Modeling post-ART AIDS survival data using Weibull parametric models with changepoints

Constantin T. Yiannoutsos, Ph.D.
Division of Biostatistics
Indiana University School of Medicine
Indianapolis, Indiana, USA

Introduction of the data analysis problem

One thousand one hundred and twenty (1,120) Human Immunodeficiency Virus (HIV)-infected patients were administered antiretroviral medications (ARV) as part of the study.

The duration of follow-up in this patient cohort was as short as ten days and as long as almost 33 months (median follow-up time 26.9 months).

One hundred and five (105) subjects died during the study.

A Kaplan-Meier plot of patient mortality is shown in the following figure.

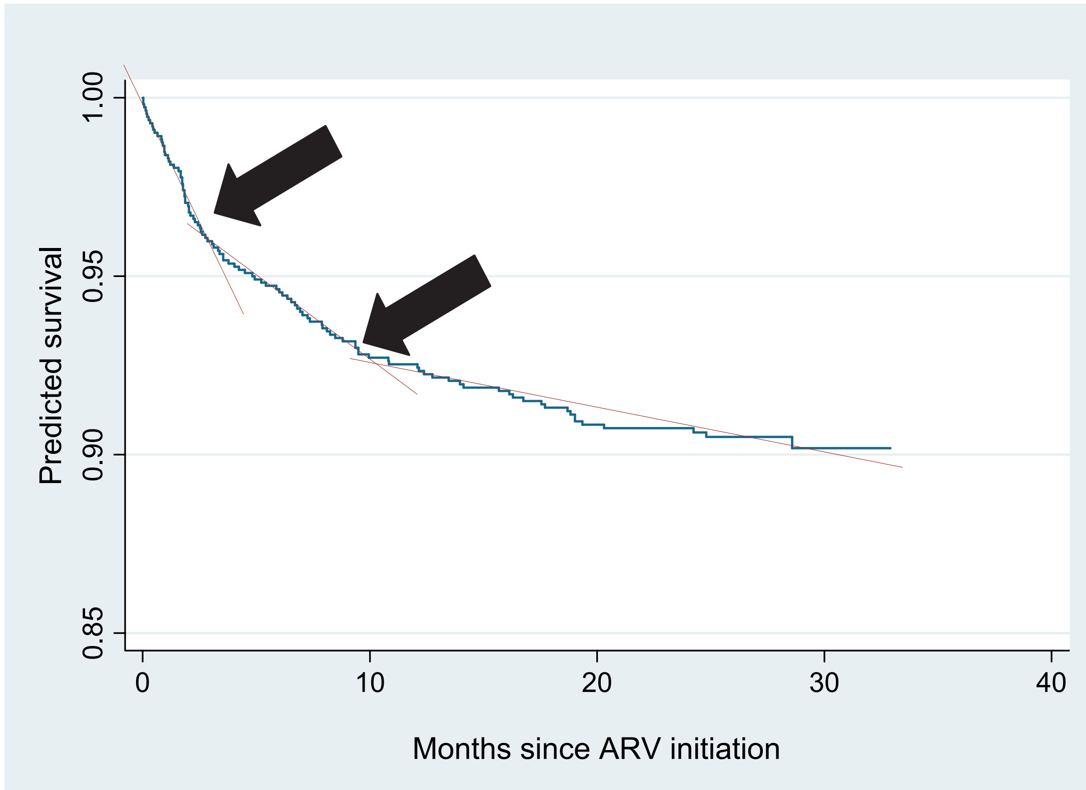

Figure 1: Kaplan Meier summary of patient mortality

Inspection of the Kaplan-Meier curve suggests that there may be times where the hazard (instantaneous risk) of death is higher than others (see arrows on the figure). The following table shows the percent of patients dying in every period post ARV initiation.

Table 1: Death rate (per 100 person-years) during the first 24 months of follow-up

Time period				
$<$ 3 months	3-6 months	6-12 months	12-18 months	18-24 months
16.4%	5.6 %	4.5%	2.65 %	1.27 %

Weibull model of patient survival

A frequently used mathematical model of patient survival is based on the Weibull probability distribution function for survival time T.

$$f(t) = \left(\frac{t}{\delta}\right)^{\alpha} \exp\left(-\frac{t}{\delta}\right)^{\alpha}$$

with δ and α the scale and shape parameters respectively.

The Weibull model is a generalization of the common exponential survival model and is reasonable to use here as, in contrast to the exponential model, it does not assume constant hazard of death.

A Weibull analysis of the study data is shown in Figure 2.

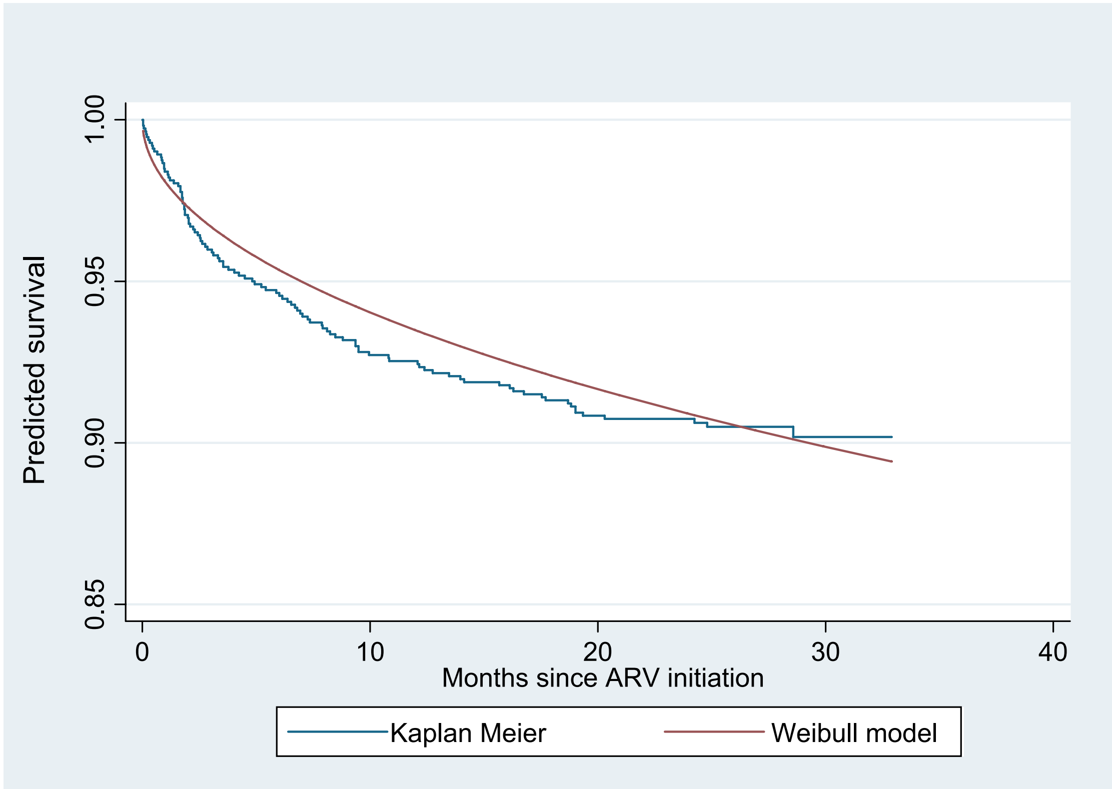

Figure 2: Kaplan Meier versus Weibull survival analysis

It is clear from the figure, that the Weibull model underestimates patient mortality, particularly early on after ARV initiation.

For this reason it would be useful to consider more flexible models that take into consideration the possible changes in hazard over various periods after initiation of therapy.

These are presented in the next section.

Weibull model with one changepoint

In the Weibull model, the cumulative hazard $H(t_i)$ for the i^{th} subject is

$$H(t_i) = \left(\frac{t_i}{\delta} \right)^{\alpha}$$

so that the log-cumulative hazard has the linear form in log time

$$g(t_i) = \alpha \ln t_i + \lambda \tag{1}$$

where $\lambda = -\alpha \ln \delta$.

In the simplest case of a single change point a, we will have two Weibull models, before and after the change point, with scale and shape parameters (δ_1, α_1) and (δ_2, α_2) respectively and log-cumulative hazards would be as in equation (1), above i.e.,

$$g(t_i) = c_i(\alpha_1 \ln t_i + \lambda_1) + (1 - c_i)(\alpha_2 \ln t_i + \lambda_2)$$

where c_i is the change point indicator i.e.,

$$c_i = \begin{cases} 1 & \text{if } \; t_i \leq a \\ 0 & \text{otherwise} \end{cases}$$

We would like to restrict the possible values of the λ's so that the log-cumulative hazards meet at the change point a. This would require that

$$\alpha_1 \ln a + \lambda_1 = \alpha_2 \ln a + \lambda_2$$

and thus,

$$\lambda_2 = \lambda_1 + (\alpha_1 - \alpha_2) \ln a$$

and the log-cumulative hazard for the single-change-point model will have the form (Jiwani, 2005)

$$g(t_i) = \lambda_1 + c_i(\alpha_1 \ln t_i) + (1 - c_i) [\alpha_2 \ln t_i + (\alpha_1 - \alpha_2) \ln a] \qquad (2)$$

The probability density function is given from the usual relationship

$$f(t_i) = h(t_i)\exp\left[-H(t_i)\right]$$

while the likelihood function in general is given by

$$L_i = [f(t_i)]^{w_i}[S(t_i)]^{(1-w_i)} = h(t_i)^{w_i}e^{-H(t_i)}$$

where

$$w_i = \begin{cases} 1 & \text{if observed death} \\ 0 & \text{if censored} \end{cases}$$

The log-likelihood in the single-changepoint model is

$$\log L = \sum_{i=1}^{n} \{w_i[c_i \ln \alpha_1 + (1-c_i)\ln \alpha_2] \\ -w_i \ln t_i + w_i \ln H(t_i) - H(t_i)\} \qquad (3)$$

By maximizing (3) with respect to α_1, α_2 and λ_1, for given values of the change point a, we obtain both the optimal change point a and the parameters of the piece-wise Weibull distributions (α_1, δ_1) and (α_2, δ_2).

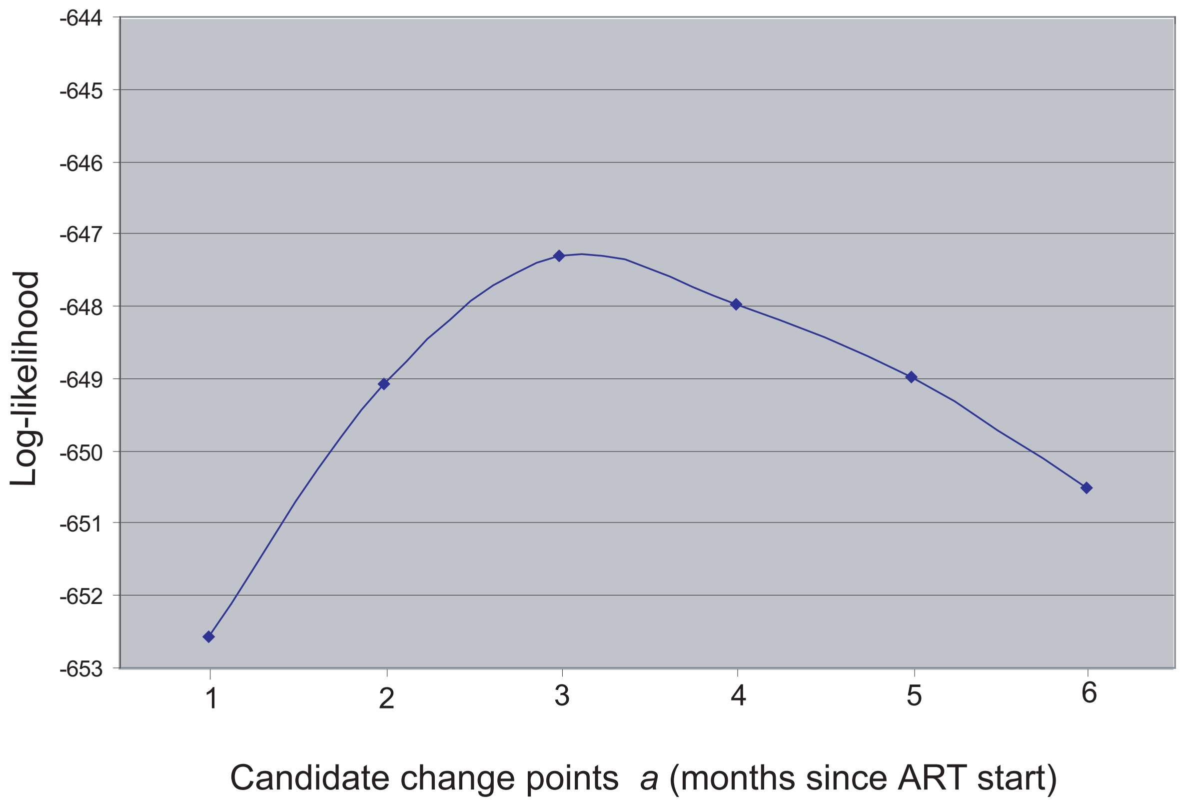

Figure 3: Log-likelihood search for optimal change point a

Carrying out this analysis, and searching through possible candidate change points a, indicates that the optimal change point is $a = 3$, that is, three months after initiation of antiretroviral therapy (ART) (Figure 3).

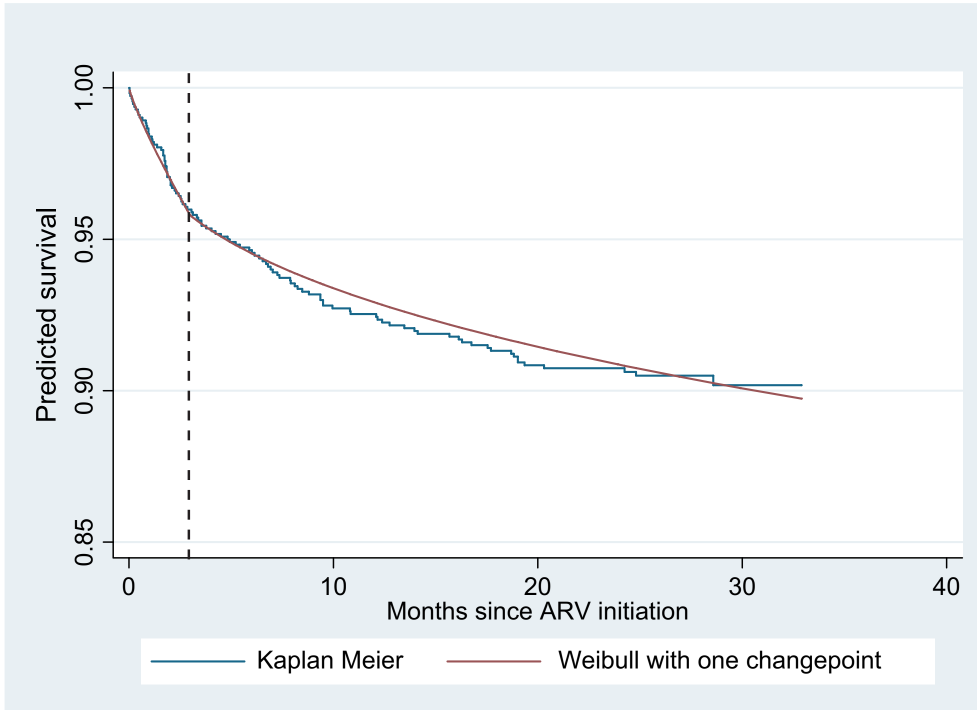

Figure 4: Kaplan Meier survival estimates versus Weibull with one change point

Hazard plot in the Weibull one-changepoint model

Another informative figure of the implication of the change point model is the hazard plot shown in Figure 5.

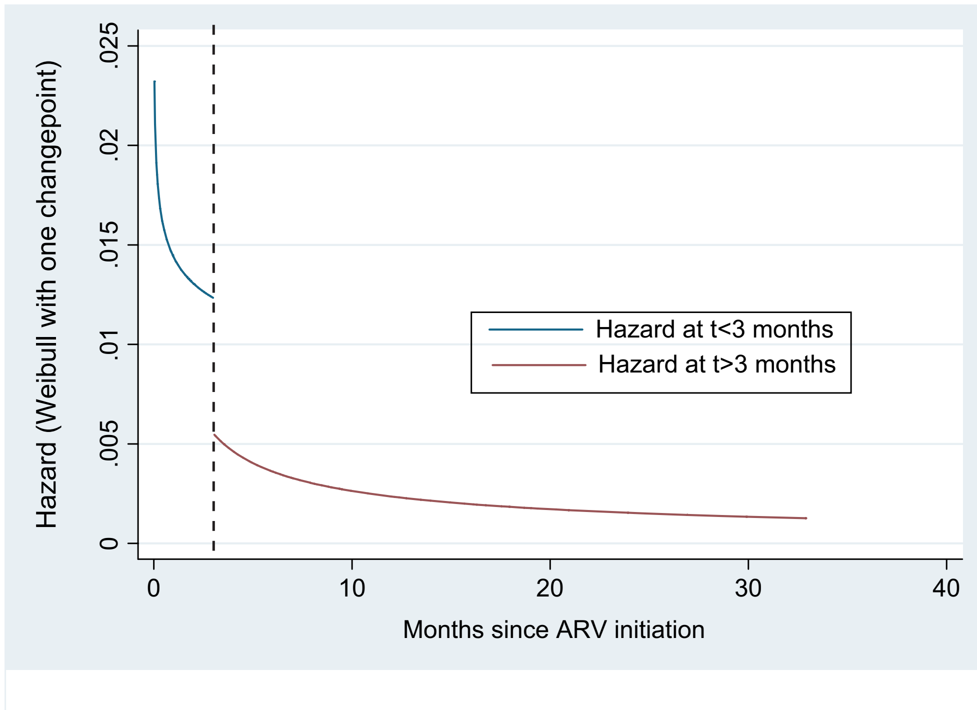

Figure 5: Hazard plot of the Weibull with one changepoint at $t = 3$ months

Discussion of the Weibull one-changepoint model

The estimated Weibull survival is shown in Figure 4 along with the Kaplan-Meier reference survival estimate.

The impression is that the fit, particularly in the first six months, is particularly good but the survival estimate still underestimates the mortality rate in subsequent months.

The conclusion from Figures 4 and 5 is that the change point model reflects a situation where a very high hazard of death in the first three months after ARV initiation is followed by a period of lower hazard.

It is also worth noting that the construction of the model ensures that the piece-wise cumulative hazards, and thus the survival curves, will meet at the change point, but this is not the case with the hazard curves.

Weibull model with two changepoints

We consider a Weibull model with two changepoints. The three-changepoint model for the Weibull is derived from Noura & Read (1990). The log-cumulative-hazard $g(t)$ for the two-changepoint case (a_1 and a_2) is given by the following relationship

$$g(t) = \lambda_1 + c_1\alpha_1 \ln(t) + c_2\alpha_2 \ln(t) + (\alpha_1 - \alpha_2)\ln(a_1) \\ +c_3\alpha_3 \ln(t) + (\alpha_1 - \alpha_2)\ln(a_1) + (\alpha_2 - \alpha_3)\ln(a_2) \tag{4}$$

where c_j is the change point indicator and $j = 1, 2, 3$. Setting $a_0 = 0$ and $a_3 = +\infty$ we have,

$$c_i = \begin{cases} 1 & \text{if } a_{j-1} \le t_i < a_j \\ 0 & \text{otherwise} \end{cases}, \ j = 1, 2, 3$$

and the log-likelihood is

$$\ln L = w_i c_1 \ln(\alpha_1) + c_2 \ln(\alpha_2) + c_3 \ln(\alpha_3) - w_i \ln(t) + w_i \ln H(t) - Ht \tag{5}$$

Performing this analysis and searching various combinations of changepoints the optimal two changepoints were found to be $a_1 = 3$ and $a_2 = 10$ (Figure 6).

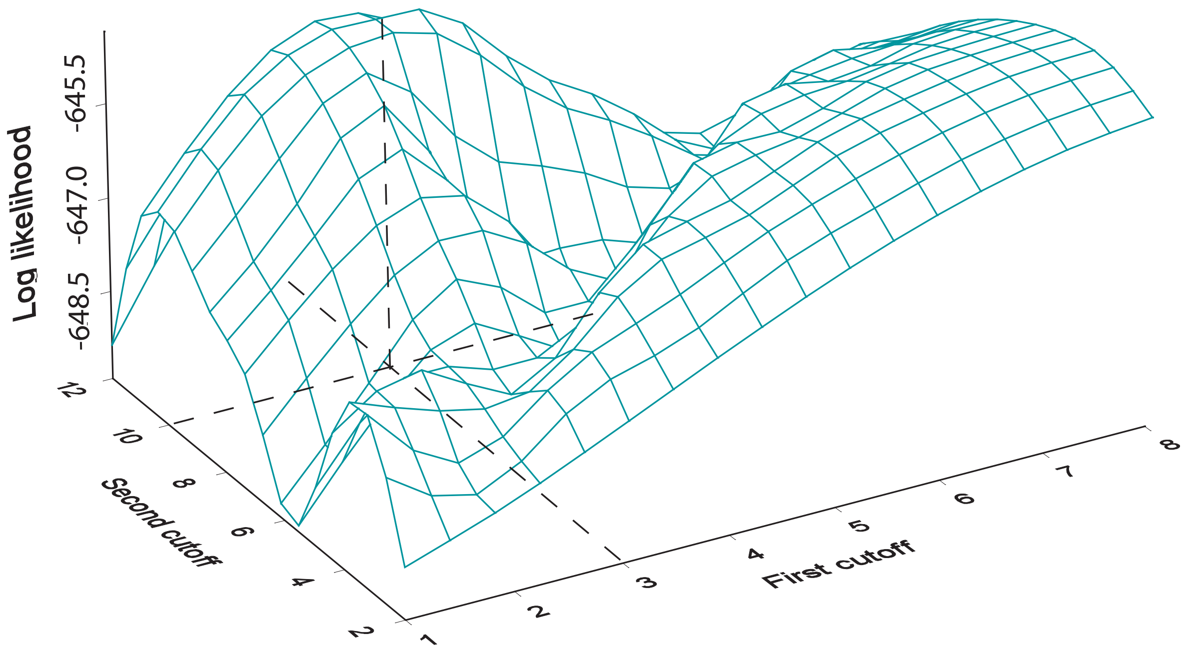

Figure 6: Log-likelihood search for optimal changepoints a_1 and a_2

The new survival estimate is shown in Figure 7.

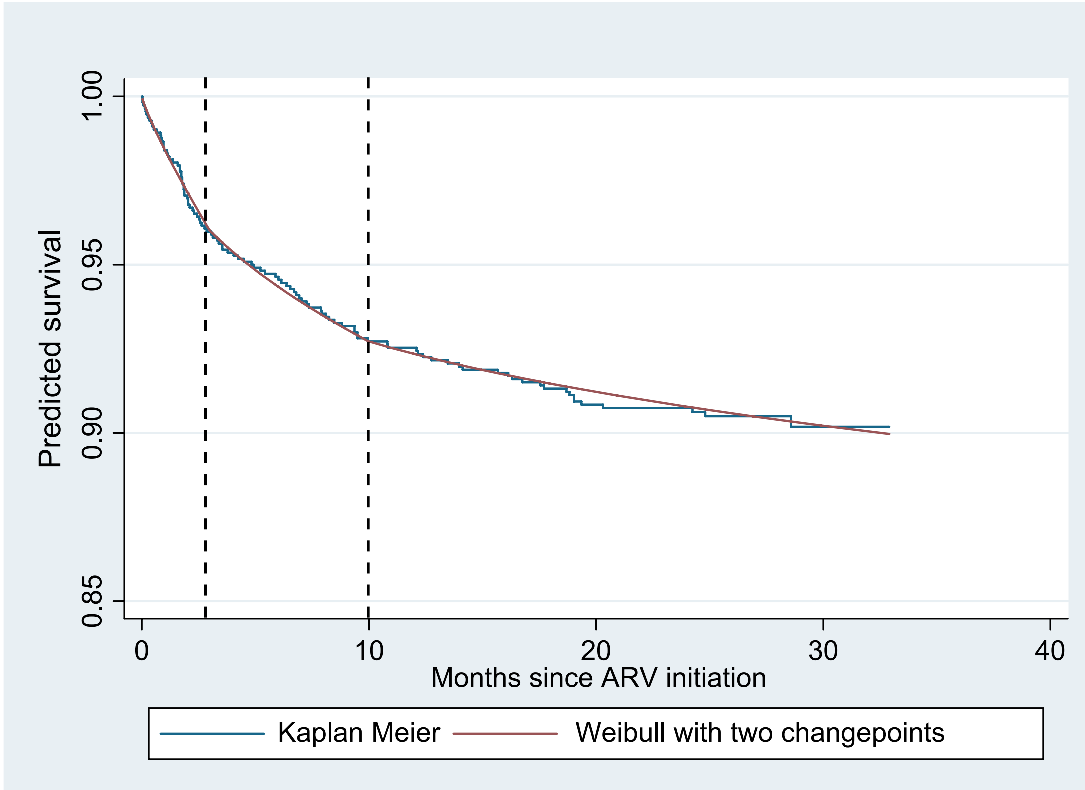

Figure 7: Kaplan Meier survival estimates versus Weibull with two change points

A hazard plot of the two-changepoint problem is shown in Figure 8.

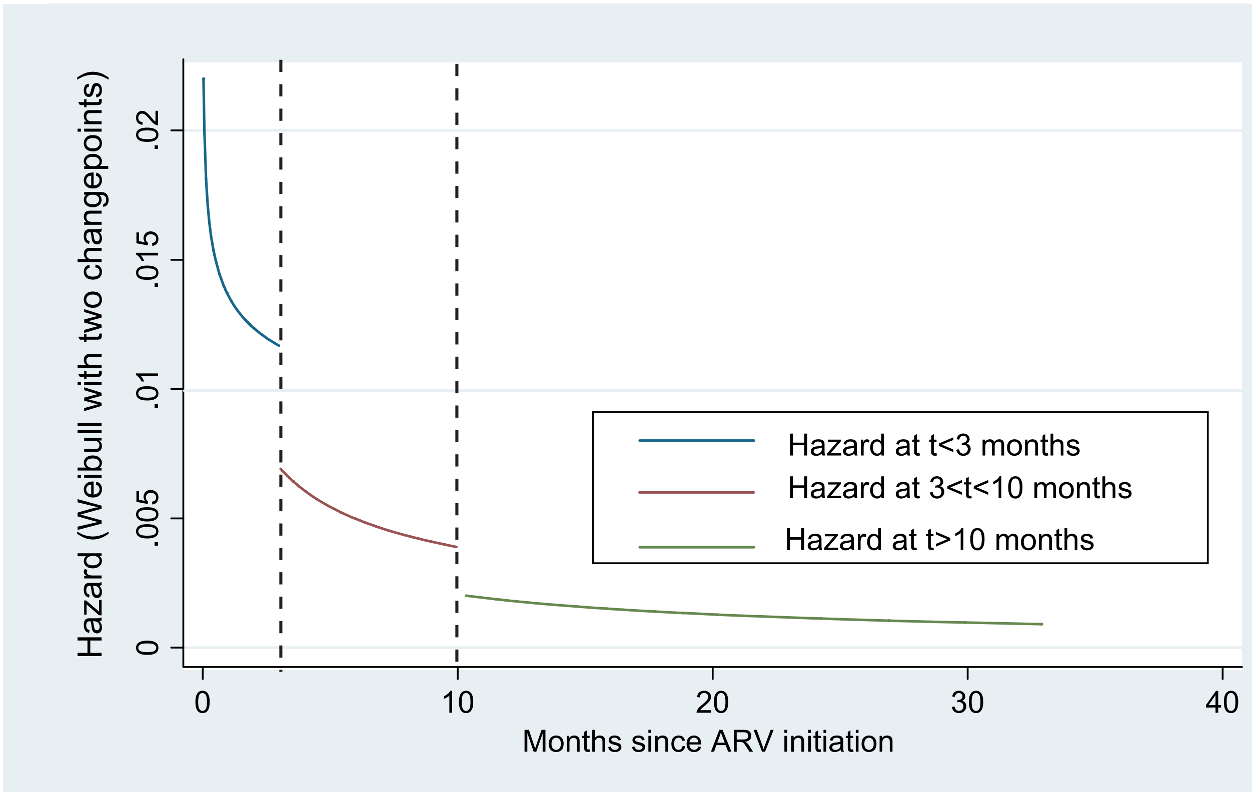

Figure 8: Hazard plot of the Weibull with two change points at $t = 3$ and $t = 10$ months

Discussion of the Weibull two-changepoint model

The fit from the two-changepoint Weibull model is very good throughout the post-ARV period.

The hazard plot implies that there are three periods after initiation of ARV: An initial period of high risk immediately after initiation ofL ARV that extends up to three months, followed by an intermediate risk period between three and ten months, which itself is followed by a period of stabilized (almost constant) low risk after 10 months.

Bibliography

Jiwani LJ Parametric Changepoint Survival Model with Application to Coronary Artery Bypass Graft Surgery Data. MS thesis, Simon Fraser University, 2005.
URL
http://www.stat.sfu.ca/people/alumni/Theses/Jiwani-2005.pdf.

Noura AA, Read KLQ. Proportional hazards changepoint models in survival analysis. *Appl Statist*, **39** pp. 241-253