival time of individual fixations as the unit of analysis; (2) we assess the survival model not only with respect to the estimated regression coefficients, but also with respect to the models' predictive performance on unseen data; (3) we use a semiparametric regression method for survival data which has not been previously applied, as far as we know, to reading-time data. It is also worth pointing out that although we believe that a

Table 1: Results of Cox proportional hazards model of fixation times in the Dundee Corpus section 01-16: hazard ratios (HR) and significance levels ($p$) for all covariates in the model, and for each individual model of reader $a$-$j$.

| | $a$ | | $b$ | | $c$ | | $d$ | | $e$ | | $f$ | | $g$ | | $h$ | | $i$ | | $j$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | HR | $p$ | HR | $p$ | HR | $p$ | HR | $p$ | HR | $p$ | HR | $p$ | HR | $p$ | HR | $p$ | HR | $p$ | HR | $p$ |
| Word length | 1.015 | $<.001$ | 0.983 | $<.001$ | 0.979 | $<.001$ | 0.988 | $<.001$ | 0.992 | $<.05$ | 0.992 | $<.01$ | 0.992 | $<.01$ | 0.985 | $<.001$ | 0.990 | $<.01$ | 0.987 | $<.001$ |
| Word frequency | 1.055 | $<.001$ | 1.042 | $<.001$ | 1.036 | $<.001$ | 1.051 | $<.001$ | 1.051 | $<.001$ | 1.014 | $<.001$ | 1.031 | $<.001$ | 1.028 | $<.001$ | 1.040 | $<.001$ | 1.044 | $<.001$ |
| Bigram probability | 1.108 | $<.001$ | 1.196 | $<.1$ | 1.092 | $<.05$ | 1.006 | $<.01$ | 1.013 | $<.05$ | 1.014 | $<.001$ | 0.953 | | 1.011 | $<.001$ | 1.003 | | 1.005 | $<.05$ |
| Surprisal | 1.001 | | 0.986 | $<.001$ | 0.994 | $<.01$ | 0.984 | $<.001$ | 0.998 | $<.01$ | 0.991 | $<.05$ | 1.002 | | 0.994 | | 0.993 | $<.05$ | 0.996 | $<.01$ |
| Entropy | 0.966 | $<.001$ | 0.986 | $<.01$ | 0.980 | $<.001$ | 0.988 | $<.01$ | 0.963 | $<.001$ | 1.002 | | 0.990 | $<.05$ | 0.992 | $<.05$ | 0.969 | $<.001$ | 0.978 | $<.001$ |

careful comparison of the results obtained using survival analysis to those reported for other regression methods would be useful and interesting, it is nevertheless beyond the scope of this paper.

Most of the stimulus variables included in the analysis have been shown to correlate with reading times in other regression studies: the number of letters in the word, the logarithm of the word's relative frequency (based on occurrences in the British National Corpus), the logarithm of the conditional (bigram) probability of the word (based on occurrences in the Google Web 1T 5-gram corpus (Brants and Franz, 2006)), the syntactic surprisal and entropy scores[3] (computed here using the probabilistic PCFG parser by Roark et al. (2009)). The surprisal (Hale, 2001) at word $w_i$ refers to the negative log probability of $w_i$ given the preceding words, computed using the prefix probabilities of the parser. A number of studies have previously established a positive relation between surprisal and word-reading times (Boston et al., 2008; Demberg and Keller, 2008; Roark et al., 2009). The entropy, as quantified here, approximates the structural uncertainty associated with the rest of the sentence, or what is yet to be computed (Roark et al., 2009).

In this experiment, we use the first 16 texts in the Dundee corpus for parameter estimation, and the following two texts, 17 and 18 for model assessment of the generalization error. To avoid introducing biases that may result from pooling distributional data together, we model each of the readers in the corpus separately. Prior to running the experiments, we also validated the Cox proportional hazards assumption using a goodness-of-fit approach based on the Schoenfeld residuals (Schoenfeld, 1982). The outcome of this test indicated a slight violation of the proportional hazards assumption. However, it is well-known that a slight violation may occur for large data samples, given that $p$-values can be driven by sample size (Kleinbaum and Klein, 2005).

[3]To ease interpretation of the estimated hazard ratios, no interaction terms were included in this model.

## 6.1 Results

### 6.1.1 Hazard Ratios

Table 1 shows the results of the Cox proportional hazards regression models. The estimated hazard ratio for each covariate along with the corresponding significance level is reported for each reader. Recall that a hazard ratio above one indicates a worse survival prognosis, i.e. shorter fixation times, while a hazard ratio below one indicates better survival, i.e. longer fixation times.

Overall, the effects go in the directions expected for these variables based on previous research. There is a significant positive effect of *word length* on survival for all but one reader. The hazard ratio for the significant effects ranges between 0.979 and 0.992. Word length thus decreases the hazard by about 1-2% for each additional letter in a word when adjusting for the effects of the other covariates in the model. *Word frequency* is significantly and negatively related to survival across all readers. More frequent words have shorter survival times. The average hazard ratio among the readers is 1.0392 and the estimated risk of a saccade increases thus on average by a factor of 1.0392 for each unit increase in log word frequency. *Bigram probability* is negatively and significantly related to survival for eight readers with an average hazard ratio of 1.0569. *Surprisal* is significantly and positively related to survival for seven readers. Among these, the hazard decreases by 1% for each unit increase in surprisal. *Entropy* has a significant and positive effect on survival on all but one readers. The hazard ratios range between 0.963 and 0.992, corresponding to a de-

| | Brier score $t$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $t$.100 | | $t$.150 | | $t$.200 | | $t$.250 | | $t$.300 | |
| Model | Cox | KM | Cox | KM | Cox | KM | Cox | KM | Cox | KM |
| a | 0.05 | 0.05 | 0.14 | 0.15 | 0.24 | 0.25 | 0.14 | 0.15 | 0.05 | 0.06 |
| b | 0.05 | 0.05 | 0.12 | 0.13 | 0.23 | 0.25 | 0.21 | 0.23 | 0.12 | 0.13 |
| c | 0.13 | 0.13 | 0.23 | 0.24 | 0.17 | 0.18 | 0.06 | 0.07 | 0.02 | 0.02 |
| d | 0.07 | 0.07 | 0.17 | 0.18 | 0.23 | 0.25 | 0.15 | 0.16 | 0.06 | 0.06 |
| e | 0.05 | 0.05 | 0.15 | 0.15 | 0.23 | 0.25 | 0.14 | 0.15 | 0.05 | 0.05 |
| f | 0.09 | 0.09 | 0.21 | 0.21 | 0.22 | 0.23 | 0.12 | 0.12 | 0.06 | 0.06 |
| g | 0.16 | 0.16 | 0.23 | 0.23 | 0.24 | 0.25 | 0.12 | 0.13 | 0.07 | 0.07 |
| h | 0.07 | 0.07 | 0.15 | 0.15 | 0.24 | 0.25 | 0.20 | 0.20 | 0.12 | 0.12 |
| i | 0.04 | 0.04 | 0.13 | 0.13 | 0.23 | 0.25 | 0.10 | 0.10 | 0.03 | 0.03 |
| j | 0.06 | 0.06 | 0.18 | 0.19 | 0.23 | 0.25 | 0.12 | 0.12 | 0.05 | 0.05 |
| Avg. | 0.077 | 0.077 | 0.171 | 0.176 | 0.226 | 0.241 | 0.136 | 0.143 | 0.063 | 0.065 |

Table 2: Prediction error on held-out data between the observed survival status and the predicted survival probability at different times $t$, for Kaplan-Meier and Cox-model adjusted survival, and for all models of readers $a$-$j$.

creased risk by 1-4% per additional unit increase, after adjusting for the effects of the other predictors. While Frank (2010) recently showed that *sentence* entropy, i.e. non-structural entropy, accounts for a significant fraction of the variance in reading times, our results provide additional support for the influence of *structural* sentence entropy on reading times. Moreover, it is noteworthy that the effect of entropy appears reliably robust in individual first fixation times, suggesting that the effects of structural processing demands can be immediate rather than delayed in the eye movement record.

### 6.1.2 Adjusted Survival

We summarize the results of the evaluation of the adjusted survival function on held-out data in table 2 and in table 3. Table 2 shows the Brier score computed at different points in time in the interval 100 to 300 ms. Results are reported both for the Kaplan-Meier estimate of the survival function and for the fitted Cox-models. We present the results for each individual model. The bottom row gives the results obtained when averaging over all models at the specified time $t$.

Recall that the Brier score, or prediction error, at any specified time $t$, is computed over *all* fixations in the held-out set and gives the average of the squared distances between the actual survival status and the predicted survival probability at time $t$. Although the differences between the Cox-model and the Kaplan-Meier estimate are small overall, there are two subtle but notable results. First, the adjusted survival model is never underperforming the Kaplan-Meier survival estimate. The prediction error of the Cox model is consistently lower or equal to the Kaplan-Meier prediction error at each time point and for each reader. Second, in comparison to the Kaplan-Meier error, the prediction error of the adjusted model is systematically lower in the time window 150-250 ms, but essentially the same prior to, and after, this time period. This is readily reflected in the average scores, for example. One interpretation of these small but systematic differences suggests that there is a limited period, approximately no earlier than 150 ms. and no later than 250 ms. on average, during which the covariates in the model are primarily influencing the survival time. Before and after this period, the stimulus variables of the fixated word appear to have little or no influence on the time when saccades are generated. In other words, we observe an improved agreement to the observed data in the interval 150-250 ms. under the assumption that each individual fixation has an independent survival function whose value at time $t$ is influenced by the specific values for the stimulus variables of the fixation. Recall that the benchmark, the Kaplan-Meier estimate, in contrast, assumes one and the same underlying survival function for all fixations, ignoring all available covariate information. By plotting the

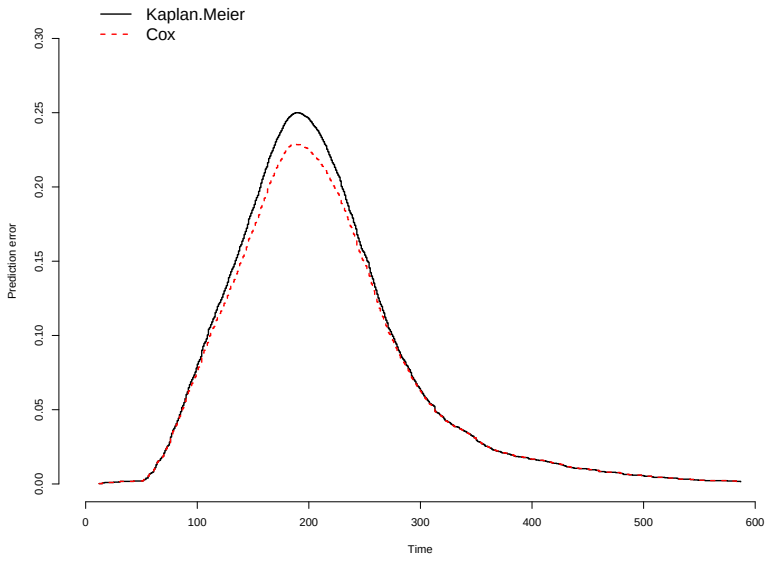


Figure 2: Prediction error curves on held-out data between the observed survival status and the predicted survival probability, for Kaplan-Meier and Cox-model adjusted survival, for the model of reader $d$.

| Model | IBSC | $C$-index |
|---|---|---|
| Kaplan-Meier | 0.041 | 0.582 |
| Cox | 0.043 | 0.598 |

Table 3: Integrated Brier score (IBSC) and Concordance index ($C$-index) on held-out data, for Kaplan-Meier and Cox-model adjusted survival, averaged over the results obtained for each model of reader $a$-$j$.

time-dependent prediction error, subtle differences in survival over the time course are more easily spotted. Figure 2 shows, as an example, the prediction error curve for one of the models.

Table 3 gives the integrated brier score, i.e., the prediction error obtained when integrating over all event times, and the concordance index $C$, for both the Kaplan-Meier estimate and the Cox model. These results are averaged over the results of the individual models. The integrated Brier score verifies that the Cox model fares somewhat better, although the impact of the model variables appears limited in time. The $C$-value for both the Kaplan-Meier and the Cox model is significantly better than chance (0.5). A $C$-value of 0.6 - 0.7 is a common result for survival data.

## 7 Conclusion

In this paper we applied survival analysis to model fixation times in reading. In particular, we modeled the survival function of fixation time distributions using the Kaplan-Meier estimate, and the Cox proportional hazards model to adjust for cognitive and linguistic effects on survival. The adjusted survival models were assessed with respect to the effect of covariates on hazard rates, and with respect to their predictive performance using evaluation metrics that are novel in the context of eye-movement and psycholinguistic modeling.

The results of the analysis suggests that: (1) structural sentence entropy influences survival, i.e., increasing structural uncertainty about the rest of the sentence decreases the risk of moving the eyes; (2) stimulus variables associated with the current fixation influence the survival of the fixation in a limited time frame, roughly between 150 and 250 ms following onset; and (3) linguistic and cognitive effects may influence the timing of saccades earlier than is sometimes assumed.

Looking ahead, important topics to be investigated in the future include *frailty models* and *competing risks survival analysis*. Frailty models are survival-based regression models with random effects, designed to account for variance due to individual-level factors otherwise unaccounted for.

Competing risks survival analysis apply to situations where a finite number of different types of events are possible, but only one of the events can actually occur per individual, e.g., dying from either lung cancer or stroke. In the current study we did not differentiate between different types of events following a fixation. A competing risks analysis, however, would let us differentiate between different types of saccades and study the influence of predictors on the survival function based on the type of the saccade following a fixation, e.g., whether it is a forward-directed saccade, refixation or regression. These and other issues will be addressed.

## References


Marisa F. Boston, John Hale, Reinhold Kliegl, Umesh Patil, and Shravan Vasishth. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *Journal of Eye Movement Reasearch*, 2:1–12.

Thorsten Brants and Alex Franz. 2006. Web 1T 5-gram Version 1. Linguistic Data Consortium.

Glenn W. Brier. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3.

David R. Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34:187–220.

Vera Demberg and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109:193–210.

Gary Feng. 2006. Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research*, 7:70–95.

Gary Feng. 2009. Time course and hazard function: A distributional analysis of fixation duration in reading. *Journal of Eye Movement Research*, 3:1–23.

Stefan L. Frank. 2010. Uncertainty reduction as a measure of cognitive processing effort. In *Proceedings of the ACL Workshop on Cognitive Modeling and Computational Linguistics*.

Erika Graf, Schmoor Claudia, Sauerbrei Will, and Schumacher Martin. 1999. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18:2529–2545.

John Hale. 2001. A probabilistic early parser as a psycholinguistic model. In *Proceedings of the second conference of the North American chapter of the Association for Computational Linguistics*, volume 2, pages 159–166.

Frank E. Jr Harrell, Robert M. Califf, David B. Pryor, Kerry L. Lee, and Rober A. Rosati. 1982. Evaluating the yield of medical tests. *Journal of the American Medical Association*, 247:2543–2546.

John D. Kalbfleisch and Ross L. Prentice. 1980. *The statistical analysis of failure time data*. Wiley.

Edward L. Kaplan and Paul Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53:457–481.

Alan Kennedy and Joel Pynte. 2005. Parafoveal-on-foveal effects in normal reading. *Vision Research*, 45:153–168.

David G. Kleinbaum and Mitchell. Klein. 2005. *Survival analysis: A self-learning text*. Springer.

George W. McConkie, Paul W. Kerr, and Brian P. Dyre. 1994. What are normal eye movements during reading: Toward a mathematical description. In J. Ygge and G. Lennerstrand (Eds.), editors, *Eye movements in reading: Perceptual and language processes*, pages 315–327. Oxford: Elsevier.

Joel Pynte and Allan Kennedy. 2006. An influence over eye movements in reading exerted from beyond the level of the word: Evidence from reading english and french. *Vision Research*, 46:3786–3801.

Brian Roark, Asaf Bachrach, Carlos Cardenas, and Christophe. Pallier. 2009. Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 324–333.

David Schoenfeld. 1982. Partial residuals for the proportional hazards model. *Biometrika*, 69:51–55.

Shun-nan Yang and George W. McConkie. 2001. Eye movements during reading: a theory of saccade initiation times. *Vision Research*, 41:3567–3585.

Shun-nan Yang. 2006. A oculomotor-based model of eye movements in reading: The competition/interaction model. *Cognitive Systems Research*, 7:56–69.