Chapter 5

APPLICATION TO DATA FROM THE INSURANCE INDUSTRY

5.1 Description of the Data Set

5.1.1 Introduction

An extensive data set from an insurance company, containing information on policies written over the last few years, is available and permission has been given by this company to use this data set to illustrate the theoretical principles developed in the previous two chapters.

5.1.2 The raw data set of policies

A subgroup of policies is formed by selecting only mortgage protection policies written during four selected months, namely March 1998, June 1998, November 1998 and March 1999. This subgroup or smaller data set consists of the lifetimes of 10077 policies, together with some concomitant information on other variables such as age of the policyholder, credit turnover of his bankaccount and a score value, determined by the company.

Consider the following experimental design as illustrated in Figure 5.1. The 10077 policies enter the study at four different times (**staggered entry**). The event to be occurred is a lapse. The lifetime of a policy is measured from inception date up to the lapsing date. If the lapsing date is prior to the pre-determined cut-off date of 15 April 2001, then the lifetime is **observed** (an uncensored observation). If a policy is still in force (alive) when the termination point is reached, the lifetime of this policy is **right-censored**.

From the 2586 policies with entry date March 1998 (inception dates between 1 March 1998 and 31 March 1998), a total of 1666 policies have lifetimes 37 months and more and thus were right-censored. From the 2809 policies with entry date June 1998 (inception dates between 1 June 1998 and 30 June 1998), a total of 1924 policies have lifetimes 34 months and more and were censored. From the 2286 policies with entry date November 1998 (inception dates between 1 November 1998 and 30 November 1998), a total of 1674 policies have lifetimes 28 months and more and were censored. From the 2396 policies with entry date March 1999 (inception dates between 1 March 1999 and 31 March 1999), a total of 1848 policies have lifetimes 24 months and more and were censored.

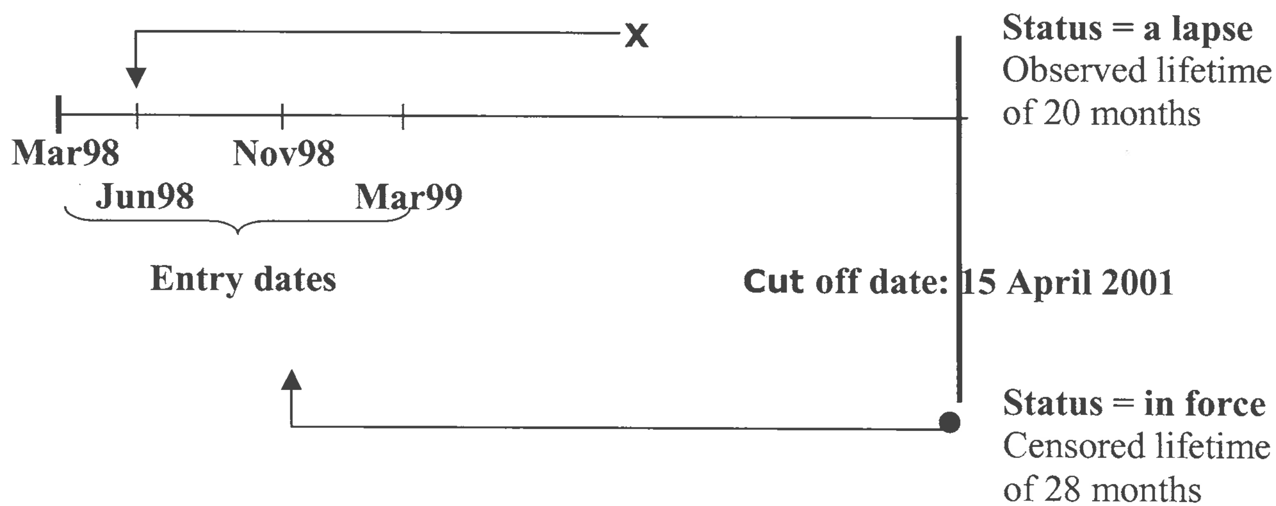

Figure 5.1: **Experimental design for illustrative data set**

5.1.3 The grouped data set of policies

The lifetimes of the policies that enter the study at March 1998 (called the first sample of size 2586) can be grouped into seven adjacent, non-overlapping fixed intervals

$$[0;12),[12;17),[17;24),[24;28),[28;34),[34;37) \quad \text{and} \quad [37;\infty).$$

The lifetimes of the policies that enter the study at June 1998 (called the second sample of size 2809) can be grouped into six adjacent, non-overlapping fixed intervals

$$[0;12),[12;17),[17;24),[24;28),[28;34) \quad \text{and} \quad [34;\infty).$$

The lifetimes of the policies that enter the study at November 1998 (called the third sample of size 2286) can be grouped into five adjacent, non-overlapping fixed intervals

$$[0;12),[12;17),[17;24),[24;28) \quad \text{and} \quad [28;\infty).$$

The lifetimes of the policies that enter the study at March 1999 (called the fourth sample of size 2396) can be grouped into four adjacent, non-overlapping fixed intervals

$$[0;12),[12;17),[17;24) \quad \text{and} \quad [24;\infty).$$

The four samples are assumed to be independent samples from multinomial populations. Four frequency distributions are formed when the observed and censored lifetimes of all the policies are grouped into the different class intervals and are shown in Table 5.1.

Table 5.1: **Frequency distributions of the four samples**

Interval	Class Intervals				Frequency Vector				Vector of Upper Bounds			
number	March 98	June 98	Nov 98	March 99	f_1	f_2	f_3	f_4	x_1	x_2	x_3	x_4
first	$[0,12)$	$[0,12)$	$[0,12)$	$[0,12)$	66	118	154	175	12	12	12	12
second	$[12,17)$	$[12,17)$	$[12,17)$	$[12,17)$	158	166	99	166	17	17	17	17
third	$[17,24)$	$[17,24)$	$[17,24)$	$[17,24)$	254	229	242	207	24	24	24	24
fourth	$[24,28)$	$[24,28)$	$[24,28)$	$[24,\infty)$	157	200	117	1848	28	28	28	
fifth	$[28,34)$	$[28,34)$	$[28,\infty)$		250	172	1674		34	34		
sixth	$[34,37)$	$[34,\infty)$			35	1924			37			
seventh	$[37,\infty)$				1666							
Total					2586	2809	2286	2396				

Figure 5.2 shows the histograms of the four relative frequency distributions.

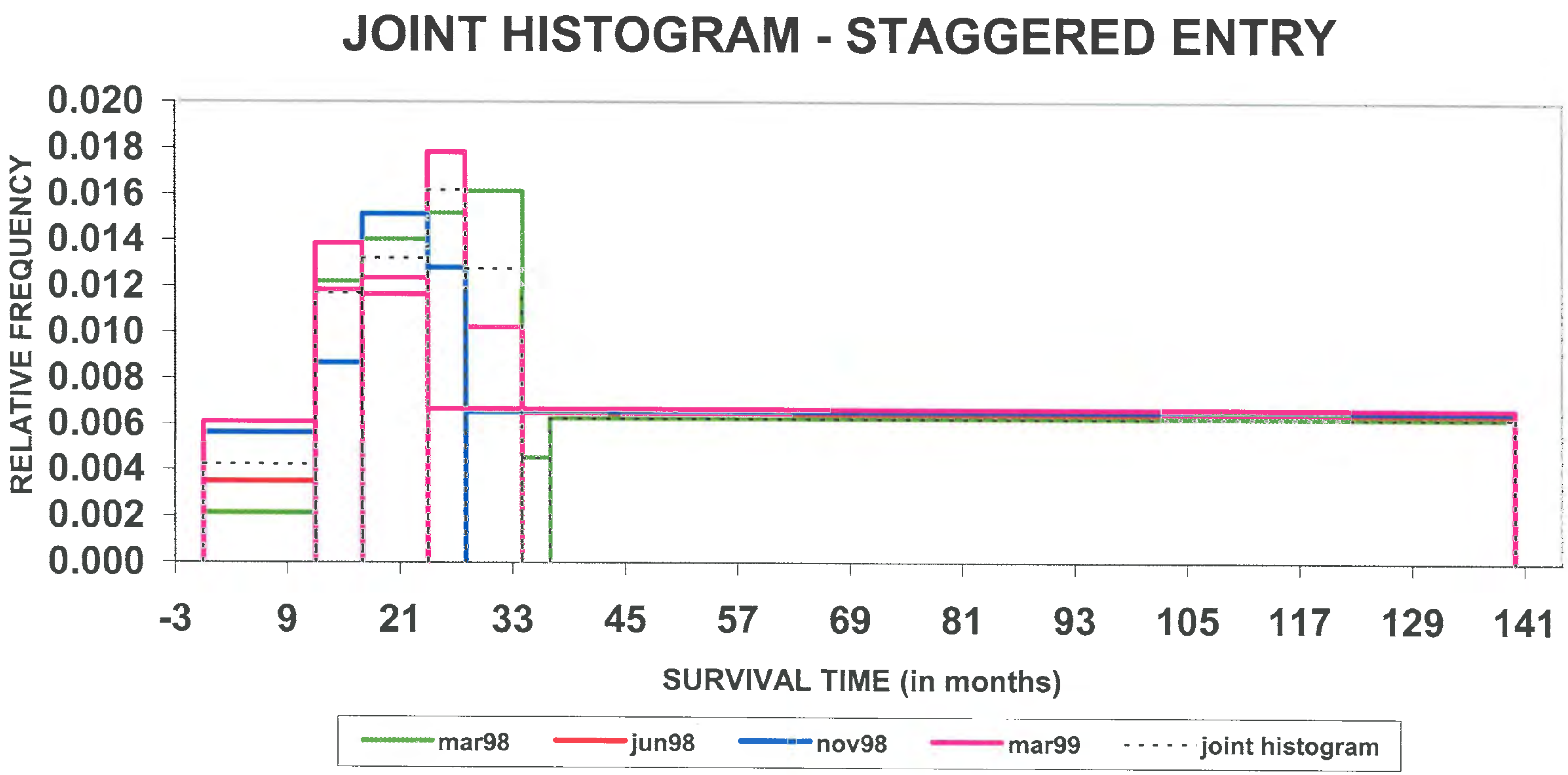

Figure 5.2: **Histograms of the four relative frequency distributions**

5.2 Fitting of a Lifetime Distribution

5.2.1 A fixed censoring time

Consider the June 1998 group of policies as an example of a data set with a pre-assigned fixed censoring time at 15 April 2001. All these 2809 policies start at the same time. The censored lifetimes are grouped into the open interval $[34;\infty)$, while the observed lifetimes are grouped into intervals $[0;12),[12;17),[17;24),[24;28)$, and $[28;34)$. Note that
$\boldsymbol{x} = (12,17,24,28,34)'$ is a 5×1 vector of upper class boundaries,
$\boldsymbol{f} = (118,166,229,200,172,1924)'$ is a 6×1 vector of frequencies and
$\boldsymbol{p} = (0.0420078, 0.0590958, 0.0815237, 0.0711997, 0.0612318, 0.6849413)'$ is a 6×1 vector of relative frequencies. This vector is used as starting value for the iterative process.

One survival model is fitted to this grouped data set under constraints imposed by the Weibull/log-logistic/lognormal distribution. The double iterative procedure, as described in section 3.3.2 of chapter three, is illustrated by the following computer output when a Weibull model is fitted to the grouped data.

```
I J        P         M | I J        P         M | I J        P         M | I J        P         M | I J        P         M
                       |                        |                        |                        |
1 1 0.0550538 0.0420078 | 2 1 0.0504223 0.0562540 | 3 1 0.0505730 0.0507667 | 4 1 0.0505685 0.0509455 | 5 1 0.0505686 0.0509401
    0.0489497 0.0590958 |     0.0465755 0.0466379 |     0.0466658 0.0456550 |     0.0466631 0.0457035 |     0.0466632 0.0457021
    0.0801283 0.0815237 |     0.0815332 0.0803011 |     0.0815189 0.0815789 |     0.0815195 0.0815707 |     0.0815195 0.0815712
    0.0504270 0.0711997 |     0.0524268 0.0513481 |     0.0523893 0.0534019 |     0.0523906 0.0533565 |     0.0523905 0.0533581
    0.0811157 0.0612318 |     0.0861566 0.0812418 |     0.0860254 0.0857363 |     0.0860296 0.0856221 |     0.0860295 0.0856258
    0.6843255 0.6849413 |     0.6828856 0.6842171 |     0.6828276 0.6828612 |     0.6828286 0.6828017 |     0.6828286 0.6828027
                       |                        |                        |                        |
1 2 0.0562478 0.0420078 | 2 2 0.0507666 0.0562540 | 3 2 0.0509453 0.0507667 | 4 2 0.0509400 0.0509455 | 5 2 0.0509401 0.0509401
    0.0466462 0.0590958 |     0.0456548 0.0466379 |     0.0457034 0.0456550 |     0.0457021 0.0457035 |     0.0457021 0.0457021
    0.0803032 0.0815237 |     0.0815797 0.0803011 |     0.0815715 0.0815789 |     0.0815720 0.0815707 |     0.0815719 0.0815712
    0.0513481 0.0711997 |     0.0534018 0.0513481 |     0.0533565 0.0534019 |     0.0533580 0.0533565 |     0.0533580 0.0533581
    0.0812371 0.0612318 |     0.0857358 0.0812418 |     0.0856215 0.0857363 |     0.0856252 0.0856221 |     0.0856251 0.0856258
    0.6842177 0.6849413 |     0.6828613 0.6842171 |     0.6828017 0.6828612 |     0.6828028 0.6828017 |     0.6828027 0.6828027
                       |                        |                        |                        |
1 3 0.0562540 0.0420078 | 2 3 0.0507667 0.0562540 | 3 3 0.0509455 0.0507667 | 4 3 0.0509401 0.0509455 | 5 3 0.0509403 0.0509401
    0.0466379 0.0590958 |     0.0456550 0.0466379 |     0.0457035 0.0456550 |     0.0457021 0.0457035 |     0.0457022 0.0457021
    0.0803011 0.0815237 |     0.0815789 0.0803011 |     0.0815707 0.0815789 |     0.0815712 0.0815707 |     0.0815712 0.0815712
    0.0513481 0.0711997 |     0.0534019 0.0513481 |     0.0533565 0.0534019 |     0.0533581 0.0533565 |     0.0533580 0.0533581
    0.0812418 0.0612318 |     0.0857363 0.0812418 |     0.0856221 0.0857363 |     0.0856258 0.0856221 |     0.0856257 0.0856258
    0.6842171 0.6849413 |     0.6828612 0.6842171 |     0.6828017 0.6828612 |     0.6828027 0.6828017 |     0.6828027 0.6828027
                       |                                                                           |
1 4 0.0562540 0.0420078 |                                                                           |
    0.0466379 0.0590958 |                                                                           |      MLE for M_c
    0.0803011 0.0815237 |                                                                           |
    0.0513481 0.0711997 |                                                                           |       0.0509403
    0.0812418 0.0612318 |                                                                           |       0.0457022
    0.6842171 0.6849413 |                                                                           |       0.0815712
                                                                                                    |       0.053358
                                                                                                    |       0.0856257
                                                                                                    |       0.6828027
```

The estimated parameters of the fitted survival models and the Wald test with the discrepancy values are reported in Table 5.2.

Table 5.2: **Maximum likelihood estimation subject to constraints: a fixed censoring time**

	Survival model		
	Weibull	Log-logistic	Lognormal
Maximum likelihood estimates	$\ln\widehat{\lambda}$=-7.693382	$\ln\widehat{\lambda}$=-8.243037	$\widehat{\mu}$ =3.8910773
	$\widehat{\alpha}$=1.9084456	$\widehat{\alpha}$=1.9084456	$\widehat{\sigma}$ =0.8319341
Wald test	51.5	39.8	25.0
Discrepancy	0.0183	0.0142	0.0089

Figure 5.3 shows the histogram of the relative frequency distribution and the fitted survival distributions. It is clear that the lognormal and log-logistic models fit very well.

5.2.2 Staggered entry

Introduction

Consider Figure 5.2, the four histograms of the relative frequency distributions. Maximum likelihood estimates are to be found in the following ways.

1. One survival model is fitted to the four histograms under constraints imposed by the Weibull/log-logistic/lognormal distribution.
2. Four survival models (Weibull/log-logistic/lognormal models), one for each entry time, are fitted under constraints imposed by the Weibull/log-logistic/lognormal distribution and under **further constraints** that
   - λ_i's are equal and α_i's are equal when fitting a Weibull or log-logistic
   - μ_i's are equal and σ_i's are equal when fitting a lognormal.
3. A joint histogram is fitted to the four histograms of the four relative frequency distributions under constraints imposed by the experimental design.

The constraints imposed by the experimental design

Consider Figure 5.4, illustrating the constraints imposed by the experimental design.

- $\pi_{1,j} = \pi_{2,j} = \pi_{3,j} = \pi_{4,j} \quad j = 1, 2, ..., 3$

- $$\begin{aligned} \pi_{1,7} + \pi_{1,6} + \pi_{1,5} + \pi_{1,4} &= \pi_{2,6} + \pi_{2,5} + \pi_{2,4} \\ &= \pi_{3,5} + \pi_{3,4} \\ &= \pi_{4,4} \end{aligned}$$

- $$\begin{aligned} \pi_{1,5} &= \pi_{2,5} \\ \pi_{1,4} &= \pi_{2,4} \end{aligned}$$

where $\pi_{i,j}$ = probability of an observation from sample i will fall in the j^{th} interval
= interval probability of j^{th} interval from sample $i \quad i = 1, 2, 3, 4 \quad j = 1, 2, ..., 7$

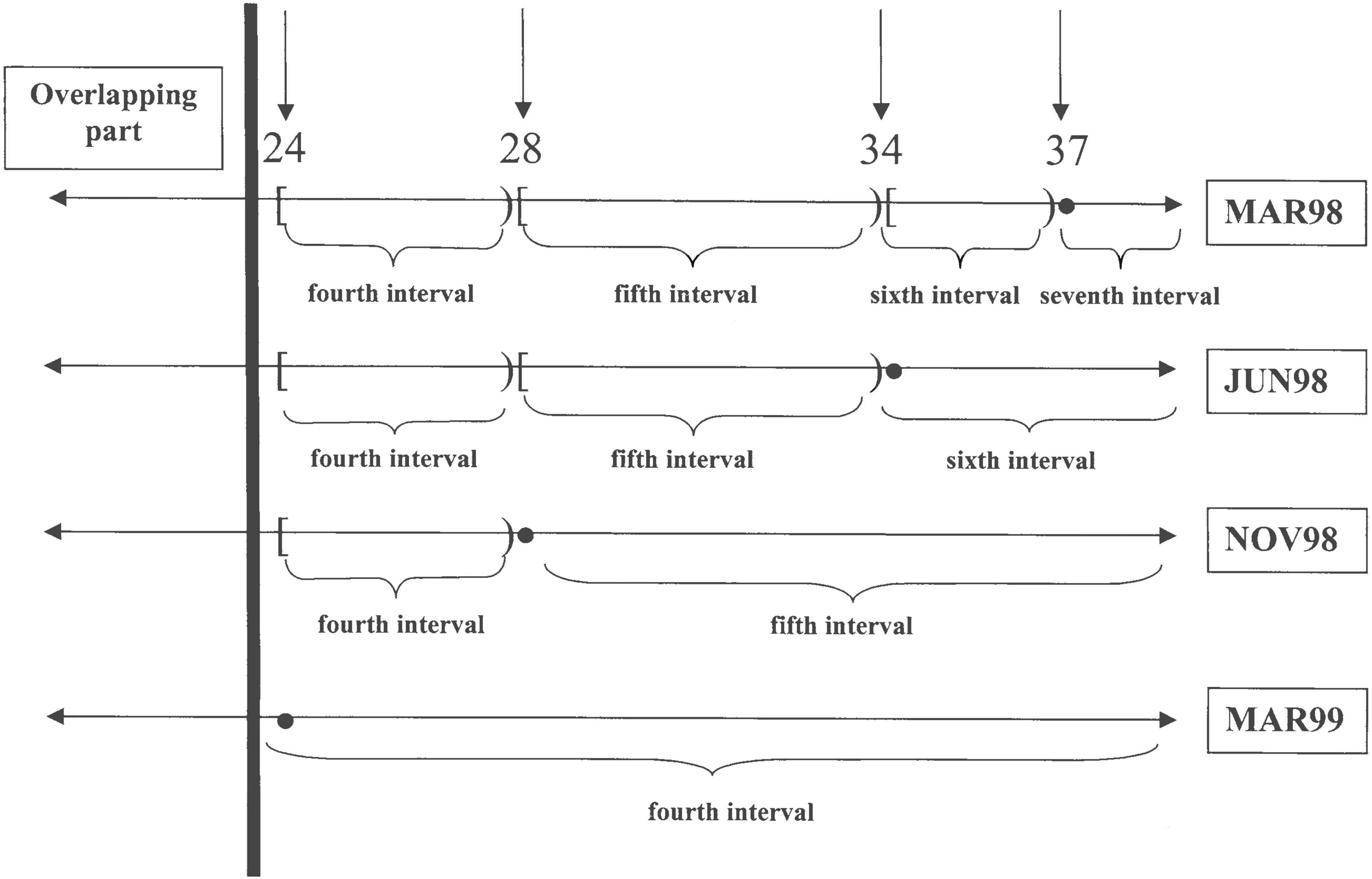

Figure 5.4: **Constraints imposed by the experimental design**

Fitting of one survival model to the four histograms

One survival model is fitted to the four histograms under constraints imposed by the Weibull/log-logistic/lognormal distribution. The estimated parameters and the Wald test with the discrepancy values are reported in Table 5.3.

Table 5.3: **Fitting of one survival model to the four histograms**

	Survival model		
	Weibull	Log-logistic	Lognormal
Maximum likelihood estimates	$\ln\widehat{\lambda}$=-7.39252	$\ln\widehat{\lambda}$=-7.959399	$\widehat{\mu}$ =3.8727296
	$\widehat{\alpha}$=1.8434286	$\widehat{\alpha}$=2.0647366	$\widehat{\sigma}$ =0.8636028
Wald test	302.5	253.6	254.9
Discrepancy	0.0300	0.0252	0.0253

The invariance property of the maximum likelihood estimator provides that the MLE of $\widehat{\ln\lambda}$ can be written as $\ln\widehat{\lambda}$.

Fitting of four survival models

Four survival models are fitted, one for each entry time, under constraints imposed by the Weibull/log-logistic/lognormal distribution and under further constraints that the parameters are equal.

The estimated parameters and the Wald test with the discrepancy values are reported in Table 5.4.

Table 5.4: **Fitting of one survival model to the four histograms**

Maximum likelihood estimates	Survival model: Weibull	Log-logistic	Lognormal
March 1998	$\ln\widehat{\lambda}$=-8.230773	$\ln\widehat{\lambda}$=-8.960949	$\widehat{\mu}$ =3.8234358
	$\widehat{\alpha}$=2.0570424	$\widehat{\alpha}$=2.3273887	$\widehat{\sigma}$ =0.7219206
June 1998	$\ln\widehat{\lambda}$=-7.693383	$\ln\widehat{\lambda}$=-8.243037	$\widehat{\mu}$ =3.8910773
	$\widehat{\alpha}$=1.9084457	$\widehat{\alpha}$=2.1214022	$\widehat{\sigma}$ =0.8319341
Nov 1998	$\ln\widehat{\lambda}$=-7.172834	$\ln\widehat{\lambda}$=-7.582113	$\widehat{\mu}$ =3.9182342
	$\widehat{\alpha}$=1.8026532	$\widehat{\alpha}$=1.9727851	$\widehat{\sigma}$ =0.9624843
March 1999	$\ln\widehat{\lambda}$=-6.781666	$\ln\widehat{\lambda}$=-7.113033	$\widehat{\mu}$ =3.9521501
	$\widehat{\alpha}$=1.7103598	$\widehat{\alpha}$=1.8569722	$\widehat{\sigma}$ =1.0417936
Over all four entry times	$\ln\widehat{\lambda}$=-7.39252	$\ln\widehat{\lambda}$=-7.959399	$\widehat{\mu}$ =3.8727296
	$\widehat{\alpha}$=1.8434286	$\widehat{\alpha}$=2.0647366	$\widehat{\sigma}$ =0.8636028
Wald test	302.5	253.6	254.9
Discrepancy	0.0300	0.0252	0.0253

A joint histogram to the four histograms is needed to make a graphical representation of the fitted models.

Fitting of a joint histogram to the four histograms

A joint histogram is fitted to the four histograms under constraints imposed by the experimental design. Table 5.5 gives the fitted joint relative frequencies to the four sets of relative frequencies of the samples. A graphical representation of the joint histogram over the four histograms appears in Figure 5.5.

Figure 5.6 shows the fitted joint histogram and the fitted survival distributions. The log-normal and log-logistic models again fit the data very well.

Table 5.5: **Fitted joint relative frequency distribution to the four samples**

Interval	Lifetime Intervals				Relative Frequency Vector				Fitted Joint
number	March 98	June 98	Nov 98	March 99	p_1	p_2	p_3	p_4	Relative Frequencies
first	$[0, 12)$	$[0, 12)$	$[0, 12)$	$[0, 12)$	0.025522	0.042008	0.067367	0.073038	0.050908
second	$[12, 17)$	$[12, 17)$	$[12, 17)$	$[12, 17)$	0.061098	0.059096	0.043307	0.069282	0.058450
third	$[17, 24)$	$[17, 24)$	$[17, 24)$	$[17, 24)$	0.098221	0.081524	0.105862	0.086394	0.092488
fourth	$[24, 28)$	$[24, 28)$	$[24, 28)$	$[24, \infty)$	0.060712	0.071200	0.051181	0.771286	0.064701
fifth	$[28, 34)$	$[28, 34)$	$[28, \infty)$		0.096674	0.061232	0.732284		0.076481
sixth	$[34, 37)$	$[34, \infty)$			0.013534	0.684941			0.013518
seventh	$[37, \infty)$				0.644238				0.643455

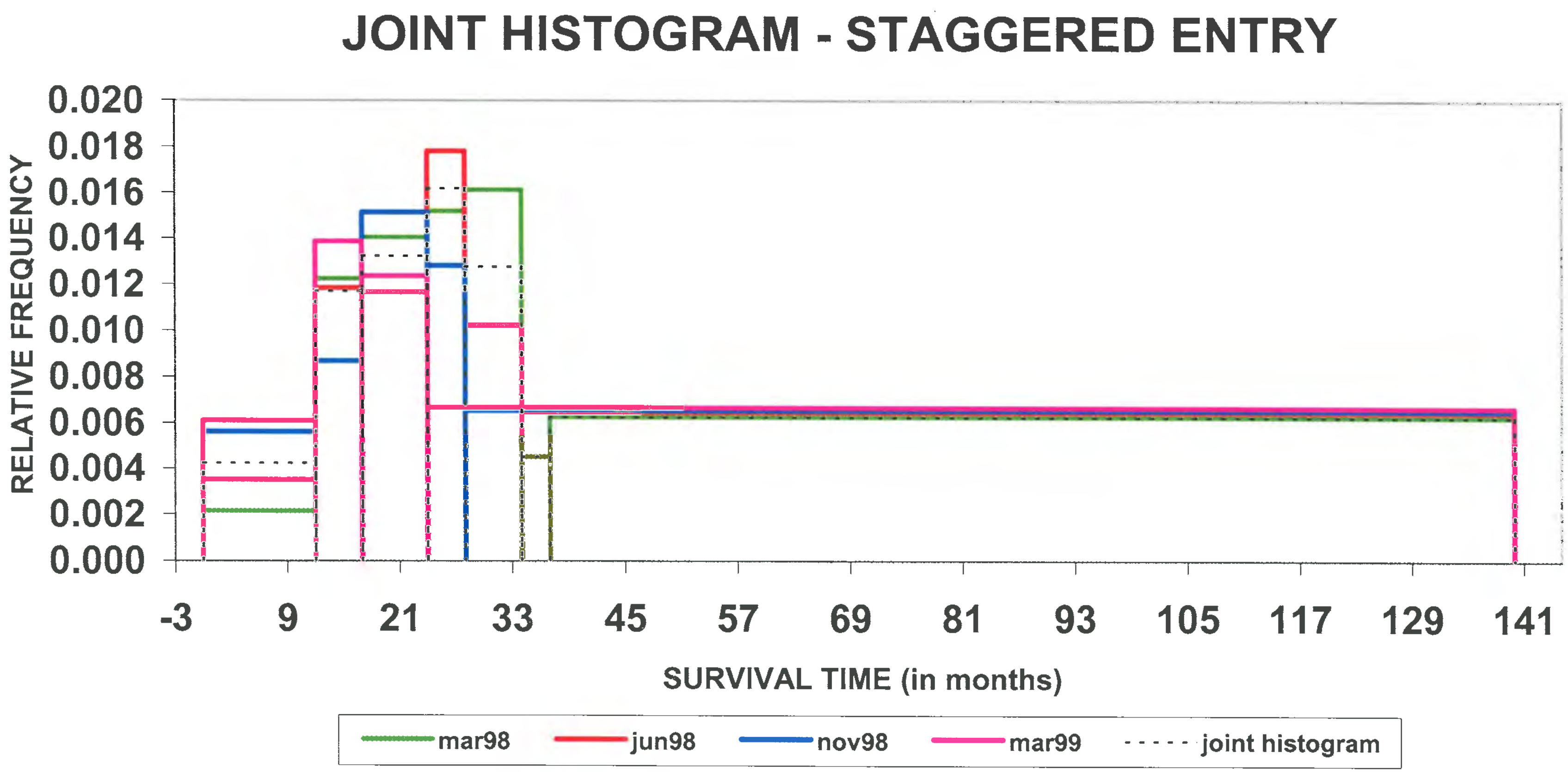

Figure 5.5: **Joint histogram over the four histograms**

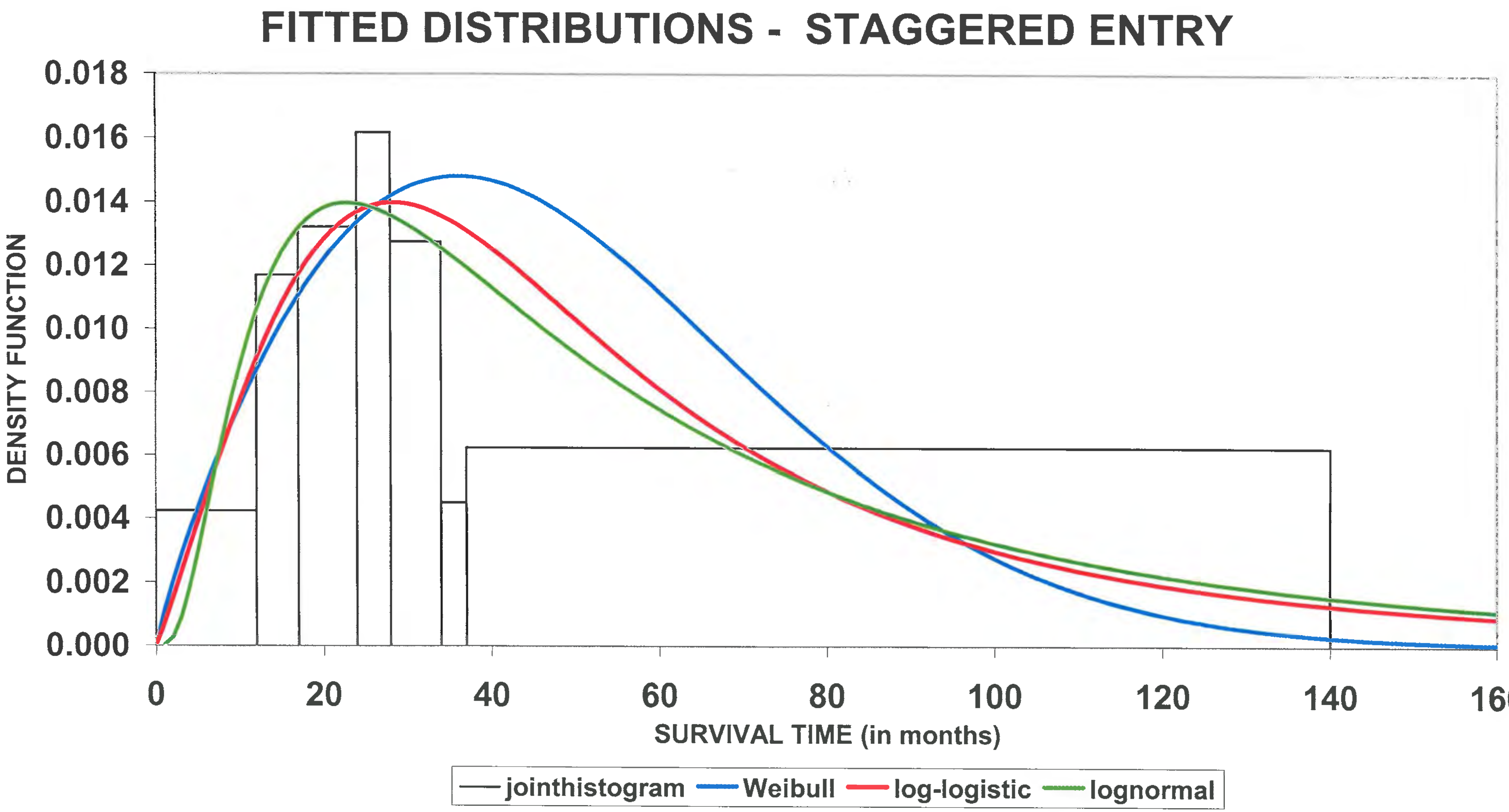

Figure 5.6: **Joint histogram and fitted survival distributions**

Estimated survivor and hazard functions and percentiles

Once the parameters of the Weibull and log-logistic survival distributions have been estimated, estimated hazard rates and survivor functions and the odds of a lapse can be calculated for time t. Percentiles of these survival distributions can also be estimated.

The formulaes and examples of calculations of the estimated hazard rates, survivor functions, odds of a lapse and percentiles of the Weibull and log-logistic survival distributions are given on the next page.

The survival curves and the graphs of the hazard rates of the fitted Weibull and log-logistic models are shown respectively in Figure 5.7 and Figure 5.8. From Figure 5.7 is it clear that the two survivor functions are equal for $t-$values up to 40 months, and then the probability for a policy to survive longer than time t with $t>$40 becomes larger for the log-logistic fitting than for the Weibull fitting. Note in Figure 5.8 the increasing trend of the Weibull hazard rates as t increases.

Survival Model

WEIBULL

Estimated hazard function

$$\boxed{\widehat{h}(t) = \widehat{\lambda} \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}$$

$$\widehat{h}(t) = e^{-7.39252} \cdot 1.8434286 \cdot t^{1.8434286-1}$$

$$\begin{aligned} \widehat{h}(12) &= 0.0092323 \\ \widehat{h}(24) &= 0.0165655 \end{aligned}$$

Estimated survivor function

$$\boxed{\widehat{S}(t) = \exp(-\widehat{\lambda} \cdot t^{\widehat{\alpha}})}$$

$$\widehat{S}(t) = \exp(-e^{-7.39252} \cdot t^{1.8434286})$$

$$\begin{aligned} \widehat{S}(12) &= 0.9416719 \\ \widehat{S}(24) &= 0.806001 \end{aligned}$$

Estimated odds of a lapse

$$\boxed{\widehat{odds}(t) = \frac{1-\widehat{S}(t)}{\widehat{S}(t)} = \exp(\widehat{\lambda} \cdot t^{\widehat{\alpha}-1})}$$

$$\widehat{odds}(t) = \frac{1-\widehat{S}(t)}{\widehat{S}(t)} = \exp(e^{-7.39252} \cdot t^{1.8434286-1})$$

$$\begin{aligned} \widehat{odds}(12) &= 0.061941 \\ \widehat{odds}(24) &= 0.240693 \end{aligned}$$

Estimated percentiles

$$\boxed{\widehat{t_p} = (\frac{1}{\widehat{\lambda}} \cdot \ln \frac{100}{100-p})^{\frac{1}{\widehat{\alpha}}}}$$

$$\widehat{t_{50}} = (\frac{1}{e^{-7.39252}} \cdot \ln \frac{100}{100-50})^{\frac{1}{1.8434286}} = 45.21$$

LOG-LOGISTIC

Estimated hazard function

$$\boxed{\widehat{h}(t) = \frac{\widehat{\lambda} \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}{(1+\widehat{\lambda} \cdot t^{\widehat{\alpha}})}}$$

$$\widehat{h}(t) = \frac{e^{-7.959399} \cdot 2.0647366 \cdot t^{2.0647366-1}}{(1+e^{-7.959399} \cdot t^{2.0647366})}$$

$$\begin{aligned} \widehat{h}(12) &= 0.0095996 \\ \widehat{h}(24) &= 0.0170517 \end{aligned}$$

Estimated survivor function

$$\boxed{\widehat{S}(t) = \frac{1}{1+\widehat{\lambda} \cdot t^{\widehat{\alpha}}}}$$

$$\widehat{S}(t) = \frac{1}{1+e^{-7.959399} \cdot t^{2.0647366}}$$

$$\begin{aligned} \widehat{S}(12) &= 0.9442083 \\ \widehat{S}(24) &= 0.8017956 \end{aligned}$$

Estimated odds of a lapse

$$\boxed{\widehat{odds}(t) = \frac{1-\widehat{S}(t)}{\widehat{S}(t)} = \widehat{\lambda} \cdot t^{\widehat{\alpha}}}$$

$$\widehat{odds}(t) = e^{-7.959399} \cdot t^{2.0647366}$$

$$\begin{aligned} \widehat{odds_0}(12) &= 0.0590884 \\ \widehat{odds_0}(24) &= 0.2472006 \end{aligned}$$

Estimated percentiles

$$\boxed{\widehat{t_p} = (\frac{1}{\widehat{\lambda}} \cdot \frac{p}{100-p})^{\frac{1}{\widehat{\alpha}}}}$$

$$\widehat{t_{50}} = (\frac{1}{e^{-7.959399}} \cdot \frac{50}{100-50})^{\frac{1}{2.0647366}} = 47.22$$

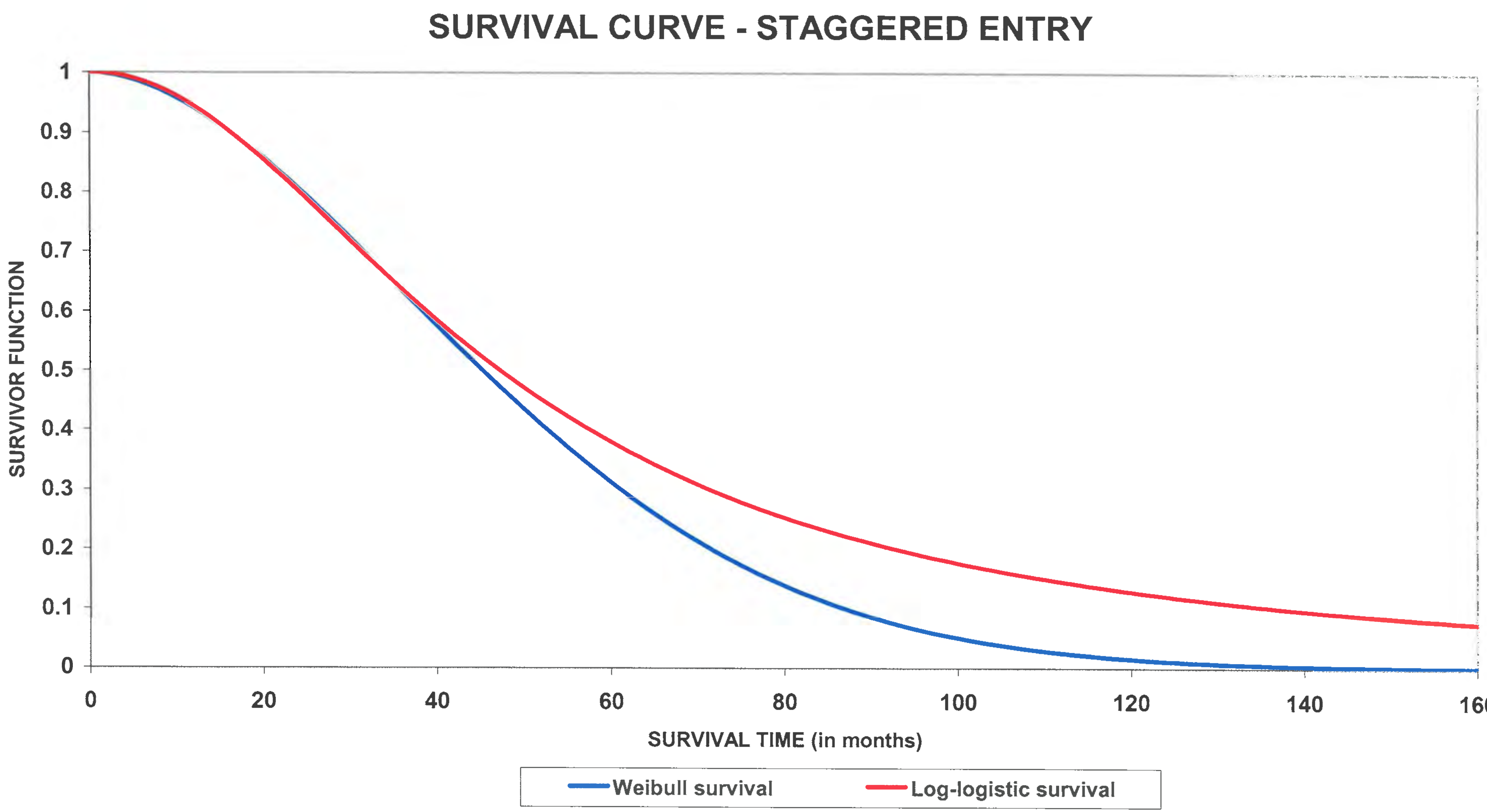

Figure 5.7: **Survival curves of fitted Weibull and log-logistic models**

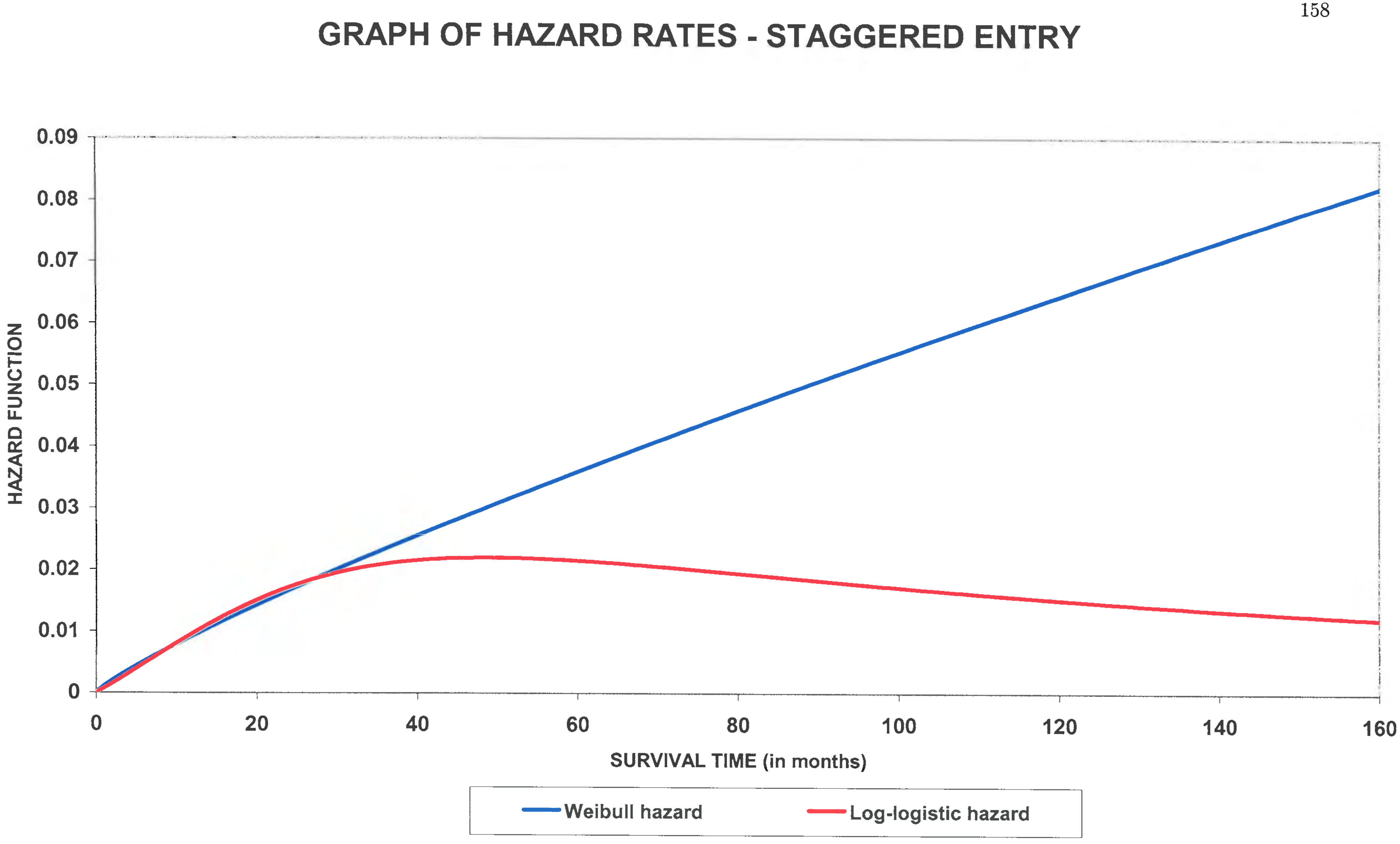

Figure 5.8: **Graphs of hazard rates of fitted Weibull and log-logistic models**

The estimated percentiles of the Weibull and log-logistic survival distributions are reported in Table 5.6.

Table 5.6: **Percentiles estimated from Weibull and log-logistic regression models**

	Survival model	
Percentile	Weibull	Log-logistic
P5	11.01	11.34
P10	16.27	16.29
P20	24.45	24.13
P25	28.06	27.74
P30	31.53	31.33
P40	38.31	38.80
P50	45.21	47.22
P60	52.60	57.47
P70	61.00	71.18
P75	65.85	80.40
P80	71.40	92.42
P90	86.71	136.88
P95	100.02	196.56

At the Weibull model, the median time to a lapse of a policy is estimated as 45.21 months and the odds of a lapse at 45.21 months is 1, that means $P(T >$45.21 months)$=P(T <$45.21 months). At the log-logistic model, the median time to a lapse of a policy is estimated as 47.22 months and the odds of a lapse at 47.22 months is 1, that means $P(T >$47.22 months)$=P(T <$47.22 months).

It is evident from the estimates of the percentile lifetimes that 20% of the policies will not lapse within 71 months under a Weibull model (see Weibull's P80), while 30% of the policies will not lapse within 71 months under a log-logistic model (see log-logistic's P70).

Note again the equal percentile estimates for the two distributions up to 40 months, confirming the pattern that was detected in the survival curves in Figure 5.7.

5.3 Fitting of Parametric Regression Models

5.3.1 Introduction

A survival model is fitted for **each level of a risk factor** or **combination of levels of risk factors** by using maximum likelihood estimation of parameters subject to constraints.

The fitting of regression models is illustrated where the effect of the risk factors (covariates) is to alter the scale parameter λ, while the shape parameter α remains constant. Applications are also done where both parameters alter.

The fitting of log-logistic regression models and Weibull regression models will be discussed only for **staggered entry of policies**.

5.3.2 A survival model for each level of a risk factor

Consider one risk factor AGE on three levels [18;35), [35;45) and [45+) years. The 10077 observations are distributed in the three age groups as follows: 3644 in age group [18;35), 3425 in age group [35;45) and 3008 in age group [45+). A regression model is fitted to the grouped survival data where each policy has information on the entry period as well as the age level. The grouped lifetimes of the policies with staggered entry as well as the concomitant information on AGE are given in Table 5.7.

The combined frequency vector $\boldsymbol{f}$ is defined as
$\boldsymbol{f}' = (\boldsymbol{f}'_{11}, \boldsymbol{f}'_{21}, \boldsymbol{f}'_{31}, \boldsymbol{f}'_{41}, \boldsymbol{f}'_{12}, \boldsymbol{f}'_{22}, \boldsymbol{f}'_{32}, \boldsymbol{f}'_{42}, \boldsymbol{f}'_{13}, \boldsymbol{f}'_{23}, \boldsymbol{f}'_{33}, \boldsymbol{f}'_{43})$
$\boldsymbol{f}_{il}$ is the frequency vector for the i^{th} entry group and the l^{th} AGE level,
$i = 1, 2, 3, 4$ and $l = 1, 2, 3$.

$\boldsymbol{f}_{11} = (29, 59, 95, 73, 108, 15, 642)'$
$\boldsymbol{f}_{12} = (21, 50, 91, 45, 75, 13, 553)'$
$\boldsymbol{f}_{13} = (16, 49, 68, 39, 67, 7, 471)'$

$\boldsymbol{f}_{21} = (41, 75, 103, 92, 83, 628)'$
$\boldsymbol{f}_{22} = (49, 62, 61, 66, 54, 753)'$

Table 5.7: **Multi-dimensional frequency table of grouped data set with one risk factor**

Entry	Age	Lifetime intervals						
March 98		$[0,12)$	$[12,17)$	$[17,24)$	$[24,28)$	$[28,34)$	$[34,37)$	$[37,\infty)$
	[18;35)	29	59	95	73	108	15	642
	[35;45)	21	50	91	45	75	13	553
	[45+)	16	49	68	39	67	7	471
June 98		$[0,12)$	$[12,17)$	$[17,24)$	$[24,28)$	$[28,34)$	$[34,\infty)$	
	[18;35)	41	75	103	92	83	628	
	[35;45)	49	62	61	66	54	753	
	[45+)	28	29	65	42	35	543	
Nov 98		$[0,12)$	$[12,17)$	$[17,24)$	$[24,28)$	$[28,\infty)$		
	[18;35)	68	34	99	57	570		
	[35;45)	40	44	83	33	533		
	[45+)	46	21	60	27	571		
March 99		$[0,12)$	$[12,17)$	$[17,24)$	$[24,\infty)$			
	[18;35)	71	60	69	573			
	[35;45)	54	61	68	616			
	[45+)	50	45	70	659			

$\boldsymbol{f}_{23} = (28, 29, 65, 42, 35, 543)'$

$\boldsymbol{f}_{31} = (68, 34, 99, 57, 570)'$
$\boldsymbol{f}_{32} = (40, 44, 83, 33, 533)'$
$\boldsymbol{f}_{33} = (46, 21, 60, 27, 571)'$

$\boldsymbol{f}_{41} = (71, 60, 69, 573)'$
$\boldsymbol{f}_{42} = (54, 61, 68, 616)'$
$\boldsymbol{f}_{43} = (50, 45, 70, 659)'$

The vectors $\boldsymbol{x}_i \quad i = 1, 2, 3, 4$ of upper class boundaries for the i^{th} entry group are

$$\boldsymbol{x}_1 = \begin{pmatrix} 12 \\ 17 \\ 24 \\ 28 \\ 34 \\ 37 \end{pmatrix} \quad \boldsymbol{x}_2 = \begin{pmatrix} 12 \\ 17 \\ 24 \\ 28 \\ 34 \end{pmatrix} \quad \boldsymbol{x}_3 = \begin{pmatrix} 12 \\ 17 \\ 24 \\ 28 \end{pmatrix} \quad \text{and} \quad \boldsymbol{x}_4 = \begin{pmatrix} 12 \\ 17 \\ 24 \end{pmatrix}.$$

From the estimated regression parameters, survival model parameters can be found for each level of this risk factor as well as for the baseline distribution.

The shape parameter remains constant

The estimated regression coefficients of the regression model where the effect of the risk factor AGE is to alter the scale parameter λ, while the shape parameter α remains constant, are reported in Table 5.8.

Table 5.8: **Fitting a regression model (constant shape) to grouped data with one risk factor**

Effect	Maximum likelihood estimates	Regression model Log-logistic	Weibull
Baseline mean	$\widehat{\ln \lambda_0} = \ln \widehat{\lambda}_0$	-7.981750	-7.404312
Age [18;35)	$\widehat{\beta}_{A_1}$	0.180958	0.159090
Age [35;45)	$\widehat{\beta}_{A_2}$	-0.034975	-0.033957
Age [45+)	$\widehat{\beta}_{A_3}$	-0.145983	-0.125133
Constant shape	$\widehat{\alpha}$	2.066384	1.8423341

The estimated lambda parameters of the three survival distributions for the three AGE levels then are

$$\begin{aligned}
\widehat{\lambda}_{A_1} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_1}) \\
\widehat{\lambda}_{A_2} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_2}) \\
\widehat{\lambda}_{A_3} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_3}).
\end{aligned}$$

with the same estimated alpha parameter $\widehat{\alpha}$. These parameters are summarized for each AGE level in Table 5.9.

Table 5.9: **Parameters of a survival model (constant shape) for each level of risk factor AGE**

AGE level	Maximum likelihood estimates	Survival model Log-logistic	Weibull
Age [18;35)	$\ln \widehat{\lambda}_{A_1}$	-7.800792	-7.245223
	$\widehat{\alpha}$	2.066384	1.842334
Age [35;45)	$\ln \widehat{\lambda}_{A_2}$	-8.016725	-7.438269
	$\widehat{\alpha}$	2.066384	1.842334
Age [45+)	$\ln \widehat{\lambda}_{A_3}$	-8.127733	-7.529445
	$\widehat{\alpha}$	2.066384	1.842334
Baseline	$\ln \widehat{\lambda}_0$	-7.981750	-7.404312
	$\widehat{\alpha}$	2.066384	1.842334

The shape parameter alters

The fitting of regression models is illustrated where the effect of the risk factors (covariates) is to alter both the scale parameter λ and the shape parameter α. The estimated regression coefficients of this regression model are reported in Table 5.10.

Table 5.10: **Fitting a regression model (shape alters) to grouped data with one risk factor**

Effect	Maximum likelihood estimates	**Regression model** Log-logistic	Weibull
Baseline mean	$\widehat{\ln \lambda_0} = \ln \widehat{\lambda}_0$	-7.943357	-7.381423
Age [18;35)	$\widehat{\beta}_{A_1}$	-0.196012	-0.075175
	$\widehat{\alpha}_{A_1}$	2.168064	1.904217
Age [35;45)	$\widehat{\beta}_{A_2}$	0.156976	0.119892
	$\widehat{\alpha}_{A_2}$	1.9974967	1.790610
Age [45+)	$\widehat{\beta}_{A_3}$	0.039035	-0.044717
	$\widehat{\alpha}_{A_3}$	1.9995073	1.811986

The weighted mean of the $\widehat{\alpha}_{A_i}$'s $\quad i = 1, 2, 3$ is used as an estimate for the shape parameter of the baseline distribution.

The estimated lambda parameters of the three survival distributions for the three AGE levels are calculated from Table 5.10 as

$$
\begin{aligned}
\widehat{\lambda}_{A_1} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_1}) \\
\widehat{\lambda}_{A_2} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_2}) \\
\widehat{\lambda}_{A_3} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_3}).
\end{aligned}
$$

Each age group survival distribution has its own estimated alpha parameter.

These parameters for each AGE level are summarized in Table 5.11, together with the Wald test and discrepancy value to compare the fitted survival distributions at each AGE level.

Table 5.11: **Parameters of a survival model (shape alters) for each level of risk factor AGE**

AGE level	Maximum likelihood estimates	**Survival model** Log-logistic	Weibull
Age [18;35)	$\ln \widehat{\lambda}_{A_1}$	-8.139369	-7.456598
	$\widehat{\alpha}_{A_1}$	2.168064	1.904217
Wald test		128.5	144.2
Discrepancy		0.0353	0.0396
Age [35;45)	$\ln \widehat{\lambda}_{A_2}$	-7.786381	-7.261531
	$\widehat{\alpha}_{A_2}$	1.997497	1.790610
Wald test		93.1	108.3
Discrepancy		0.0271	0.0316
Age [45+)	$\ln \widehat{\lambda}_{A_3}$	-7.904321	-7.426139
	$\widehat{\alpha}_{A_3}$	1.999507	1.811986
Wald test		95.5	109.5
Discrepancy		0.0317	0.0364
Baseline	$\ln \widehat{\lambda}_0$	-7.943357	-7.381423
	$\widehat{\alpha}_0$	2.0597767	1.8380729

A joint histogram to the data of each AGE level **over the four entry groups** is needed to make a graphical representation of the fitted models for each AGE level. Table 5.12 gives the three sets of fitted joint frequencies for the three AGE levels. This fitting was done by maximum likelihood estimation subject to constraints imposed by the experimental design. The Wald test and discrepancy value measure the goodness-of-fit.

Table 5.12: **Fitted joint frequency distributions for the three AGE levels**

Interval number	Interval of survival times	Fitted Joint Frequencies Age [18;35) years	Age [35;45) years	Age [45+) years
first	$[0, 12)$	209	164	140
second	$[12, 17)$	228	217	144
third	$[17, 24)$	366	303	263
fourth	$[24, 28)$	285.65814	195.15779	165.56561
fifth	$[28, 34)$	330.67093	226.805	208.49003
sixth	$[34, 37)$	50.791574	53.264105	30.561947
seventh	$[37, \infty)$	2173.8794	2265.7731	2056.3824
Wald		70.63	53.92	46.62
Discrepancy		0.0194	0.0157	0.0155

Figure 5.9 shows the fitted joint histogram and the fitted survival distributions for age group [18;35).

Figure 5.10 shows the fitted joint histogram and the fitted survival distributions for age group [35;45).

Figure 5.11 shows the fitted joint histogram and the fitted survival distributions for age group [45+).

In all the cases the survival models fit very well, with the log-logistic model slightly better than the Weibull model, as indicated by the discrepancy values in Table 5.11.

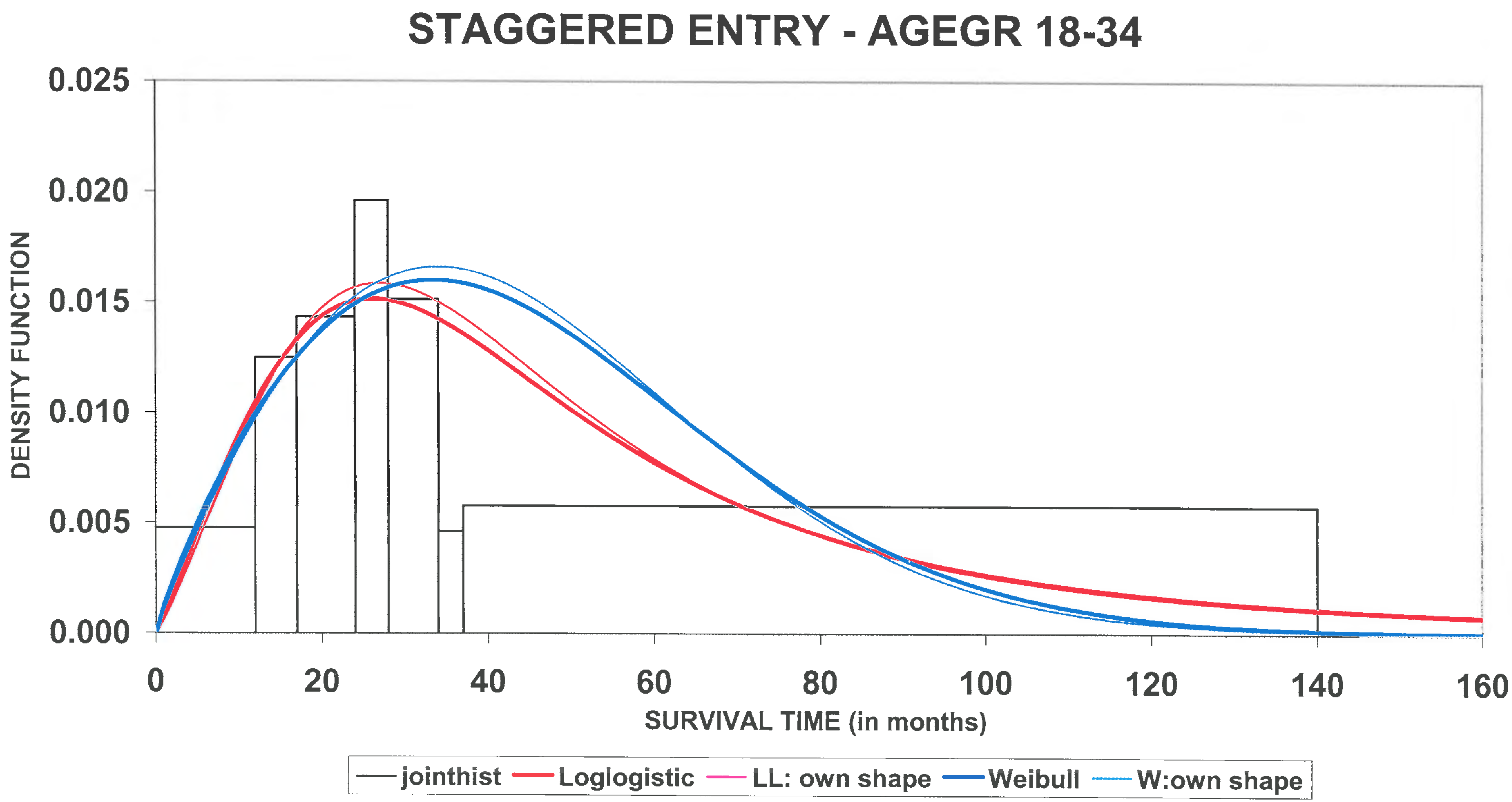

Figure 5.9: **Joint histogram and fitted survival distributions for age group [18;35)**

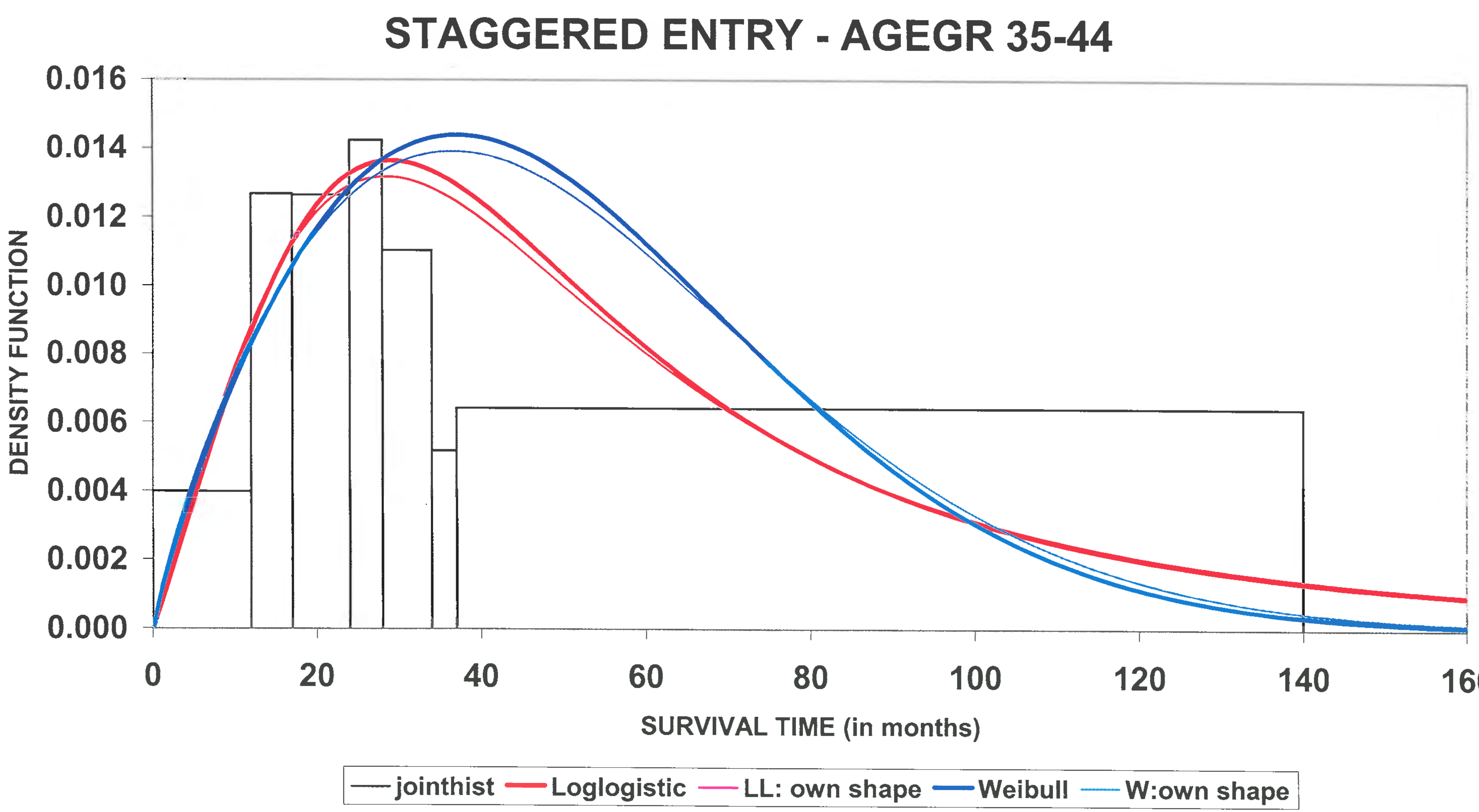

Figure 5.10: **Joint histogram and fitted survival distributions for age group [35;45)**

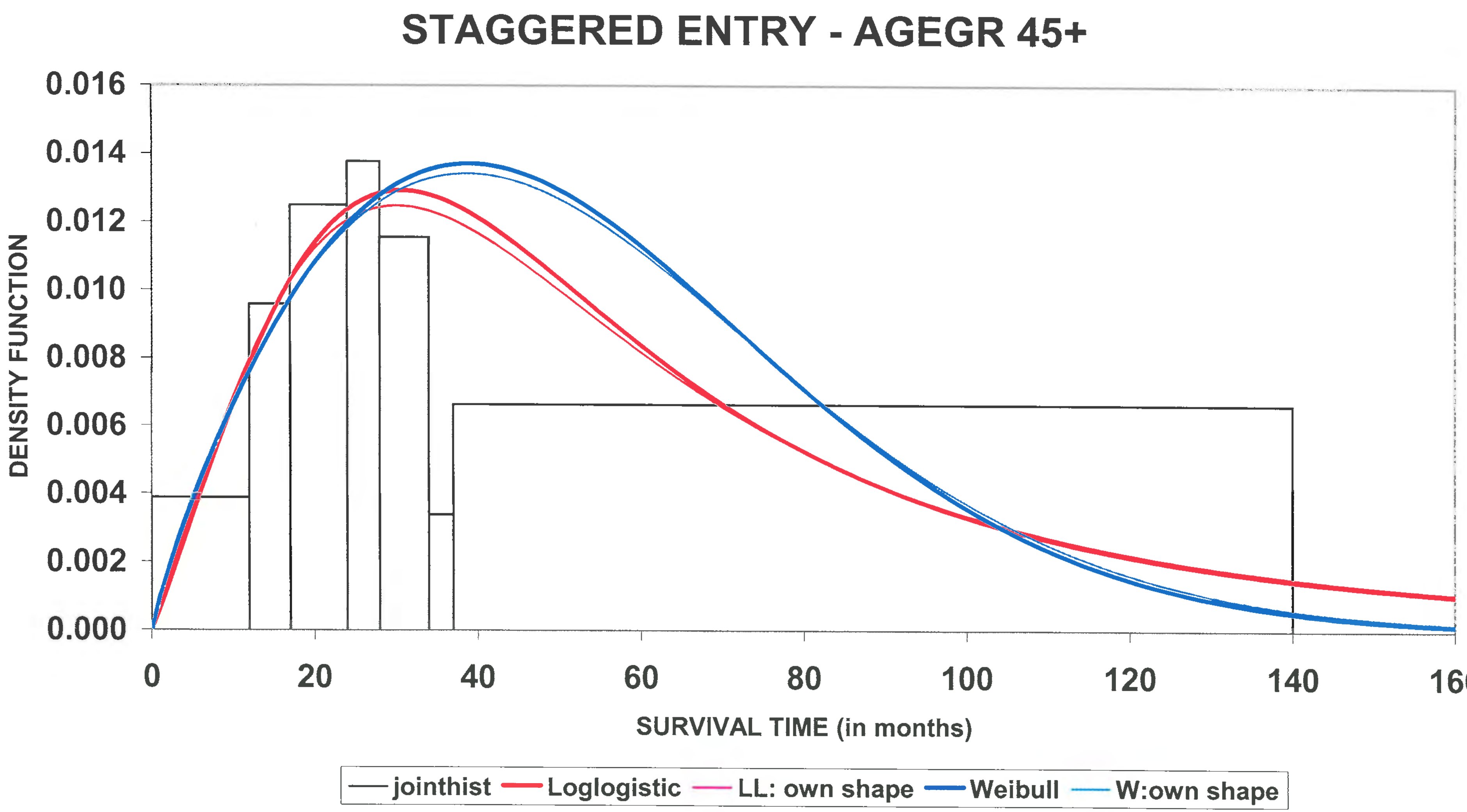

Figure 5.11: **Joint histogram and fitted survival distributions for age group [45+)**

5.3.3 Deriving of indices and risk scores from log-logistic regression model

Once the parameters of the log-logistic baseline distribution and log-logistic age group distributions have been estimated, estimated hazard and survivor functions, odds of a lapse, odds ratios and hazard ratios at time t can be calculated.

The odds ratio for age group [18;35) is the relative odds of a lapse at time t of a policy, with the age of the policyholder in [18;35), compared to a policy with the baseline characteristics. The odds ratios for the three age groups result in a set of indices, showing the effect of each age group on the baseline odds of a lapse at time t.

The hazard ratio for age group [18;35) is the relative hazard rate of a lapse at time t of a policy, with the age of the policyholder in [18;35), compared to a policy with the baseline characteristics. The hazard ratios for the three age groups result in a set of risk scores, showing the effect of each age group on the baseline hazard rate of a lapse at time t.

Percentiles of the four log-logistic survival distributions can also be estimated.

The calculations of estimated hazard and survivor functions, odds of a lapse, odds ratios and hazard ratios are illustrated on the following five pages.

The survival curves and the graphs of the hazard rates of the fitted log-logistic age group models are shown in Figure 5.12 and Figure 5.13 with measure of comparison the baseline curves. These survival curves can be described as graphs of the covariate-adjusted survivor functions, and the other graphs as graphs of the covariate-adjusted hazard rates.

The effects of the agegroups on the baseline distribution are clearly depicted in these figures. It is evident from these two figures that the policyholders in the age group 45+ have the lowest risk for their policies to lapse. Note the the survival curve of this age group lies above the baseline survival curve in Figure 5.12, while the curve of the hazard rates for this age group lies the furthest distance beneath the baseline curve of hazard rates. Similarly age group [18;35) has the highest risk for their policies to lapse.

Estimated Hazard Function at log-logistic regression model

Shape remains constant

$$\boxed{\widehat{h}_0(t) = \frac{\widehat{\lambda}_0 \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}{(1+\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}})}}$$

$$\widehat{h}_0(t) = \frac{e^{-7.98175} \cdot 2.0663843 \cdot t^{2.0663843-1}}{(1+e^{-7.98175} \cdot t^{2.0663843})}$$

$$\widehat{h}_0(12) = 0.009443$$
$$\widehat{h}_0(24) = 0.0168323$$

$$\boxed{\widehat{h}_{A_1}(t) = \frac{\widehat{\lambda}_{A_1} \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}{(1+\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}})}}$$

$$\widehat{h}_{A_1}(t) = \frac{e^{-7.800792} \cdot 2.0663843 \cdot t^{2.0663843-1}}{(1+e^{-7.800792} \cdot t^{2.0663843})}$$

$$\widehat{h}_{A_1}(12) = 0.0111944$$
$$\widehat{h}_{A_1}(24) = 0.0194182$$

$$\boxed{\widehat{h}_{A_2}(t) = \frac{\widehat{\lambda}_{A_2} \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}{(1+\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}})}}$$

$$\widehat{h}_{A_2}(t) = \frac{e^{-8.016725} \cdot 2.0663843 \cdot t^{2.0663843-1}}{(1+e^{-8.016725} \cdot t^{2.0663843})}$$

$$\widehat{h}_{A_2}(12) = 0.0091356$$
$$\widehat{h}_{A_2}(24) = 0.0163637$$

$$\boxed{\widehat{h}_{A_3}(t) = \frac{\widehat{\lambda}_{A_3} \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}{(1+\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}})}}$$

$$\widehat{h}_{A_3}(t) = \frac{e^{-8.127733} \cdot 2.0663843 \cdot t^{2.0663843-1}}{(1+e^{-8.127733} \cdot t^{2.0663843})}$$

$$\widehat{h}_{A_3}(12) = 0.0082216$$
$$\widehat{h}_{A_3}(24) = 0.0149428$$

Shape alters

$$\boxed{\widehat{h}_0(t) = \frac{\widehat{\lambda}_0 \cdot \widehat{\alpha}_0 \cdot t^{\widehat{\alpha}_0-1}}{(1+\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0})}}$$

$$\widehat{h}_0(t) = \frac{e^{-7.943357} \cdot 2.0597767 \cdot t^{2.0597767-1}}{(1+e^{-7.943357} \cdot t^{2.0597767})}$$

$$\widehat{h}_0(12) = 0.0096102$$
$$\widehat{h}_0(24) = 0.0170145$$

$$\boxed{\widehat{h}_{A_1}(t) = \frac{\widehat{\lambda}_{A_1} \cdot \widehat{\alpha}_{A_1} \cdot t^{\widehat{\alpha}_{A_1}-1}}{(1+\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}_{A_1}})}}$$

$$\widehat{h}_{A_1}(t) = \frac{e^{-8.139369} \cdot 2.1680640 \cdot t^{2.1680640-1}}{(1+e^{-8.139369} \cdot t^{2.1680640})}$$

$$\widehat{h}_{A_1}(12) = 0.0108363$$
$$\widehat{h}_{A_1}(24) = 0.0201312$$

$$\boxed{\widehat{h}_{A_2}(t) = \frac{\widehat{\lambda}_{A_2} \cdot \widehat{\alpha}_{A_2} \cdot t^{\widehat{\alpha}_{A_2}-1}}{(1+\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}_{A_2}})}}$$

$$\widehat{h}_{A_2}(t) = \frac{e^{-7.786381} \cdot 1.9974967 \cdot t^{1.9974967-1}}{(1+e^{-7.786381} \cdot t^{1.9974967})}$$

$$\widehat{h}_{A_2}(12) = 0.0093391$$
$$\widehat{h}_{A_2}(24) = 0.015965$$

$$\boxed{\widehat{h}_{A_3}(t) = \frac{\widehat{\lambda}_{A_3} \cdot \widehat{\alpha}_{A_3} \cdot t^{\widehat{\alpha}_{A_3}-1}}{(1+\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}_{A_3}})}}$$

$$\widehat{h}_{A_3}(t) = \frac{e^{-7.904321} \cdot 1.9995073 \cdot t^{1.9995073-1}}{(1+e^{-7.904321} \cdot t^{1.9995073})}$$

$$\widehat{h}_{A_3}(12) = 0.0084005$$
$$\widehat{h}_{A_3}(24) = 0.0145896$$

Estimated Survivor Function at log-logistic regression model

Shape remains constant

$$\boxed{\widehat{S}_0(t) = \frac{1}{1+\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}}}}$$

$$\widehat{S}_0(t) = \frac{1}{1+e^{-7.981750} \cdot t^{2.066384}}$$

$$\begin{aligned} \widehat{S}_0(12) &= 0.9451622 \\ \widehat{S}_0(24) &= 0.8045013 \end{aligned}$$

$$\boxed{\widehat{S}_{A_1}(t) = \frac{1}{1+\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}}}}$$

$$\widehat{S}_{A_1}(t) = \frac{1}{1+e^{-7.800792} \cdot t^{2.066384}}$$

$$\begin{aligned} \widehat{S}_{A_1}(12) &= 0.9349915 \\ \widehat{S}_{A_1}(24) &= 0.7744674 \end{aligned}$$

$$\boxed{\widehat{S}_{A_2}(t) = \frac{1}{1+\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}}}}$$

$$\widehat{S}_{A_2}(t) = \frac{1}{1+e^{-8.016725} \cdot t^{2.066384}}$$

$$\begin{aligned} \widehat{S}_{A_2}(12) &= 0.946947 \\ \widehat{S}_{A_2}(24) &= 0.809944 \end{aligned}$$

$$\boxed{\widehat{S}_{A_3}(t) = \frac{1}{1+\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}}}}$$

$$\widehat{S}_{A_3}(t) = \frac{1}{1+e^{-8.127733} \cdot t^{2.066384}}$$

$$\begin{aligned} \widehat{S}_{A_3}(12) &= 0.9522551 \\ \widehat{S}_{A_3}(24) &= 0.8264469 \end{aligned}$$

Shape alters

$$\boxed{\widehat{S}_0(t) = \frac{1}{1+\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0}}}$$

$$\widehat{S}_0(t) = \frac{1}{1+e^{-7.943357} \cdot t^{2.0597767}}$$

$$\begin{aligned} \widehat{S}_0(12) &= 0.9440121 \\ \widehat{S}_0(24) &= 0.8017512 \end{aligned}$$

$$\boxed{\widehat{S}_{A_1}(t) = \frac{1}{1+\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}_{A_1}}}}$$

$$\widehat{S}_{A_1}(t) = \frac{1}{1+e^{-8.139369} \cdot t^{2.168064}}$$

$$\begin{aligned} \widehat{S}_{A_1}(12) &= 0.9400224 \\ \widehat{S}_{A_1}(24) &= 0.7771517 \end{aligned}$$

$$\boxed{\widehat{S}_{A_2}(t) = \frac{1}{1+\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}_{A_2}}}}$$

$$\widehat{S}_{A_2}(t) = \frac{1}{1+e^{-7.786381} \cdot t^{1.997497}}$$

$$\begin{aligned} \widehat{S}_{A_2}(12) &= 0.9438949 \\ \widehat{S}_{A_2}(24) &= 0.8081802 \end{aligned}$$

$$\boxed{\widehat{S}_{A_3}(t) = \frac{1}{1+\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}_{A_3}}}}$$

$$\widehat{S}_{A_3}(t) = \frac{1}{1+e^{-7.904321} \cdot t^{1.999507}}$$

$$\begin{aligned} \widehat{S}_{A_3}(12) &= 0.9495848 \\ \widehat{S}_{A_3}(24) &= 0.8248819 \end{aligned}$$

Estimated Odds at log-logistic regression model

Shape remains constant

$$\boxed{\widehat{odds}_0(t) = \frac{1-\widehat{S}_0(t)}{\widehat{S}_0(t)} = \widehat{\lambda}_0 \cdot t^{\widehat{\alpha}}}$$

$$\widehat{odds}_0(t) = e^{-7.981750} \cdot t^{2.066384}$$

$$\begin{aligned} \widehat{odds}_0(12) &= 0.0580194 \\ \widehat{odds}_0(24) &= 0.243006 \end{aligned}$$

$$\boxed{\widehat{odds}_{A_1}(t) = \frac{1-\widehat{S}_{A_1}(t)}{\widehat{S}_{A_1}(t)} = \widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}}}$$

$$\widehat{odds}_{A_1}(t) = e^{-7.800792} \cdot t^{2.066384}$$

$$\begin{aligned} \widehat{odds}_{A_1}(12) &= 0.0695284 \\ \widehat{odds}_{A_1}(24) &= 0.2912099 \end{aligned}$$

$$\boxed{\widehat{odds}_{A_2}(t) = \frac{1-\widehat{S}_{A_2}(t)}{\widehat{S}_{A_2}(t)} = \widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}}}$$

$$\widehat{odds}_{A_2}(t) = e^{-8.016725} \cdot t^{2.066384}$$

$$\begin{aligned} \widehat{odds}_{A_2}(12) &= 0.056025 \\ \widehat{odds}_{A_2}(24) &= 0.234654 \end{aligned}$$

$$\boxed{\widehat{odds}_{A_3}(t) = \frac{1-\widehat{S}_{A_3}(t)}{\widehat{S}_{A_3}(t)} = \widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}}}$$

$$\widehat{odds}_{A_3}(t) = e^{-8.127733} \cdot t^{2.066384}$$

$$\begin{aligned} \widehat{odds}_{A_3}(12) &= 0.0501388 \\ \widehat{odds}_{A_3}(24) &= 0.2099991 \end{aligned}$$

Shape alters

$$\boxed{\widehat{odds}_0(t) = \frac{1-\widehat{S}_0(t)}{\widehat{S}_0(t)} = \widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0}}$$

$$\widehat{odds}_0(t) = e^{-7.943357} \cdot t^{2.059777}$$

$$\begin{aligned} \widehat{odds}_0(12) &= 0.0593084 \\ \widehat{odds}_0(24) &= 0.2472697 \end{aligned}$$

$$\boxed{\widehat{odds}_{A_1}(t) = \frac{1-\widehat{S}_{A_1}(t)}{\widehat{S}_{A_1}(t)} = \widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}_{A_1}}}$$

$$\widehat{odds}_{A_1}(t) = e^{-8.139369} \cdot t^{2.168064}$$

$$\begin{aligned} \widehat{odds}_{A_1}(12) &= 0.0638045 \\ \widehat{odds}_{A_1}(24) &= 0.2867500 \end{aligned}$$

$$\boxed{\widehat{odds}_{A_2}(t) = \frac{1-\widehat{S}_{A_2}(t)}{\widehat{S}_{A_2}(t)} = \widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}_{A_2}}}$$

$$\widehat{odds}_{A_2}(t) = e^{-7.786381} \cdot t^{1.997497}$$

$$\begin{aligned} \widehat{odds}_{A_2}(12) &= 0.05944 \\ \widehat{odds}_{A_2}(24) &= 0.237348 \end{aligned}$$

$$\boxed{\widehat{odds}_{A_3}(t) = \frac{1-\widehat{S}_{A_3}(t)}{\widehat{S}_{A_3}(t)} = \widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}_{A_3}}}$$

$$\widehat{odds}_{A_3}(t) = e^{-7.904321} \cdot t^{1.999507}$$

$$\begin{aligned} \widehat{odds}_{A_3}(12) &= 0.0530918 \\ \widehat{odds}_{A_3}(24) &= 0.2122948 \end{aligned}$$

Estimated Odds Ratio at log-logistic regression model

Shape remains constant

$$\boxed{\widehat{oddsratio}_{A_1}(t) = \frac{\widehat{odds}_{A_1}(t)}{\widehat{odds}_0(t)} = \frac{\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}}}{\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}}}}$$

$$\begin{aligned} \widehat{oddsratio}_{A_1}(12) &= \frac{0.0695284}{0.0580194} \\ &= 1.198365 \\ \widehat{oddsratio}_{A_1}(24) &= \frac{0.2912099}{0.243006} \\ &= 1.198365 \end{aligned}$$

$$\boxed{\widehat{oddsratio}_{A_2}(t) = \frac{\widehat{odds}_{A_2}(t)}{\widehat{odds}_0(t)} = \frac{\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}}}{\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}}}}$$

$$\begin{aligned} \widehat{oddsratio}_{A_2}(12) &= \frac{0.0560253}{0.0580194} \\ &= 0.965629 \\ \widehat{oddsratio}_{A_2}(24) &= \frac{0.2346538}{0.243006} \\ &= 0.965629 \end{aligned}$$

$$\boxed{\widehat{oddsratio}_{A_3}(t) = \frac{\widehat{odds}_{A_3}(t)}{\widehat{odds}_0(t)} = \frac{\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}}}{\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}}}}$$

$$\begin{aligned} \widehat{oddsratio}_{A_3}(12) &= \frac{0.0501388}{0.0580194} \\ &= 0.864172 \\ \widehat{oddsratio}_{A_3}(24) &= \frac{0.2099991}{0.243006} \\ &= 0.864172 \end{aligned}$$

Odds ratio of an age group is constant over time

Odds ratio is called an index

One set of indices, irrespective of time

Shape alters

$$\boxed{\widehat{oddsratio}_{A_1}(t) = \frac{\widehat{odds}_{A_1}(t)}{\widehat{odds}_0(t)} = \frac{\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}_{A_1}}}{\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0}}}$$

$$\begin{aligned} \widehat{oddsratio}_{A_1}(12) &= \frac{0.0638045}{0.0593084} \\ &= 1.075808 \\ \widehat{oddsratio}_{A_1}(24) &= \frac{0.286750}{0.2472697} \\ &= 1.159665 \end{aligned}$$

$$\boxed{\widehat{oddsratio}_{A_2}(t) = \frac{\widehat{odds}_{A_2}(t)}{\widehat{odds}_0(t)} = \frac{\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}_{A_2}}}{\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0}}}$$

$$\begin{aligned} \widehat{oddsratio}_{A_2}(12) &= \frac{0.05944}{0.0593084} \\ &= 1.002219 \\ \widehat{oddsratio}_{A_2}(24) &= \frac{0.237348}{0.2472697} \\ &= 0.959874 \end{aligned}$$

$$\boxed{\widehat{oddsratio}_{A_3}(t) = \frac{\widehat{odds}_{A_3}(t)}{\widehat{odds}_0(t)} = \frac{\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}_{A_3}}}{\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0}}}$$

$$\begin{aligned} \widehat{oddsratio}_{A_3}(12) &= \frac{0.0530918}{0.0593084} \\ &= 0.895182 \\ \widehat{oddsratio}_{A_3}(24) &= \frac{0.2122948}{0.2472697} \\ &= 0.8585556 \end{aligned}$$

Odds ratio of an age group depends on time

Odds ratio is called an index

Two sets of indices, one for t=12 and one for t=24

Estimated Hazard Ratio at log-logistic regression model

Shape remains constant

$$\widehat{hazardratio}_{A_1}(t) = \frac{\widehat{h}_{A_1}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_1}}{\widehat{\lambda}_0} \cdot \frac{(1+\widehat{\lambda}_0 t^{\widehat{\alpha}})}{(1+\widehat{\lambda}_{A_1} t^{\widehat{\alpha}})}$$

$$\begin{aligned}\widehat{hazardratio}_{A_1}(12) &= \tfrac{0.0111944}{0.009443} \\ &= 1.1854696 \\ \widehat{hazardratio}_{A_1}(24) &= \tfrac{0.0194182}{0.0168323} \\ &= 1.1536273\end{aligned}$$

$$\widehat{hazardratio}_{A_2}(t) = \frac{\widehat{h}_{A_2}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_2}}{\widehat{\lambda}_0} \cdot \frac{(1+\widehat{\lambda}_0 t^{\widehat{\alpha}})}{(1+\widehat{\lambda}_{A_2} t^{\widehat{\alpha}})}$$

$$\begin{aligned}\widehat{hazardratio}_{A_2}(12) &= \tfrac{0.0087902}{0.009443} \\ &= 0.967453 \\ \widehat{hazardratio}_{A_2}(24) &= \tfrac{0.0157604}{0.0168323} \\ &= 0.9721618\end{aligned}$$

$$\widehat{hazardratio}_{A_3}(t) = \frac{\widehat{h}_{A_3}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_3}}{\widehat{\lambda}_0} \cdot \frac{(1+\widehat{\lambda}_0 t^{\widehat{\alpha}})}{(1+\widehat{\lambda}_{A_3} t^{\widehat{\alpha}})}$$

$$\begin{aligned}\widehat{hazardratio}_{A_3}(12) &= \tfrac{0.0080242}{0.009443} \\ &= 0.8706574 \\ \widehat{hazardratio}_{A_3}(24) &= \tfrac{0.014387}{0.0168323} \\ &= 0.8877456\end{aligned}$$

Hazard ratio of an age group depends on time

Hazard ratio is called a risk score

Two sets of risk scores, one for t=12 and one for t=24

Shape alters

$$\widehat{hazardratio}_{A_1}(t) = \frac{\widehat{h}_{A_1}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_1}}{\widehat{\lambda}_0} \cdot \frac{(1+\widehat{\lambda}_0 t^{\widehat{\alpha}_0})}{(1+\widehat{\lambda}_{A_1} t^{\widehat{\alpha}_{A_1}})}$$

$$\begin{aligned}\widehat{hazardratio}_{A_1}(12) &= \tfrac{0.0108363}{0.0096102} \\ &= 1.1275801 \\ \widehat{hazardratio}_{A_1}(24) &= \tfrac{0.0201312}{0.0170145} \\ &= 1.1831797\end{aligned}$$

$$\widehat{hazardratio}_{A_2}(t) = \frac{\widehat{h}_{A_2}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_2}}{\widehat{\lambda}_0} \cdot \frac{(1+\widehat{\lambda}_0 t^{\widehat{\alpha}_0})}{(1+\widehat{\lambda}_{A_2} t^{\widehat{\alpha}_{A_2}})}$$

$$\begin{aligned}\widehat{hazardratio}_{A_2}(12) &= \tfrac{0.0093391}{0.0096102} \\ &= 0.9717948 \\ \widehat{hazardratio}_{A_2}(24) &= \tfrac{0.015965}{0.0170145} \\ &= 0.9383156\end{aligned}$$

$$\widehat{hazardratio}_{A_3}(t) = \frac{\widehat{h}_{A_3}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_3}}{\widehat{\lambda}_0} \cdot \frac{(1+\widehat{\lambda}_0 t^{\widehat{\alpha}_0})}{(1+\widehat{\lambda}_{A_3} t^{\widehat{\alpha}_{A_3}})}$$

$$\begin{aligned}\widehat{hazardratio}_{A_3}(12) &= \tfrac{0.0084005}{0.0096102} \\ &= 0.8741186 \\ \widehat{hazardratio}_{A_3}(24) &= \tfrac{0.0145896}{0.0170145} \\ &= 0.8574789\end{aligned}$$

Hazard ratio of an age group depends on time

Hazard ratio is called a risk score

Two sets of risk scores, one for t=12 and one for t=24

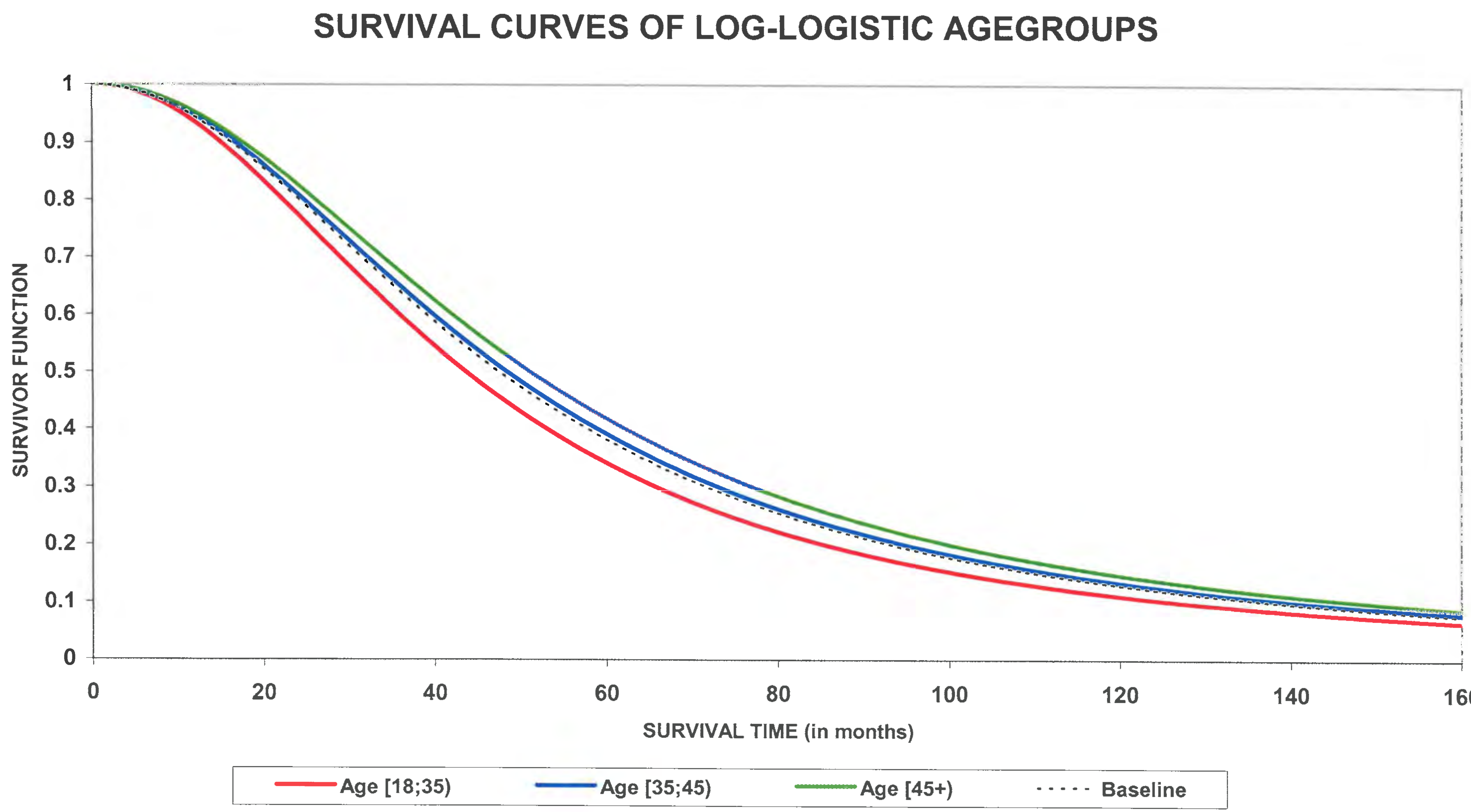

Figure 5.12: **Survival curves of fitted log-logistic age group models**

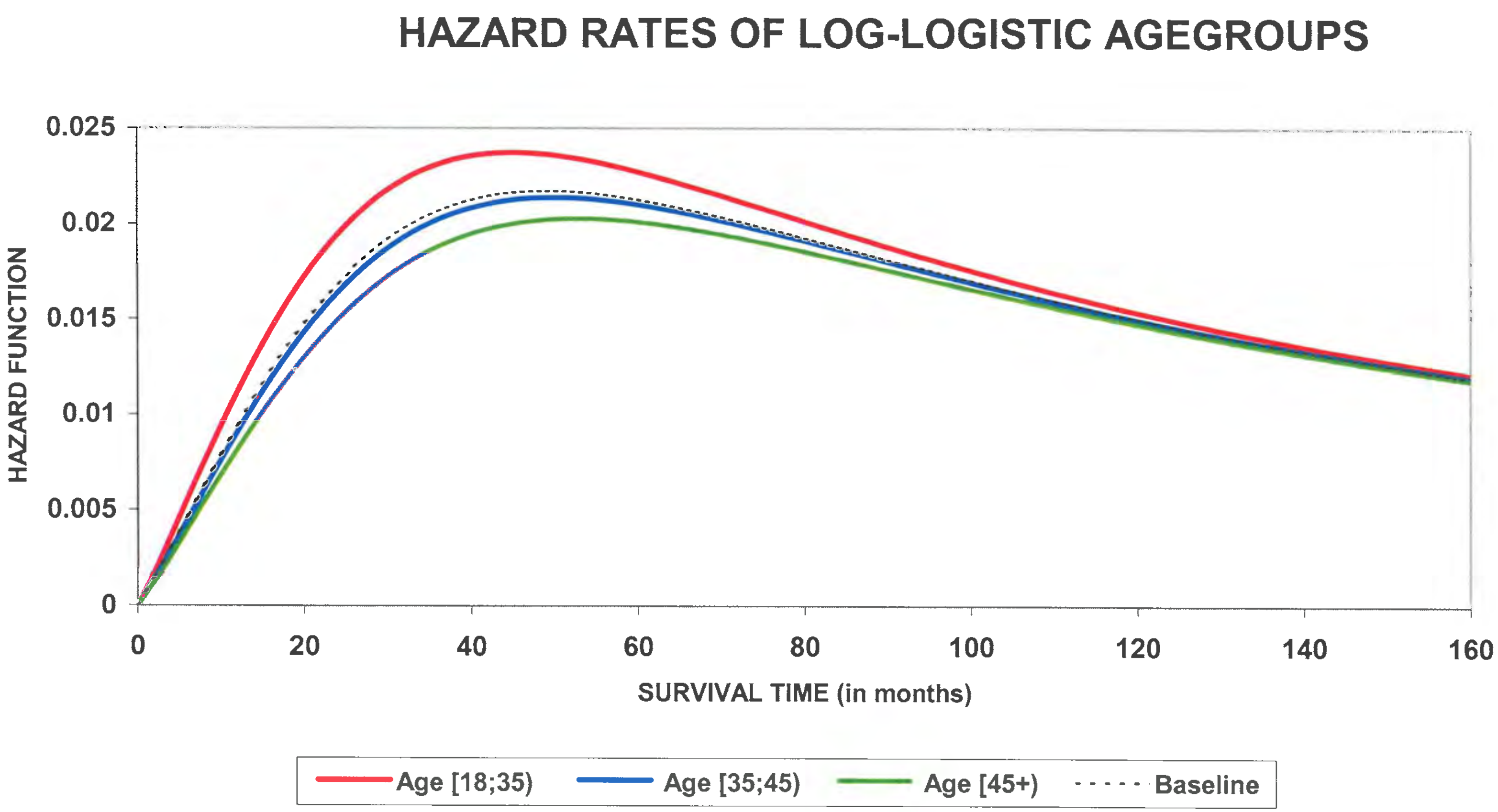

Figure 5.13: **Graphs of hazard rates of fitted log-logistic age group models**

For a **constant shape parameter** in the log-logistic distributions, the indices (estimated odds ratios) may be obtained also from the exponent of the $\widehat{\beta}$-values in the log-logistic regression model, for example

$$\begin{aligned} e^{\widehat{\beta}_{A_1}} &= e^{0.180958} = 1.198365 \\ e^{\widehat{\beta}_{A_2}} &= e^{-0.034975} = 0.965629 \\ e^{\widehat{\beta}_{A_3}} &= e^{-0.145983} = 0.864172. \end{aligned}$$

The indices of the three age groups, estimated from the log-logistic regression model, are compared to the indices, obtained from the logit model, in Table 5.13.

Table 5.13: **Comparison of indices: log-logistic regression model and logit model**

		Log-logistic regression model				Logit model	
		Shape remains constant		Shape alters			
Effect	n	Index		Index		Index	
		t=12	t=24	t=12	t=24	t=12	t=24
Baseline odds	10077	0.058019	0.243006	0.059308	0.247270	0.0543	0.2570
Age [18;35)	3644	1.198365	1.198365	1.075808	1.159665	1.1321	1.1370
Age [35;45)	3425	0.965629	0.965629	1.002219	0.959874	0.9679	0.9817
Age [45+)	3008	0.864172	0.864172	0.898556	0.858556	0.9126	0.8960

The index of age group [18;35) of 1.198365 shows the effect of this age group on the baseline odds of a lapse. This effect is multiplicative on the baseline odds of a lapse. Thus the effect of age group [18;35) is to increase the baseline odds of a lapse by a factor 1.198365 .

From Table 5.13 follows that one log-logistic regression model provides indices for any time value, while a new logit model has to be built for a fixed time value, say t=12 months, conditional on a restricted experimental design where all the policies must have an exposure of at least one year when investigating the lapses of policies in the first year. There is no such restrictions in the more general experimental design for the log-logistic regression model where all the policies can be used in the analysis, even those policies with inception dates very close to the cut-off point.

Predicted indices from the log-logistic regression model, for varying time values, are shown in Table 5.14 (constant shape) and in Table 5.15 (shape alters).

Predicted risk scores from the log-logistic regression model, for varying time values, are shown in Table 5.16 (constant shape) and in Table 5.17 (shape alters).

Table 5.13: **Predicted indices from log-logistic regression model (constant shape)**

	Predicted indices from log-logistic regression model									
Effect	t=6	t=12	t=18	t=24	t=30	t=36	t=42	t=48	t=54	t=60
Baseline odds	0.013853	0.058019	0.134105	0.243006	0.385363	0.561680	0.772373	1.017796	1.298259	1.614039
Age [18;35)	1.198365	1.198365	1.198365	1.198365	1.198365	1.198365	1.198365	1.198365	1.198365	1.198365
Age [35;45)	0.965629	0.965629	0.965629	0.965629	0.965629	0.965629	0.965629	0.965629	0.965629	0.965629
Age [45+)	0.864172	0.864172	0.864172	0.864172	0.864172	0.864172	0.864172	0.864172	0.864172	0.864172

Table 5.14: **Predicted indices from log-logistic regression model (shape alters)**

	Predicted indices from log-logistic regression model									
Effect	t=6	t=12	t=18	t=24	t=30	t=36	t=42	t=48	t=54	t=60
Baseline odds	0.014225	0.059308	0.136718	0.247270	0.391547	0.570006	0.783034	1.030921	1.313978	1.632444
Age [18;35)	0.998015	1.075808	1.124096	1.159665	1.188028	1.211716	1.232113	1.250058	1.266104	1.280632
Age [35;45)	1.046431	1.002219	0.977227	0.959874	0.946627	0.935939	0.926996	0.919319	0.912600	0.906631
Age [45+)	0.933371	0.898556	0.873571	0.858556	0.847086	0.837829	0.830081	0.823428	0.817603	0.812428

Table 5.15: **Predicted risk scores from log-logistic regression model (constant shape)**

Effect	Predicted risk scores from log-logistic regression model t=6	t=12	t=18	t=24	t=30	t=36	t=42	t=48	t=54	t=60
Baseline hazard rate	0.004706	0.009443	0.013575	0.016832	0.01916	0.020645	0.021440	0.021715	0.021616	0.021265
Age [18;35)	1.195126	1.185470	1.170900	1.153627	1.135699	1.118561	1.103015	1.089366	1.077614	1.067604
Age [35;45)	0.966083	0.967453	0.969570	0.972162	0.974951	0.977716	0.980313	0.982666	0.984749	0.986566
Age [45+)	0.865779	0.870657	0.878279	0.887746	0.898105	0.908557	0.918542	0.927734	0.935988	0.943282

Table 5.16: **Predicted risk scores from log-logistic regression model (shape alters)**

Effect	Predicted risk scores from log-logistic regression model t=6	t=12	t=18	t=24	t=30	t=36	t=42	t=48	t=54	t=60
Baseline hazard rate	0.004815	0.009610	0.013763	0.017014	0.019319	0.020773	0.021537	0.021783	0.021660	0.021289
Age [18;35)	1.050512	1.127580	1.165792	1.183180	1.187651	1.184381	1.176920	1.167574	1.157727	1.148149
Age [35;45)	1.014131	0.971795	0.950282	0.938316	0.932001	0.929252	0.928743	0.929593	0.932223	0.933254
Age [45+)	0.906908	0.874119	0.861105	0.857479	0.859272	0.864196	0.870770	0.878032	0.885382	0.892466

For a **constant shape parameter** in the log-logistic distributions, the pth percentile of the baseline lifetime distribution can be calculated from

$$\left(\frac{1}{\widehat{\lambda}_0}\cdot\frac{p}{100-p}\right)^{\frac{1}{\widehat{\alpha}}}$$

and that of the age group distributions from

$$\left(\frac{1}{\widehat{\lambda}_{A_i}}\cdot\frac{p}{100-p}\right)^{\frac{1}{\widehat{\alpha}}} \quad i=1,2,3.$$

Note that the latter is equal to the pth percentile of the baseline distribution multiplied by the specific index to the power $-\frac{1}{\widehat{\alpha}}$.

For a **shape parameter that alters**, the formulaes change to

$$\left(\frac{1}{\widehat{\lambda}_0}\cdot\frac{p}{100-p}\right)^{\frac{1}{\widehat{\alpha}_0}} \quad \text{and} \quad \left(\frac{1}{\widehat{\lambda}_{A_i}}\cdot\frac{p}{100-p}\right)^{\frac{1}{\widehat{\alpha}_{A_i}}} \quad i=1,2,3.$$

The estimated percentiles of the baseline and the age group log-logistic distributions for a constant shape parameter as well as for different shape parameters are reported in Table 5.18.

Table 5.18: **Lifetime percentiles estimated from log-logistic regression model**

	Log-logistic regression model							
	Lifetime distribution: constant shape				Lifetime distribution: shape alters			
Lifetime percentile	Baseline	Age group [18;35)	Age group [35;45)	Age group [45+)	Baseline	Age group [18;35)	Age group [35;45)	Age group [45+)
P5	11.45	10.49	11.64	12.28	11.32	10.98	11.29	11.95
P10	16.43	15.05	16.71	17.64	16.27	15.50	16.41	17.36
P20	24.33	22.29	24.75	26.11	24.13	22.53	24.63	26.04
P25	27.97	25.62	28.44	30.01	27.74	25.73	28.45	30.07
P30	31.58	28.93	32.12	33.89	31.34	28.89	32.26	34.10
P40	39.11	35.83	39.78	41.97	38.84	35.42	40.25	42.54
P50	47.59	43.60	48.40	51.07	47.29	42.70	49.31	52.10
P60	57.91	53.05	58.90	62.15	57.58	51.48	60.40	63.81
P70	71.71	65.70	72.94	76.96	71.36	63.12	75.36	79.59
P75	80.99	74.20	82.37	86.92	80.62	70.88	85.46	90.25
P80	93.09	85.28	94.68	99.90	92.71	80.93	98.70	104.21
P90	137.82	126.27	140.18	147.91	137.43	117.64	148.12	156.34
P95	197.87	181.27	201.24	212.35	197.53	166.05	215.32	227.17

The median time to a lapse of a policy, over all three age groups, is 47.59 months. The baseline odds of a lapse at 47.59 months is 1, that means that $P(T>$47.59 months$)=P(T<$47.59 months).

5.3.4 Deriving of indices and risk scores from Weibull regression model

Once the parameters of the Weibull baseline distribution and Weibull age group distributions have been estimated, estimated hazard and survivor functions, odds of a lapse, odds ratios and hazard ratios at time t can be calculated.

The odds ratio for age group [18;35) is the relative odds of a lapse at time t of a policy, with the age of the policyholder in [18;35), compared to a policy with the baseline characteristics. The odds ratios for the three age groups result in a set of indices, showing the effect of each age group on the baseline odds of a lapse at time t.

The hazard ratio for age group [18;35) is the relative hazard rate of a lapse at time t of a policy, with the age of the policyholder in [18;35), compared to a policy with the baseline characteristics. The hazard ratios for the three age groups result in a set of risk scores, showing the effect of each age group on the baseline hazard rate of a lapse at time t.

Percentiles of the four Weibull survival distributions can also be estimated.

The calculations of estimated hazard and survivor functions, odds of a lapse, odds ratios and hazard ratios are illustrated on the following five pages.

The survival curves and the graphs of the hazard rates of the fitted Weibull age group models are shown in Figure 5.14 and Figure 5.15.

Estimated Hazard Function at Weibull regression model

Shape remains constant

$$\boxed{\widehat{h}_0(t) = \widehat{\lambda}_0 \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}$$

$$\widehat{h}_0(t) = e^{-7.404312} \cdot 1.842334 \cdot t^{1.842334-1}$$

$$\begin{aligned} \widehat{h}_0(12) &= 0.0090938 \\ \widehat{h}_0(24) &= 0.0163048 \end{aligned}$$

$$\boxed{\widehat{h}_{A_1}(t) = \widehat{\lambda}_{A_1} \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}$$

$$\widehat{h}_{A_1}(t) = e^{-7.245223} \cdot 1.842334 \cdot t^{1.842334-1}$$

$$\begin{aligned} \widehat{h}_{A_1}(12) &= 0.010662 \\ \widehat{h}_{A_1}(24) &= 0.0191164 \end{aligned}$$

$$\boxed{\widehat{h}_{A_2}(t) = \widehat{\lambda}_{A_2} \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}$$

$$\widehat{h}_{A_2}(t) = e^{-7.438269} \cdot 1.842334 \cdot t^{1.842334-1}$$

$$\begin{aligned} \widehat{h}_{A_2}(12) &= 0.0087902 \\ \widehat{h}_{A_2}(24) &= 0.0157604 \end{aligned}$$

$$\boxed{\widehat{h}_{A_3}(t) = \widehat{\lambda}_{A_3} \cdot \widehat{\alpha} \cdot t^{\widehat{\alpha}-1}}$$

$$\widehat{h}_{A_3}(t) = e^{-7.529445} \cdot 1.842334 \cdot t^{1.842334-1}$$

$$\begin{aligned} \widehat{h}_{A_3}(12) &= 0.0080242 \\ \widehat{h}_{A_3}(24) &= 0.014387 \end{aligned}$$

Shape alters

$$\boxed{\widehat{h}_0(t) = \widehat{\lambda}_0 \cdot \widehat{\alpha}_0 t^{\widehat{\alpha}_0-1}}$$

$$\widehat{h}_0(t) = e^{-7.381423} \cdot 1.838073 \cdot t^{1.838073-1}$$

$$\begin{aligned} \widehat{h}_0(12) &= 0.0091851 \\ \widehat{h}_0(24) &= 0.0164199 \end{aligned}$$

$$\boxed{\widehat{h}_{A_1}(t) = \widehat{\lambda}_{A_1} \cdot \widehat{\alpha}_{A_1} \cdot t^{\widehat{\alpha}_{A_1}-1}}$$

$$\widehat{h}_{A_1}(t) = e^{-7.456598} \cdot 1.904217 \cdot t^{1.904217-1}$$

$$\begin{aligned} \widehat{h}_{A_1}(12) &= 0.0104033 \\ \widehat{h}_{A_1}(24) &= 0.0194701 \end{aligned}$$

$$\boxed{\widehat{h}_{A_2}(t) = \widehat{\lambda}_{A_2} \cdot \widehat{\alpha}_{A_2} \cdot t^{\widehat{\alpha}_{A_2}-1}}$$

$$\widehat{h}_{A_2}(t) = e^{-7.438269} \cdot 1.842334 \cdot t^{1.842334-1}$$

$$\begin{aligned} \widehat{h}_{A_2}(12) &= 0.0089654 \\ \widehat{h}_{A_2}(24) &= 0.0155084 \end{aligned}$$

$$\boxed{\widehat{h}_{A_3}(t) = \widehat{\lambda}_{A_3} \cdot \widehat{\alpha}_{A_3} \cdot t^{\widehat{\alpha}_{A_3}-1}}$$

$$\widehat{h}_{A_3}(t) = e^{-7.426139} \cdot 1.811986 \cdot t^{1.811986-1}$$

$$\begin{aligned} \widehat{h}_{A_3}(12) &= 0.0081153 \\ \widehat{h}_{A_3}(24) &= 0.0142474 \end{aligned}$$

Estimated Survivor Function at Weibull regression model

Shape remains constant

$$\boxed{\widehat{S}_0(t) = \exp(-\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}})}$$

$$\widehat{S}_0(t) = \exp(-e^{-7.404312} \cdot t^{1.842334})$$

$$\begin{aligned} \widehat{S}_0(12) &= 0.9424876 \\ \widehat{S}_0(24) &= 0.8086397 \end{aligned}$$

$$\boxed{\widehat{S}_{A_1}(t) = \exp(-\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}})}$$

$$\widehat{S}_{A_1}(t) = \exp(-e^{-7.245223} \cdot t^{1.842334})$$

$$\begin{aligned} \widehat{S}_{A_1}(12) &= 0.9329098 \\ \widehat{S}_{A_1}(24) &= 0.7795573 \end{aligned}$$

$$\boxed{\widehat{S}_{A_2}(t) = \exp(-\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}})}$$

$$\widehat{S}_{A_2}(t) = \exp(-e^{-7.438269} \cdot t^{1.842334})$$

$$\begin{aligned} \widehat{S}_{A_2}(12) &= 0.9443533 \\ \widehat{S}_{A_2}(24) &= 0.8143945 \end{aligned}$$

$$\boxed{\widehat{S}_{A_3}(t) = \exp(-\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}})}$$

$$\widehat{S}_{A_3}(t) = \exp(-e^{-7.529445} \cdot t^{1.842334})$$

$$\begin{aligned} \widehat{S}_{A_3}(12) &= 0.9490768 \\ \widehat{S}_{A_3}(24) &= 0.8290963 \end{aligned}$$

Shape alters

$$\boxed{\widehat{S}_0(t) = \exp(-\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0})}$$

$$\widehat{S}_0(t) = \exp(-e^{-7.381423} \cdot t^{1.8380729})$$

$$\begin{aligned} \widehat{S}_0(12) &= 0.9417969 \\ \widehat{S}_0(24) &= 0.8070283 \end{aligned}$$

$$\boxed{\widehat{S}_{A_1}(t) = \exp(-\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}_{A_1}})}$$

$$\widehat{S}_{A_1}(t) = \exp(-e^{-7.456598} \cdot t^{1.904217})$$

$$\begin{aligned} \widehat{S}_{A_1}(12) &= 0.9365433 \\ \widehat{S}_{A_1}(24) &= 0.7823969 \end{aligned}$$

$$\boxed{\widehat{S}_{A_2}(t) = \exp(-\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}_{A_2}})}$$

$$\widehat{S}_{A_2}(t) = \exp(-e^{-7.261531} \cdot t^{1.790610})$$

$$\begin{aligned} \widehat{S}_{A_2}(12) &= 0.9416866 \\ \widehat{S}_{A_2}(24) &= 0.8123183 \end{aligned}$$

$$\boxed{\widehat{S}_{A_3}(t) = \exp(-\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}_{A_3}})}$$

$$\widehat{S}_{A_3}(t) = \exp(-e^{-7.426139} \cdot t^{1.811986})$$

$$\begin{aligned} \widehat{S}_{A_3}(12) &= 0.9476747 \\ \widehat{S}_{A_3}(24) &= 0.8280275 \end{aligned}$$

Estimated Odds at Weibull regression model

Shape remains constant

$$\boxed{\widehat{odds}_0(t) = \frac{1-\widehat{S}_0(t)}{\widehat{S}_0(t)} = \exp(\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}-1})}$$

$$\widehat{odds}_0(t) = \exp(e^{-7.404312} \cdot t^{1.842334-1})$$

$$\begin{aligned}\widehat{odds}_0(12) &= 0.0610219\\ \widehat{odds}_0(24) &= 0.2366447\end{aligned}$$

$$\boxed{\widehat{odds}_{A_1}(t) = \frac{1-\widehat{S}_{A_1}(t)}{\widehat{S}_{A_1}(t)} = \exp(\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}-1})}$$

$$\widehat{odds}_{A_1}(t) = \exp(e^{-7.245223} \cdot t^{1.842334-1})$$

$$\begin{aligned}\widehat{odds}_{A_1}(12) &= 0.071915\\ \widehat{odds}_{A_1}(24) &= 0.2827793\end{aligned}$$

$$\boxed{\widehat{odds}_{A_2}(t) = \frac{1-\widehat{S}_{A_2}(t)}{\widehat{S}_{A_2}(t)} = \exp(\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}-1})}$$

$$\widehat{odds}_{A_2}(t) = \exp(e^{-7.438269} \cdot t^{1.842334-1})$$

$$\begin{aligned}\widehat{odds}_{A_2}(12) &= 0.0589258\\ \widehat{odds}_{A_2}(24) &= 0.2279062\end{aligned}$$

$$\boxed{\widehat{odds}_{A_3}(t) = \frac{1-\widehat{S}_{A_3}(t)}{\widehat{S}_{A_3}(t)} = \exp(\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}-1})}$$

$$\widehat{odds}_{A_3}(t) = \exp(e^{-7.529445} \cdot t^{1.842334-1})$$

$$\begin{aligned}\widehat{odds}_{A_3}(12) &= 0.0536555\\ \widehat{odds}_{A_3}(24) &= 0.2061326\end{aligned}$$

Shape alters

$$\boxed{\widehat{odds}_0(t) = \frac{1-\widehat{S}_0(t)}{\widehat{S}_0(t)} = \exp(\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0-1})}$$

$$\widehat{odds}_0(t) = \exp(e^{-7.381423} \cdot t^{1.8380729-1})$$

$$\begin{aligned}\widehat{odds}_0(12) &= 0.0618001\\ \widehat{odds}_0(24) &= 0.2391139\end{aligned}$$

$$\boxed{\widehat{odds}_{A_1}(t) = \frac{1-\widehat{S}_{A_1}(t)}{\widehat{S}_{A_1}(t)} = \exp(\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}_{A_1}-1})}$$

$$\widehat{odds}_{A_1}(t) = \exp(e^{-7.456598} \cdot t^{1.904217-1})$$

$$\begin{aligned}\widehat{odds}_{A_1}(12) &= 0.0677563\\ \widehat{odds}_{A_1}(24) &= 0.2781237\end{aligned}$$

$$\boxed{\widehat{odds}_{A_2}(t) = \frac{1-\widehat{S}_{A_2}(t)}{\widehat{S}_{A_2}(t)} = \exp(\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}_{A_2}-1})}$$

$$\widehat{odds}_{A_2}(t) = \exp(e^{-7.261531} \cdot t^{1.790610-1})$$

$$\begin{aligned}\widehat{odds}_{A_2}(12) &= 0.0619244\\ \widehat{odds}_{A_2}(24) &= 0.2310445\end{aligned}$$

$$\boxed{\widehat{odds}_{A_3}(t) = \frac{1-\widehat{S}_{A_3}(t)}{\widehat{S}_{A_3}(t)} = \exp(\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}_{A_3}-1})}$$

$$\widehat{odds}_{A_3}(t) = \exp(e^{-7.426139} \cdot t^{1.811986-1})$$

$$\begin{aligned}\widehat{odds}_{A_3}(12) &= 0.0552145\\ \widehat{odds}_{A_3}(24) &= 0.2076894\end{aligned}$$

Estimated Odds Ratio at Weibull regression model

Shape remains constant

$$\widehat{oddsratio}_{A_1}(t) = \frac{\widehat{odds}_{A_1}(t)}{\widehat{odds}_0(t)} = \frac{\exp(\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}-1})}{\exp(\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}-1})}$$

$$\begin{aligned} \widehat{oddsratio}_{A_1}(12) &= \tfrac{0.071915}{0.0610219} \\ &= 1.1785109 \\ \widehat{oddsratio}_{A_1}(24) &= \tfrac{0.2827793}{0.2366447} \\ &= 1.1949532 \end{aligned}$$

$$\widehat{oddsratio}_{A_2}(t) = \frac{\widehat{odds}_{A_2}(t)}{\widehat{odds}_0(t)} = \frac{\exp(\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}-1})}{\exp(\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}-1})}$$

$$\begin{aligned} \widehat{oddsratio}_{A_2}(12) &= \tfrac{0.0589258}{0.0610219} \\ &= 0.9656487 \\ \widehat{oddsratio}_{A_2}(24) &= \tfrac{0.2279062}{0.2366447} \\ &= 0.9630732 \end{aligned}$$

$$\widehat{oddsratio}_{A_3}(t) = \frac{\widehat{odds}_{A_3}(t)}{\widehat{odds}_0(t)} = \frac{\exp(\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}-1})}{\exp(\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}-1})}$$

$$\begin{aligned} \widehat{oddsratio}_{A_3}(12) &= \tfrac{0.0536555}{0.0610219} \\ &= 0.8792827 \\ \widehat{oddsratio}_{A_3}(24) &= \tfrac{0.2061326}{0.2366447} \\ &= 0.8710636 \end{aligned}$$

Odds ratio of an age group depends on time

Odds ratio is called an index

Two sets of indices, one for t=12 and one for t=24

Shape alters

$$\widehat{oddsratio}_{A_1}(t) = \frac{\widehat{odds}_{A_1}(t)}{\widehat{odds}_0(t)} = \frac{\exp(\widehat{\lambda}_{A_1} \cdot t^{\widehat{\alpha}_{A_1}-1})}{\exp(\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0-1})}$$

$$\begin{aligned} \widehat{oddsratio}_{A_1}(12) &= \tfrac{0.0677563}{0.0618001} \\ &= 1.0963792 \\ \widehat{oddsratio}_{A_1}(24) &= \tfrac{0.2781237}{0.2391139} \\ &= 1.1631435 \end{aligned}$$

$$\widehat{oddsratio}_{A_2}(t) = \frac{\widehat{odds}_{A_2}(t)}{\widehat{odds}_0(t)} = \frac{\exp(\widehat{\lambda}_{A_2} \cdot t^{\widehat{\alpha}_{A_2}-1})}{\exp(\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0-1})}$$

$$\begin{aligned} \widehat{oddsratio}_{A_2}(12) &= \tfrac{0.0619244}{0.0618001} \\ &= 1.0020119 \\ \widehat{oddsratio}_{A_2}(24) &= \tfrac{0.2310445}{0.2391139} \\ &= 0.9662529 \end{aligned}$$

$$\widehat{oddsratio}_{A_3}(t) = \frac{\widehat{odds}_{A_3}(t)}{\widehat{odds}_0(t)} = \frac{\exp(\widehat{\lambda}_{A_3} \cdot t^{\widehat{\alpha}_{A_3}-1})}{\exp(\widehat{\lambda}_0 \cdot t^{\widehat{\alpha}_0-1})}$$

$$\begin{aligned} \widehat{oddsratio}_{A_3}(12) &= \tfrac{0.0552145}{0.0618001} \\ &= 0.8934366 \\ \widehat{oddsratio}_{A_3}(24) &= \tfrac{0.2076894}{0.2391139} \\ &= 0.8685792 \end{aligned}$$

Odds ratio of an age group depends on time

Odds ratio is called an index

Two sets of indices, one for t=12 and one for t=24

Estimated Hazard Ratio at Weibull regression model

Shape remains constant

$$\widehat{hazardratio}_{A_1}(t) = \frac{\widehat{h}_{A_1}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_1}\widehat{\alpha}\cdot t^{\widehat{\alpha}-1}}{\widehat{\lambda}_0\widehat{\alpha}\cdot t^{\widehat{\alpha}-1}}$$

$$\begin{aligned}\widehat{hazardratio}_{A_1}(12) &= \frac{0.010662}{0.0090938}\\ &= 1.1724432\\ \widehat{hazardratio}_{A_1}(24) &= \frac{0.0191164}{0.0163048}\\ &= 1.1724432\end{aligned}$$

$$\widehat{hazardratio}_{A_2}(t) = \frac{\widehat{h}_{A_2}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_2}\widehat{\alpha}\cdot t^{\widehat{\alpha}-1}}{\widehat{\lambda}_0\widehat{\alpha}\cdot t^{\widehat{\alpha}-1}}$$

$$\begin{aligned}\widehat{hazardratio}_{A_2}(12) &= \frac{0.0087902}{0.0090938}\\ &= 0.9666133\\ \widehat{hazardratio}_{A_2}(24) &= \frac{0.0157604}{0.0163048}\\ &= 0.9666133\end{aligned}$$

$$\widehat{hazardratio}_{A_3}(t) = \frac{\widehat{h}_{A_3}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_3}\widehat{\alpha}\cdot t^{\widehat{\alpha}-1}}{\widehat{\lambda}_0\widehat{\alpha}\cdot t^{\widehat{\alpha}-1}}$$

$$\begin{aligned}\widehat{hazardratio}_{A_3}(12) &= \frac{0.0080242}{0.0090938}\\ &= 0.8823795\\ \widehat{hazardratio}_{A_3}(24) &= \frac{0.014387}{0.0163048}\\ &= 0.8823795\end{aligned}$$

Hazard ratio of an age group is constant over time

Hazard ratio is called a risk score

One set of risk scores, irrespective of time

Shape alters

$$\widehat{hazardratio}_{A_1}(t) = \frac{\widehat{h}_{A_1}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_1}\widehat{\alpha}_{A_1}\cdot t^{\widehat{\alpha}_{A_1}-1}}{\widehat{\lambda}_0\widehat{\alpha}_0\cdot t^{\widehat{\alpha}_0-1}}$$

$$\begin{aligned}\widehat{hazardratio}_{A_1}(12) &= \frac{0.0104033}{0.0091851}\\ &= 1.1326275\\ \widehat{hazardratio}_{A_1}(24) &= \frac{0.0194701}{0.0164199}\\ &= 1.1857647\end{aligned}$$

$$\widehat{hazardratio}_{A_2}(t) = \frac{\widehat{h}_{A_2}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_2}\widehat{\alpha}_{A_2}\cdot t^{\widehat{\alpha}_{A_2}-1}}{\widehat{\lambda}_0\widehat{\alpha}_0\cdot t^{\widehat{\alpha}_0-1}}$$

$$\begin{aligned}\widehat{hazardratio}_{A_2}(12) &= \frac{0.0089654}{0.0091851}\\ &= 0.9760801\\ \widehat{hazardratio}_{A_2}(24) &= \frac{0.0155084}{0.0164199}\\ &= 0.9444907\end{aligned}$$

$$\widehat{hazardratio}_{A_3}(t) = \frac{\widehat{h}_{A_3}(t)}{\widehat{h}_0(t)} = \frac{\widehat{\lambda}_{A_3}\widehat{\alpha}_{A_3}\cdot t^{\widehat{\alpha}_{A_3}-1}}{\widehat{\lambda}_0\widehat{\alpha}_0\cdot t^{\widehat{\alpha}_0-1}}$$

$$\begin{aligned}\widehat{hazardratio}_{A_3}(12) &= \frac{0.0081153}{0.0091851}\\ &= 0.8835268\\ \widehat{hazardratio}_{A_3}(24) &= \frac{0.0142474}{0.0164199}\\ &= 0.8676945\end{aligned}$$

Hazard ratio of an age group depends on time

Hazard ratio is called a risk score

Two sets of risk scores, one for t=12 and one for t=24

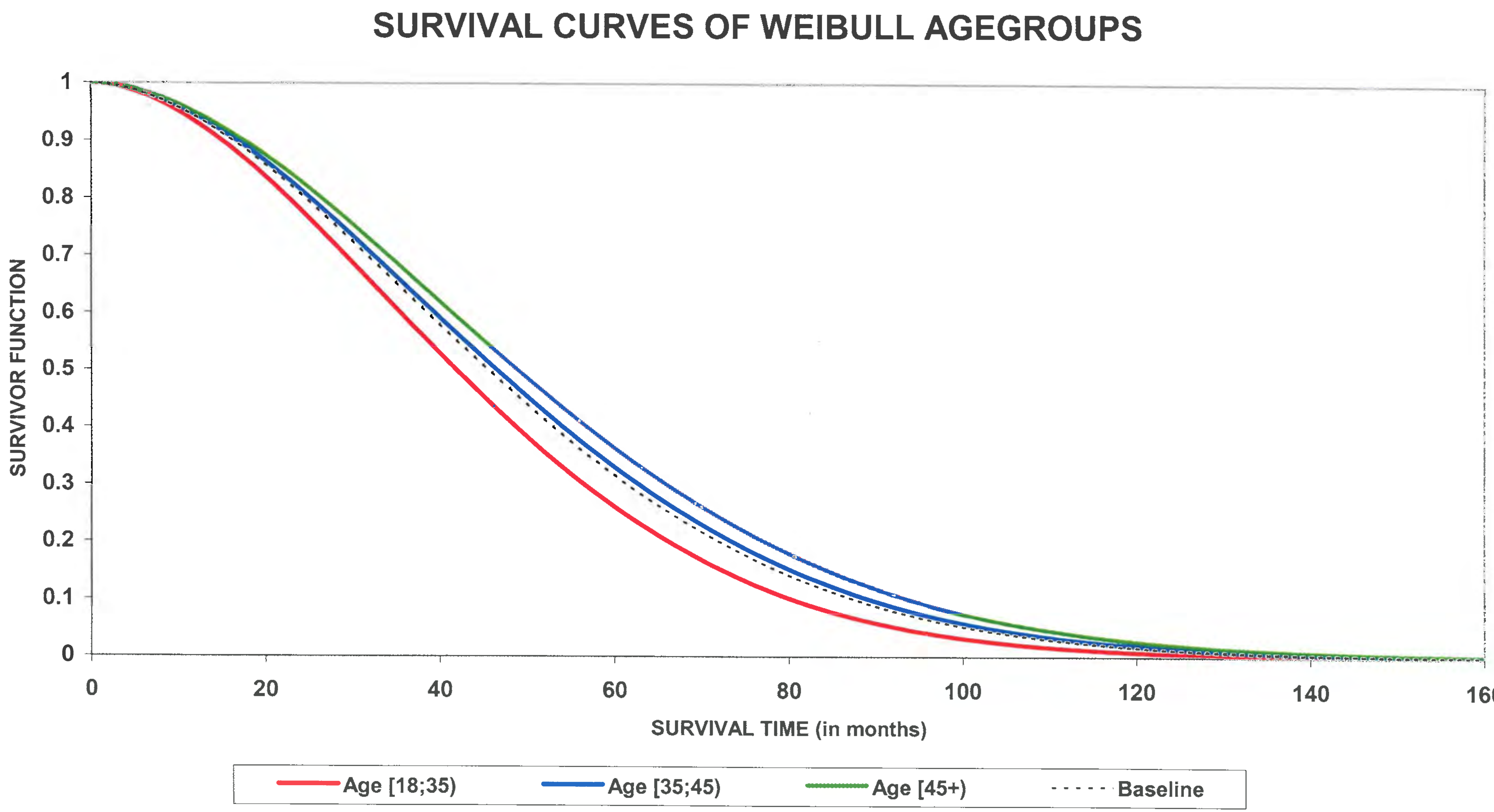

Figure 5.14: **Survival curves of fitted Weibull age group models**

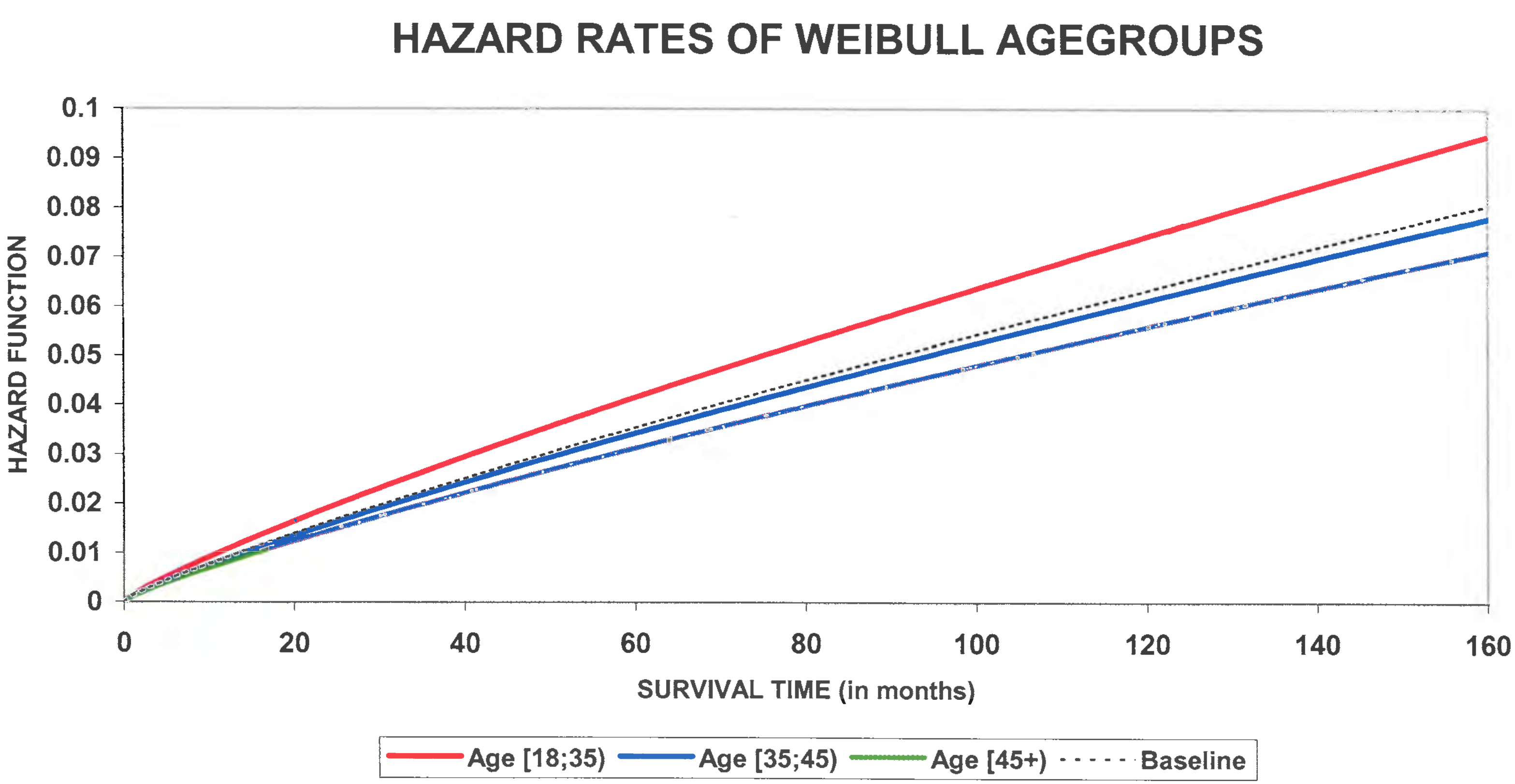

Figure 5.15: **Graphs of hazard rates of fitted Weibull age group models**

For a **constant shape parameter** in the Weibull distributions, the risk scores (estimated hazard ratios) may be obtained also from the exponent of the $\widehat{\beta}$-values in the Weibull regression model, for example

$$\begin{aligned} e^{\widehat{\beta}_{A_1}} &= e^{0.1590898} &= 1.1724432 \\ e^{\widehat{\beta}_{A_2}} &= e^{-0.033957} &= 0.9666133 \\ e^{\widehat{\beta}_{A_3}} &= e^{-0.125133} &= 0.8823795. \end{aligned}$$

The indices (estimated odds ratios) of the three age groups, estimated from the Weibull regression model, are compared to the indices, obtained from the logit model, in Table 5.19.

The index of age group [45+) of 0.879283 shows the effect of this age group on the baseline

Table 5.19: **Comparison of indices: Weibull regression model and logit model**

		Weibull regression model				Logit model	
		Shape remains constant		Shape alters			
Effect	n	Index		Index		Index	
		t=12	t=24	t=12	t=24	t=12	t=24
Baseline odds	10077	0.061022	0.236645	0.061800	0.239114	0.0543	0.2570
Age [18;35)	3644	1.178511	1.194953	1.096379	1.163143	1.1321	1.1370
Age [35;45)	3425	0.965649	0.963073	1.002012	0.966253	0.9679	0.9817
Age [45+)	3008	0.879283	0.871064	0.893437	0.868579	0.9126	0.8960

odds of a lapse. This effect is multiplicative on the baseline odds of a lapse. Thus the effect of age group [45+) is to decrease the baseline odds of a lapse by a factor 0.879283 .

From Table 5.19 follows that one Weibull regression model provides indices for any time value, while a new logitmodel has to be built for a fixed time value, say t=12 months, conditional on a restricted experimental design where all the policies must have an exposure of at least one year when investigating the lapses of policies in the first year. There is no such restrictions in the more general experimental design for the Weibull regression model where all the policies can be used in the analysis, even those policies with inception dates very close to the cut-off point.

Predicted indices from the Weibull regression model, for varying time values, are shown in Table 5.20 (constant shape) and in Table 5.21 (shape alters).

Predicted risk scores from the Weibull regression model, for varying time values, are shown in Table 5.22 (constant shape) and in Table 5.23 (shape alters).

Table 5.19: **Predicted indices from Weibull regression model (constant shape)**

Effect	t=6	t=12	t=18	t=24	t=30	t=36	t=42	t=48	t=54	t=60
	Predicted indices from Weibull regression model									
Baseline odds	0.016655	0.061022	0.133171	0.236645	0.377685	0.565662	0.814022	1.141810	1.575977	2.154842
Age [18;35)	1.174119	1.178511	1.185439	1.194953	1.207213	1.222464	1.241024	1.263282	1.289700	1.320802
Age [35;45)	0.966346	0.965649	0.964557	0.963073	0.961187	0.958881	0.956128	0.952901	0.94917	0.944903
Age [45+)	0.881521	0.879283	0.875789	0.871064	0.86509	0.857832	0.849241	0.839262	0.827842	0.814934

Table 5.20: **Predicted indices from Weibull regression model (shape alters)**

Effect	t=6	t=12	t=18	t=24	t=30	t=36	t=42	t=48	t=54	t=60
	Predicted indices from Weibull regression model									
Baseline odds	0.016913	0.061800	0.134678	0.239114	0.381412	0.571040	0.821586	1.152299	1.590420	2.174686
Age [18;35)	1.044681	1.096379	1.131960	1.163143	1.194129	1.227247	1.264219	1.306589	1.355927	1.413938
Age [35;45)	1.035773	1.002012	0.981769	0.966253	0.952697	0.939827	0.926922	0.913510	0.899255	0.883902
Age [45+)	0.911930	0.893437	0.880373	0.868579	0.856722	0.844171	0.830554	0.815621	0.799192	0.781140

Table 5.21: **Predicted risk scores from Weibull regression model (constant shape)**

Effect	Predicted risk scores from Weibull regression model									
	t=6	t=12	t=18	t=24	t=30	t=36	t=42	t=48	t=54	t=60
Baseline hazard rate	0.005072	0.009094	0.012796	0.016305	0.019676	0.022943	0.026124	0.029234	0.032283	0.035279
Age [18;35)	1.172443	1.172443	1.172443	1.172443	1.172443	1.172443	1.172443	1.172443	1.172443	1.172443
Age [35;45)	0.966613	0.966613	0.966613	0.966613	0.966613	0.966613	0.966613	0.966613	0.966613	0.966613
Age [45+)	0.882380	0.882380	0.882380	0.882380	0.882380	0.882380	0.882380	0.882380	0.882380	0.882380

Table 5.22: **Predicted risk scores from Weibull regression model (shape alters)**

Effect	Predicted risk scores from Weibull regression model									
	t=6	t=12	t=18	t=24	t=30	t=36	t=42	t=48	t=54	t=60
Baseline hazard rate	0.005138	0.009185	0.012902	0.016420	0.019796	0.023065	0.026245	0.029353	0.032398	0.035389
Age [18;35)	1.081872	1.132627	1.163415	1.185765	1.203396	1.217996	1.230479	1.241395	1.251104	1.259853
Age [35;45)	1.008726	0.976080	0.957475	0.944491	0.934540	0.926488	0.919734	0.913924	0.908829	0.904295
Age [45+)	0.899648	0.883527	0.874231	0.867694	0.862658	0.858565	0.855119	0.852146	0.849532	0.847200

For a **constant shape parameter** in the Weibull distributions, the pth percentile of the baseline lifetime distribution can be calculated from

$$\frac{1}{\widehat{\lambda}_0} \cdot \ln \frac{100}{100-p})^{\frac{1}{\widehat{\alpha}}}$$

and that of the age group distributions from

$$\frac{1}{\widehat{\lambda}_{A_i}} \cdot \ln \frac{100}{100-p})^{\frac{1}{\widehat{\alpha}}} \quad i = 1, 2, 3.$$

Note that the latter is equal to the pth percentile of the baseline distribution multiplied by the specific index to the power $-\frac{1}{\alpha}$.

For a **shape parameter that alters**, the formulaes change to

$$\frac{1}{\widehat{\lambda}_0} \cdot \ln \frac{100}{100-p})^{\frac{1}{\widehat{\alpha}_0}} \quad \text{and} \quad \frac{1}{\widehat{\lambda}_{A_i}} \cdot \ln \frac{100}{100-p})^{\frac{1}{\widehat{\alpha}_{A_i}}} \quad i = 1, 2, 3.$$

The estimated percentiles of the baseline and the age group Weibull distributions for a constant shape parameter as well as for different shape parameters are reported in Table 5.24.

Table 5.24: **Lifetime percentiles estimated from Weibull regression model**

	Weibull regression model							
	Lifetime distribution: constant shape				Lifetime distribution: shape alters			
Lifetime percentile	Baseline	Age group [18;35)	[35;45)	[45+)	Baseline	Age group [18;35)	[35;45)	[45+)
P5	11.10	10.18	11.30	11.88	11.02	10.55	10.98	11.69
P10	16.40	15.05	16.71	17.56	16.31	15.39	16.42	17.40
P20	24.65	22.61	25.11	26.38	24.53	22.83	24.97	26.32
P25	28.30	25.95	28.82	30.28	28.16	26.09	28.78	30.29
P30	31.80	29.17	32.39	34.03	31.66	29.21	32.45	34.10
P40	38.64	35.45	39.36	41.36	38.49	35.27	39.65	41.58
P50	45.61	41.83	46.45	48.81	45.44	41.40	47.02	49.21
P60	53.06	48.67	54.05	56.79	52.89	47.94	54.95	57.40
P70	61.54	56.45	62.69	65.87	61.36	55.33	64.01	66.74
P75	66.44	60.94	67.67	71.11	66.26	59.58	69.25	72.14
P80	72.04	66.08	73.38	77.11	71.86	64.44	75.27	78.33
P90	87.50	80.26	89.13	93.65	87.32	77.77	91.94	95.45
P95	100.94	92.59	102.82	108.03	100.76	89.30	106.49	110.37

The median time to a lapse of a policy, over all three age groups, is 45.61 months. The baseline odds of a lapse at 45.61 months is 1, that means that $P(T >$45.61 months)$=P(T <$45.61 months). The lowest estimated percentile lifetime values in the third column of Table 5.24 again confirm the highest risk of a policy to lapse if the policyholder is in the youngest agegroup.

5.3.5 The fitting of a regression model with a continuous predictor

Consider AGE as a continuous predictor that can be categorized into three age groups. The ordinal covariate Z takes on the values

z=1 for the age group [18;35)
z=2 for the age group [35;45)
z=3 for age group [45+).

The midpoints of the age group intervals can also be used as values of the continuous predictor AGE, that means

$z = \frac{18+34}{2} = 26$ for age group $[18;35)$
$z = \frac{35+44}{2} = 39.5$ for age group $[35;45)$
$z = \frac{45+59}{2} = 52$ for age group $[45+)$
if 60 months is assumed to be an upper limit for the open interval.

A log-logistic as well as a Weibull regression model are fitted to the grouped survival data with known z-values. From the estimated regression parameters, survival model parameters can be found for each age group as well as for the baseline distribution.

The estimated regression coefficients of these two regression models are reported in Table 5.25.

Table 5.25: **Fitting a regression model (constant shape) to grouped data with one continuous predictor**

		Regression model			
Effect	Maximum likelihood estimates	Log-logistic		Weibull	
		z-values 1 2 3	z-values 26 39.5 52	z-values 1 2 3	z-values 26 39.5 52
Baseline mean	$\widehat{\ln \lambda_0} = \ln \widehat{\lambda}_0$	-7.647250	-7.477800	-7.111259	-6.962854
Constant shape	$\widehat{\alpha}$	2.066059	2.066104	1.841998	1.842030
Age	$\widehat{\beta}$	-0.166957	-0.012856	-0.146264	-0.011261

The estimated lambda parameters of the three survival distributions for the three age groups then are

$$\widehat{\lambda}_{Age1} = \exp(\ln\widehat{\lambda}_0 + \widehat{\beta} * 1)$$
$$\widehat{\lambda}_{Age2} = \exp(\ln\widehat{\lambda}_0 + \widehat{\beta} * 2)$$
$$\widehat{\lambda}_{Age3} = \exp(\ln\widehat{\lambda}_0 + \widehat{\beta} * 3)$$

or

$$\widehat{\lambda}_{Age1} = \exp(\ln\widehat{\lambda}_0 + \widehat{\beta} * 26)$$
$$\widehat{\lambda}_{Age2} = \exp(\ln\widehat{\lambda}_0 + \widehat{\beta} * 39.5)$$
$$\widehat{\lambda}_{Age3} = \exp(\ln\widehat{\lambda}_0 + \widehat{\beta} * 52)$$

with the same estimated alpha parameter $\widehat{\alpha}$. These parameters are summarized for each age group in Table 5.26.

Table 5.26: **Parameters of a survival model (constant shape) for each age group**

AGE group	Maximum likelihood estimates	**Survival model**			
		Log-logistic		Weibull	
		z-values 1 2 3	z-values 26 39.5 52	z-values 1 2 3	z-values 26 39.5 52
Age [18;35)	$\ln\widehat{\lambda}_z$	-7.814208	-7.490656	-7.257523	-7.800792
	$\widehat{\alpha}$	2.0660588	2.0661037	1.841998	1.8421299
Age [35;45)	$\ln\widehat{\lambda}_z$	-7.901165	-7.503511	-7.403787	-8.016725
	$\widehat{\alpha}$	2.0660588	2.0661037	1.841998	1.8421299
Age [45+)	$\ln\widehat{\lambda}_z$	-8.148122	-7.516367	-7.550051	-8.127733
	$\widehat{\alpha}$	2.0660588	2.0661037	1.841998	1.8421299
Baseline	$\ln\widehat{\lambda}_0$	-7.647250	-7.477800	-7.111259	-6.962854
	$\widehat{\alpha}$	2.0660588	2.0661037	1.841998	1.8421299

5.3.6 A survival model for each combination of levels of two risk factors

Consider two risk factors AGE and SCORE where AGE has three levels [18;35), [35;45) and [45+) years and SCORE has three levels 'Low', 'Medium' and 'High'.

A cross tabulation of AGE and SCORE for the 10077 observations are given in Table 5.27.

Table 5.27: **Cross table of Age and Score**

	Score			
	Low	Medium	High	Total
Age [18;35)	833	1758	1053	3644
Age [35;45)	769	1546	1110	3425
Age [45+)	813	1541	654	3008
Total	2415	4845	2817	10077

A regression model is fitted to the grouped survival data where each policy has information on the entry period, age level and score level. The grouped lifetimes of the policies with staggered entry as well as the concomitant information on AGE and SCORE are given in Table 5.28.

The combined frequency vector is

$$\boldsymbol{f}' = (\boldsymbol{f}'_{111}, \boldsymbol{f}'_{211}, \boldsymbol{f}'_{311}, \boldsymbol{f}'_{411}, \boldsymbol{f}'_{112}, \boldsymbol{f}'_{212}, \boldsymbol{f}'_{312}, \boldsymbol{f}'_{412}, \boldsymbol{f}'_{113}, \boldsymbol{f}'_{213}, \boldsymbol{f}'_{313}, \boldsymbol{f}'_{413}, \boldsymbol{f}'_{121}, \boldsymbol{f}'_{221}, \boldsymbol{f}'_{321}, \boldsymbol{f}'_{421}, \boldsymbol{f}'_{122}, \boldsymbol{f}'_{222}, \boldsymbol{f}'_{322}, \boldsymbol{f}'_{422}, \boldsymbol{f}'_{123}, \boldsymbol{f}'_{223}, \boldsymbol{f}'_{323}, \boldsymbol{f}'_{423}, \boldsymbol{f}'_{131}, \boldsymbol{f}'_{231}, \boldsymbol{f}'_{331}, \boldsymbol{f}'_{431}, \boldsymbol{f}'_{132}, \boldsymbol{f}'_{232}, \boldsymbol{f}'_{332}, \boldsymbol{f}'_{432}, \boldsymbol{f}'_{133}, \boldsymbol{f}'_{233}, \boldsymbol{f}'_{333}, \boldsymbol{f}'_{433}).$$

$\boldsymbol{f}_{ilm}$ is the frequency vector for the i^{th} entry group, the l^{th} AGE level and the m^{th} SCORE level

$i = 1, 2, 3, 4$ and $l = 1, 2, 3$ and $m = 1, 2, 3$.

Table 5.28: **Multi-dimensional frequency table of grouped data set with two risk factors**

Entry	Age	Score	Lifetime intervals						
March 98			$[0,12)$	$[12,17)$	$[17,24)$	$[24,28)$	$[28,34)$	$[34,37)$	$[37,\infty)$
	[18;35)	Low	12	34	51	39	57	11	59
		Medium	10	12	22	19	32	4	418
		High	7	13	22	15	19	0	165
	[35;45)	Low	13	14	45	27	33	4	66
		Medium	4	22	22	8	25	4	297
		High	4	14	24	10	17	5	190
	[45+)	Low	10	25	29	17	46	2	116
		Medium	6	13	28	16	16	5	273
		High	0	11	11	6	5	0	82
June 98			$[0,12)$	$[12,17)$	$[17,24)$	$[24,28)$	$[28,34)$	$[34,\infty)$	
	[18;35)	Low	22	25	58	53	40	45	
		Medium	10	26	32	20	29	379	
		High	9	24	13	19	14	204	
	[35;45)	Low	24	24	28	30	25	106	
		Medium	12	20	14	17	16	409	
		High	13	18	19	19	13	238	
	[45+)	Low	13	15	32	19	17	107	
		Medium	11	13	22	17	12	319	
		High	4	1	11	6	6	117	
Nov 98			$[0,12)$	$[12,17)$	$[17,24)$	$[24,28)$	$[28,\infty)$		
	[18;35)	Low	34	16	50	23	54		
		Medium	19	2	32	24	317		
		High	15	16	17	10	199		
	[35;45)	Low	19	18	38	16	75		
		Medium	16	14	25	10	263		
		High	5	12	20	7	195		
	[45+)	Low	28	16	22	12	98		
		Medium	13	0	24	4	323		
		High	5	5	14	11	150		
March 99			$[0,12)$	$[12,17)$	$[17,24)$	$[24,\infty)$			
	[18;35)	Low	40	30	30	50			
		Medium	9	14	27	301			
		High	22	16	12	222			
	[35;45)	Low	24	30	29	81			
		Medium	14	15	12	307			
		High	16	16	27	228			
	[45+)	Low	20	22	28	119			
		Medium	19	12	26	369			
		High	11	11	16	171			

$$\boldsymbol{f}_{111} = (12, 34, 51, 39, 57, 11, 59)'$$
$$\boldsymbol{f}_{112} = (10, 12, 22, 19, 32, 4, 418)'$$
$$\boldsymbol{f}_{113} = (7, 13, 22, 15, 19, 0, 165)'$$
$$\boldsymbol{f}_{121} = (13, 14, 45, 27, 33, 4, 66)'$$
$$\boldsymbol{f}_{122} = (4, 22, 22, 8, 25, 4, 297)'$$
$$\boldsymbol{f}_{123} = (4, 14, 24, 10, 17, 5, 190)'$$
$$\boldsymbol{f}_{131} = (10, 25, 29, 17, 46, 2, 116)'$$
$$\boldsymbol{f}_{132} = (6, 13, 28, 16, 16, 5, 273)'$$
$$\boldsymbol{f}_{133} = (0, 11, 11, 6, 5, 0, 82)'$$

$$\boldsymbol{f}_{211} = (22, 25, 58, 53, 40, 45)'$$
$$\boldsymbol{f}_{212} = (10, 26, 32, 20, 29, 379)'$$
$$\boldsymbol{f}_{213} = (9, 24, 13, 19, 14, 204)'$$
$$\boldsymbol{f}_{221} = (24, 24, 28, 30, 25, 106)'$$
$$\boldsymbol{f}_{222} = (12, 20, 14, 17, 16, 409)'$$
$$\boldsymbol{f}_{223} = (13, 18, 19, 19, 13, 238)'$$
$$\boldsymbol{f}_{231} = (13, 15, 32, 19, 17, 107)'$$
$$\boldsymbol{f}_{232} = (11, 13, 22, 17, 12, 319)'$$
$$\boldsymbol{f}_{233} = (4, 1, 11, 6, 6, 117)'$$

$$\boldsymbol{f}_{311} = (34, 16, 50, 23, 54)'$$
$$\boldsymbol{f}_{312} = (19, 2, 32, 24, 317)'$$
$$\boldsymbol{f}_{313} = (15, 16, 17, 10, 199)'$$
$$\boldsymbol{f}_{321} = (19, 18, 38, 16, 75)'$$
$$\boldsymbol{f}_{322} = (16, 14, 25, 10, 263)'$$
$$\boldsymbol{f}_{323} = (5, 12, 20, 7, 195)'$$
$$\boldsymbol{f}_{331} = (28, 16, 22, 12, 98)'$$
$$\boldsymbol{f}_{332} = (13, 0, 24, 4, 323)'$$
$$\boldsymbol{f}_{333} = (5, 5, 14, 11, 150)'$$

$$\boldsymbol{f}_{411} = (40, 30, 30, 50)'$$
$$\boldsymbol{f}_{412} = (9, 14, 27, 301)'$$
$$\boldsymbol{f}_{413} = (22, 16, 12, 222)'$$
$$\boldsymbol{f}_{421} = (24, 30, 29, 81)'$$
$$\boldsymbol{f}_{422} = (14, 15, 12, 307)'$$

$\boldsymbol{f}_{423} = (16, 16, 27, 228)'$
$\boldsymbol{f}_{431} = (20, 22, 28, 119)'$
$\boldsymbol{f}_{432} = (19, 12, 26, 369)'$
$\boldsymbol{f}_{433} = (11, 11, 16, 171)'$

The vectors $\boldsymbol{x}_i \quad i = 1, 2, 3, 4$ of upper class boundaries for the i^{th} entry group are

$$\boldsymbol{x}_1 = \begin{pmatrix} 12 \\ 17 \\ 24 \\ 28 \\ 34 \\ 37 \end{pmatrix} \quad \boldsymbol{x}_2 = \begin{pmatrix} 12 \\ 17 \\ 24 \\ 28 \\ 34 \end{pmatrix} \quad \boldsymbol{x}_3 = \begin{pmatrix} 12 \\ 17 \\ 24 \\ 28 \end{pmatrix} \quad \text{and} \quad \boldsymbol{x}_4 = \begin{pmatrix} 12 \\ 17 \\ 24 \end{pmatrix}.$$

From the estimated regression parameters, survival model parameters can be found for each combination of the levels of AGE and SCORE as well as for the baseline distribution. The estimated regression coefficients of the regression model with two risk factors AGE and SCORE are reported in Table 5.29.

Table 5.29: **Fitting a regression model with two risk factors**

Effect	Maximum likelihood estimates	**Regression model** Log-logistic	Weibull
Baseline mean	$\widehat{\ln \lambda_0} = \ln \widehat{\lambda}_0$	-8.550810	-7.709833
Age [18;35)	$\widehat{\beta}_{A_1}$	0.205367	0.212709
Age [35;45)	$\widehat{\beta}_{A_2}$	-0.011853	-0.014725
Age [45+)	$\widehat{\beta}_{A_3}$	-0.193514	-0.197984
Score 'Low'	$\widehat{\beta}_{B_1}$	1.047861	0.897721
Score 'Medium'	$\widehat{\beta}_{B_2}$	-0.714941	-0.612472
Score 'High'	$\widehat{\beta}_{B_3}$	-0.332746	-0.285249
Constant shape	$\widehat{\alpha}$	2.249510	1.938292

The estimated lambda parameters of the nine survival distributions for the nine combinations of AGE and SCORE levels are

$$\begin{aligned}
\widehat{\lambda}_{A_1B_1} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_1} + \widehat{\beta}_{B_1}) \\
\widehat{\lambda}_{A_1B_2} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_1} + \widehat{\beta}_{B_2}) \\
\widehat{\lambda}_{A_1B_3} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_1} + \widehat{\beta}_{B_3}) \\
\widehat{\lambda}_{A_2B_1} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_2} + \widehat{\beta}_{B_1}) \\
\widehat{\lambda}_{A_2B_2} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_2} + \widehat{\beta}_{B_2}) \\
\widehat{\lambda}_{A_2B_3} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_2} + \widehat{\beta}_{B_3}) \\
\widehat{\lambda}_{A_3B_1} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_3} + \widehat{\beta}_{B_1}) \\
\widehat{\lambda}_{A_3B_2} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_3} + \widehat{\beta}_{B_2}) \\
\widehat{\lambda}_{A_3B_3} &= \exp(\ln \widehat{\lambda}_0 + \widehat{\beta}_{A_3} + \widehat{\beta}_{B_3})
\end{aligned}$$

with the same estimated alpha parameter $\widehat{\alpha}$. These parameters are summarized for each combination of AGE and SCORE levels in Table 5.30.

Table 5.30: **Parameters of a survival model for each combination of AGE and SCORE levels**

Combination of Age and Score	Maximum likelihood estimates	Survival model Log-logistic	Survival model Weibull
Age [18;35) and Low score	$\ln\widehat{\lambda}_{A_1B_1}$	-7.297757	-6.599403
	$\widehat{\alpha}$	2.249510	1.938292
Age [18;35) and Medium score	$\ln\widehat{\lambda}_{A_1B_2}$	-9.060383	-8.109596
	$\widehat{\alpha}$	2.249510	1.938292
Age [18;35) and High score	$\ln\widehat{\lambda}_{A_1B_3}$	-8.678188	-7.782373
	$\widehat{\alpha}$	2.249510	1.938292
Age [35;45) and Low score	$\ln\widehat{\lambda}_{A_2B_1}$	-7.514976	-6.826837
	$\widehat{\alpha}$	2.249510	1.938292
Age [35;45) and Medium score	$\ln\widehat{\lambda}_{A_2B_2}$	-9.277603	-8.337030
	$\widehat{\alpha}$	2.249510	1.938292
Age [35;45) and High score	$\ln\widehat{\lambda}_{A_2B_3}$	-8.895408	-8.009807
	$\widehat{\alpha}$	2.249510	1.938292
Age [45+) and Low score	$\ln\widehat{\lambda}_{A_3B_1}$	-7.696638	-7.010096
	$\widehat{\alpha}$	2.249510	1.938292
Age [45+) and Medium score	$\ln\widehat{\lambda}_{A_3B_2}$	-9.459265	-8.520288
	$\widehat{\alpha}$	2.249510	1.938292
Age [45+) and High score	$\ln\widehat{\lambda}_{A_3B_3}$	-9.077070	-8.193066
	$\widehat{\alpha}$	2.249510	1.938292
Baseline	$\ln\widehat{\lambda}_0$	-8.55081	-7.709833
	$\widehat{\alpha}$	2.249510	1.938292

A joint histogram to the data of each combination of AGE and SCORE levels **over the four entry groups** is needed to make a graphical representation of the fitted models for each combination of AGE and SCORE levels. Table 5.31 gives the nine sets of fitted joint frequencies for the nine combinations of AGE and SCORE levels. This fitting was done by maximum likelihood estimation subject to constraints imposed by the experimental design. The Wald test and discrepancy value measure the goodness-of-fit.

Table 5.31: **Fitted joint frequency distributions for the nine combinations of AGE and SCORE levels**

Interval number	Interval of survival times	Fitted Joint Frequencies Age [18;35) Low score	Age [35;45) Low score	Age [45+) Low score
first	$[0,12)$	108	80	71
second	$[12,17)$	105	86	78
third	$[17,24)$	189	140	111
fourth	$[24,28)$	130.43421	90.690721	61.44444
fifth	$[28,34)$	137.52303	92.281787	107.52778
sixth	$[34,37)$	25.621006	16.001571	6.508945
seventh	$[37,\infty)$	137.42176	264.02592	377.51883
Wald		87.06	38.99	35.20
Discrepancy		0.1045	0.0507	0.0432974

Interval number	Interval of survival times	Fitted Joint Frequencies Age [18;35) Medium score	Age [35;45) Medium score	Age [45+) Medium score
first	$[0,12)$	48	46	49
second	$[12,17)$	54	71	38.000364
third	$[17,24)$	113	73	100
fourth	$[24,28)$	66.789123	43.685567	67.905775
fifth	$[28,34)$	104.46504	71.64433	57.617021
sixth	$[34,37)$	13.00233	16.48731	22.094914
seventh	$[37,\infty)$	1358.7435	1224.1828	1206.3823
Wald		34.26	31.51	20.50
Discrepancy		0.0195	0.0204	0.0133

Interval number	Interval of survival times	Fitted Joint Frequencies Age [18;35) High score	Age [35;45) High score	Age [45+) High score
first	$[0,12)$	53	38	20.000115
second	$[12,17)$	69	60	28
third	$[17,24)$	64	90	52
fourth	$[24,28)$	67.610092	54.345528	29.945937
fifth	$[28,34)$	65.62156	56.219512	27.450442
sixth	$[34,37)$	0	20.806025	0.0006965
seventh	$[37,\infty)$	733.76835	790.62893	496.60304
Wald		20.07	14.81	22.08
Discrepancy		0.0191	0.0133428	0.0338

Figure 5.16 shows the fitted joint histogram and the fitted survival distributions for age group [18;35) and low score.

Figure 5.17 shows the fitted joint histogram and the fitted survival distributions for age group [18;35) and medium score.

Figure 5.18 shows the fitted joint histogram and the fitted survival distributions for age group [18;35) and high score.

Figure 5.19 shows the fitted joint histogram and the fitted survival distributions for age group [35;45) and low score.

Figure 5.20 shows the fitted joint histogram and the fitted survival distributions for age group [35;45) and medium score.

Figure 5.21 shows the fitted joint histogram and the fitted survival distributions for age group [35;45) and high score.

Figure 5.22 shows the fitted joint histogram and the fitted survival distributions for age group [45+) and low score.

Figure 5.23 shows the fitted joint histogram and the fitted survival distributions for age group [45+) and medium score.

Figure 5.24 shows the fitted joint histogram and the fitted survival distributions for age group [45+) and high score.

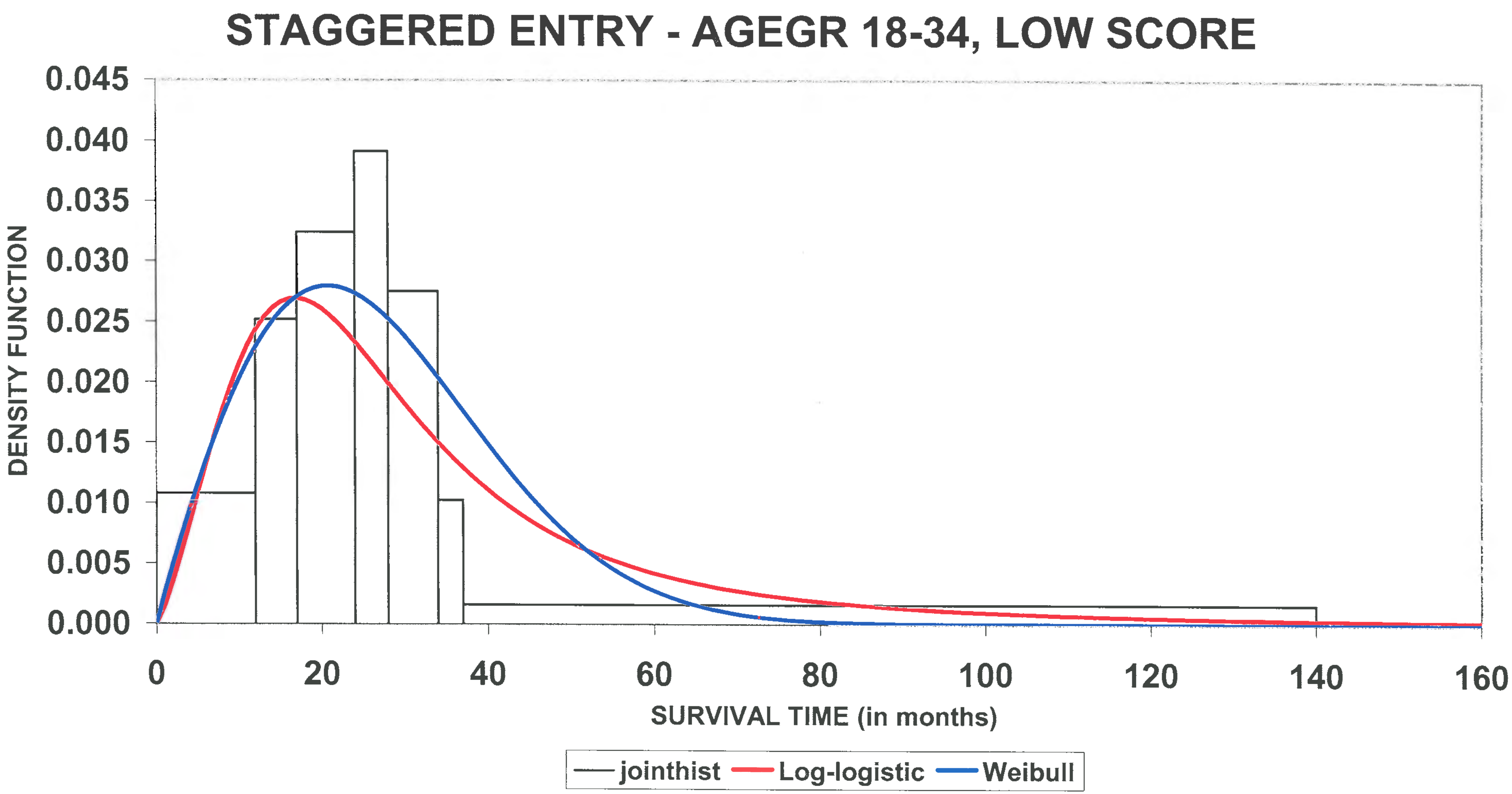

Figure 5.16: **Joint histogram and fitted survival distributions for age group [18;35) and low score**

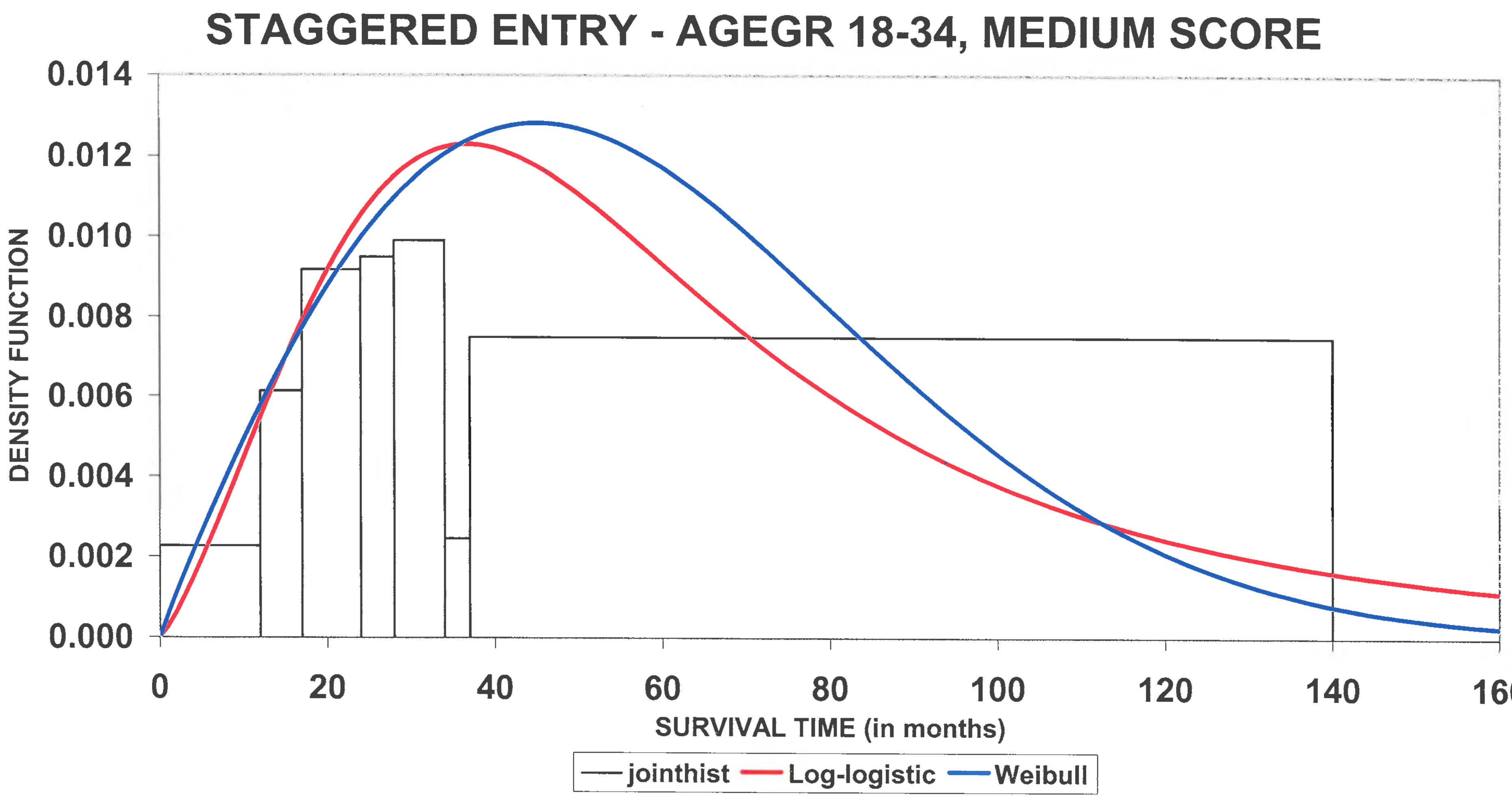

Figure 5.17: **Joint histogram and fitted survival distributions for age group [18;35) and medium score**

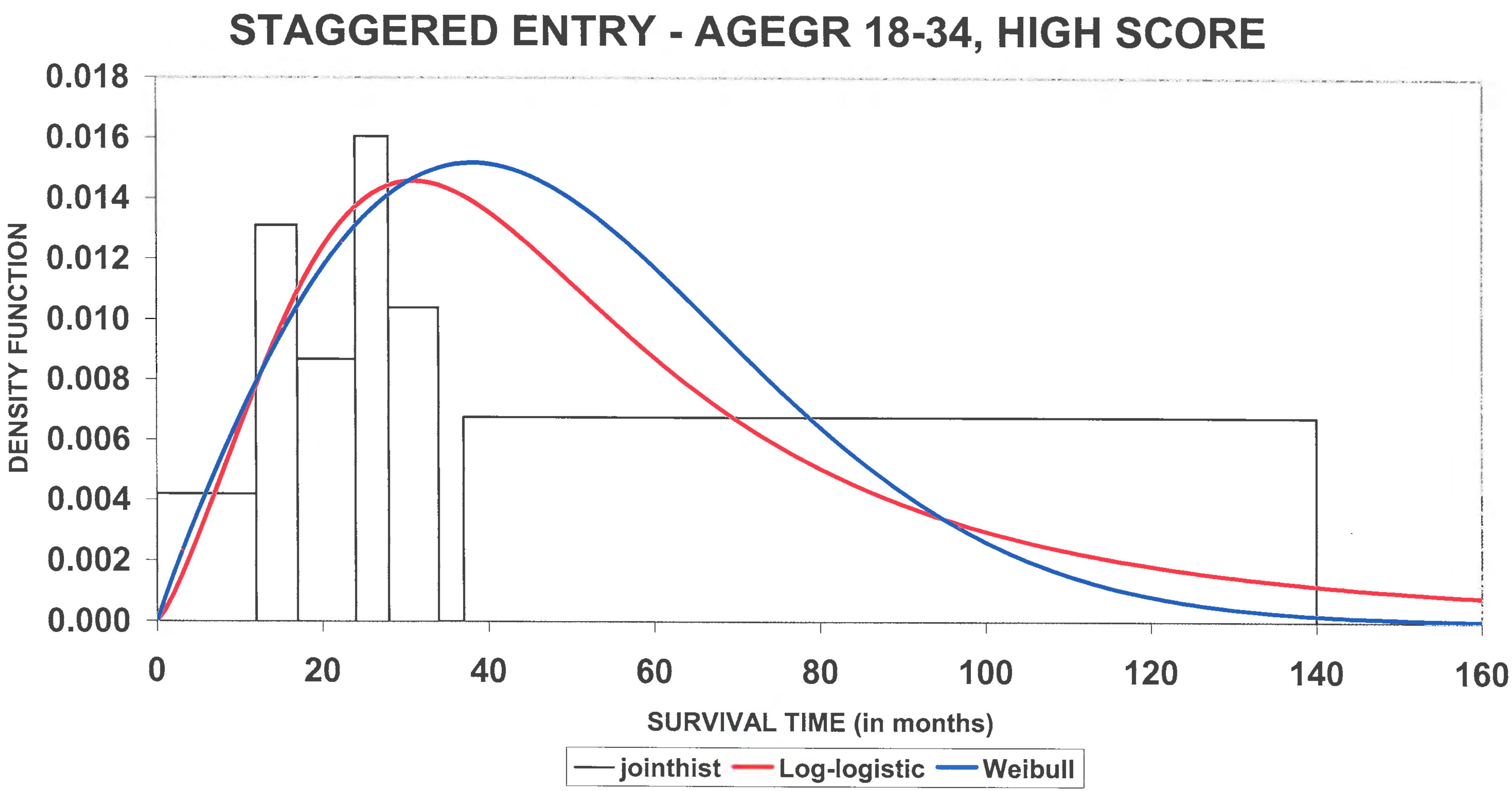

Figure 5.18: **Joint histogram and fitted survival distributions for age group [18;35) and high score**

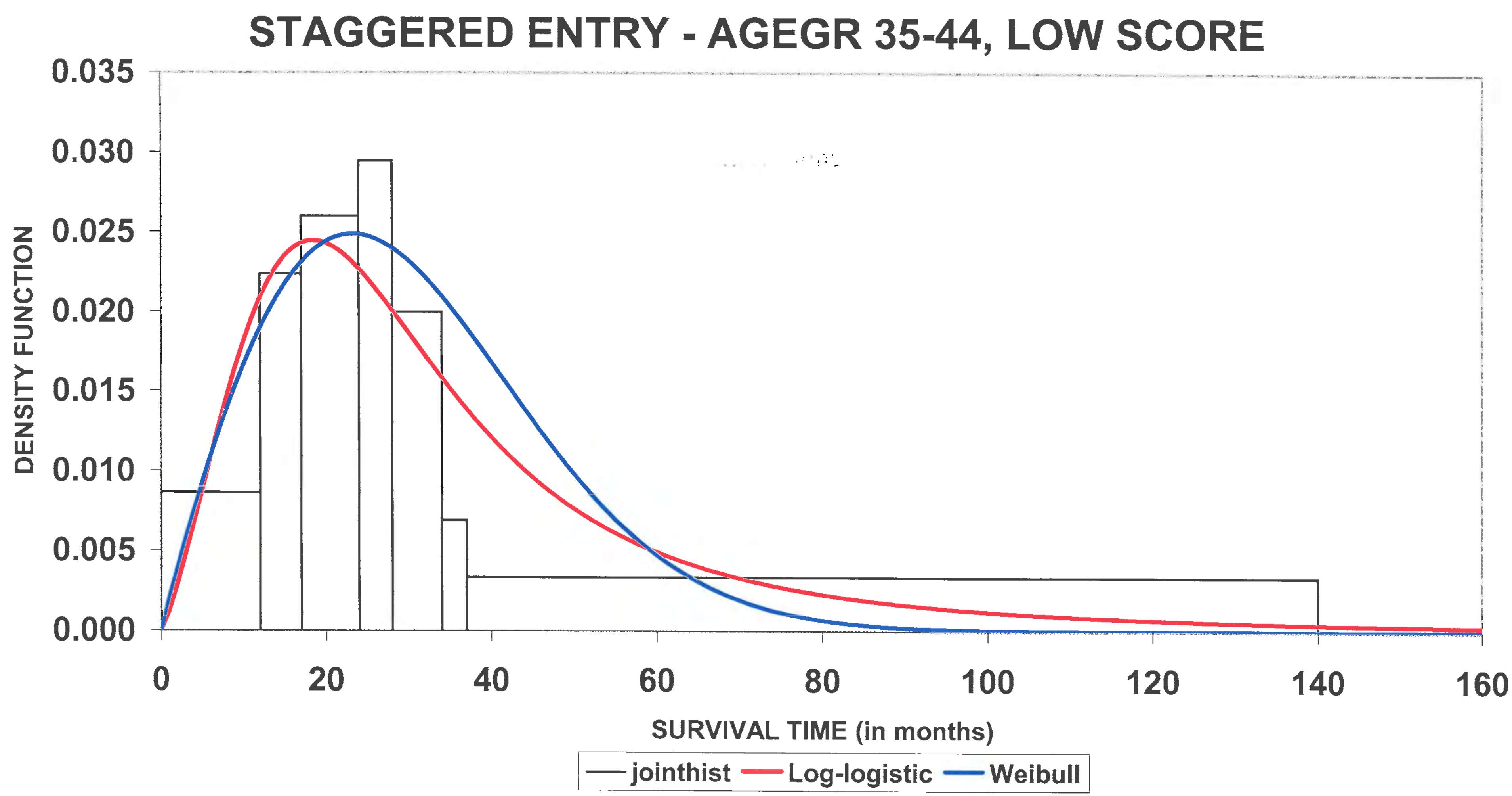

Figure 5.19: **Joint histogram and fitted survival distributions for age group [35;45) and low score**

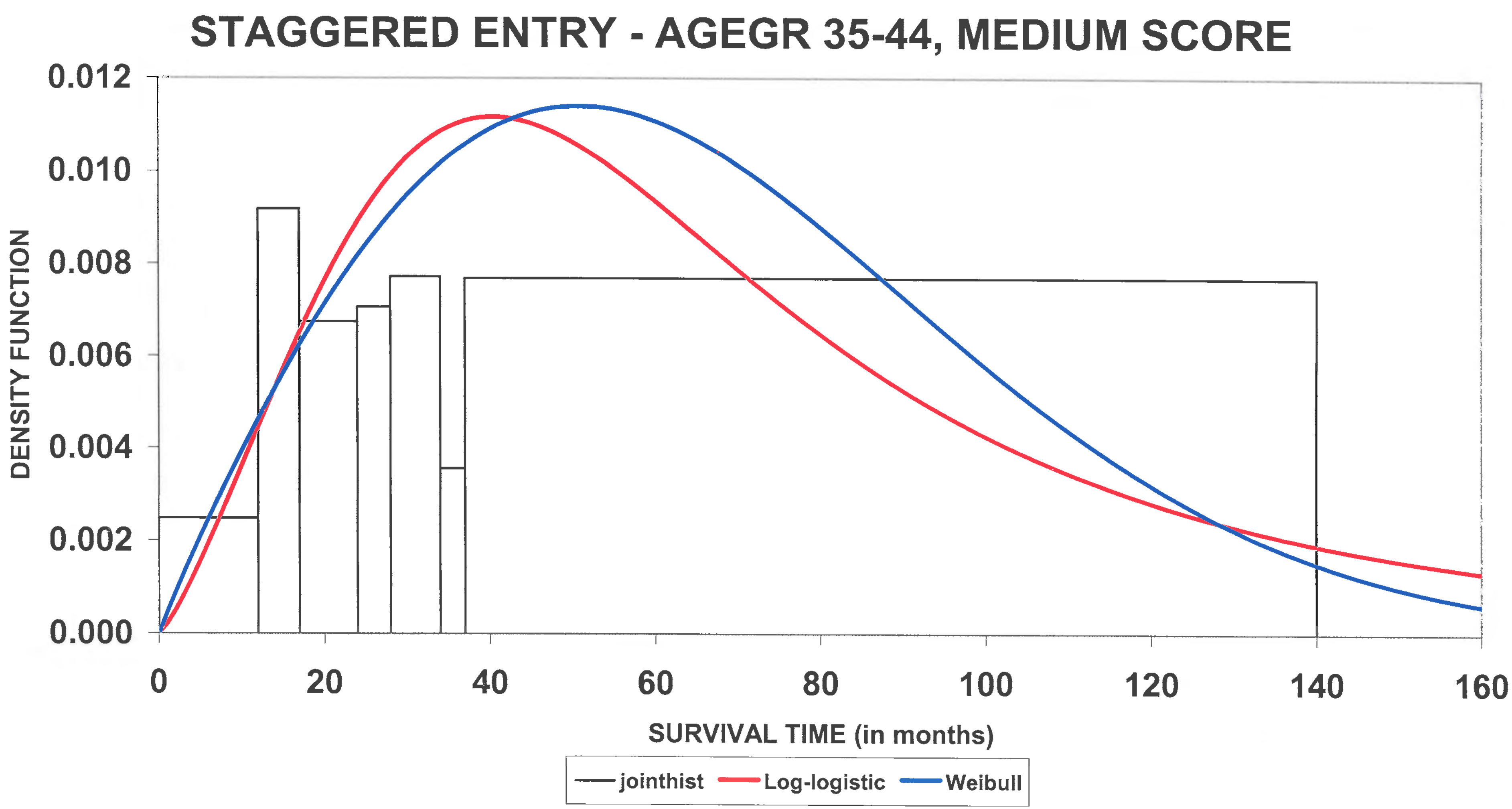

Figure 5.20: **Joint histogram and fitted survival distributions for age group [35;45) and medium score**

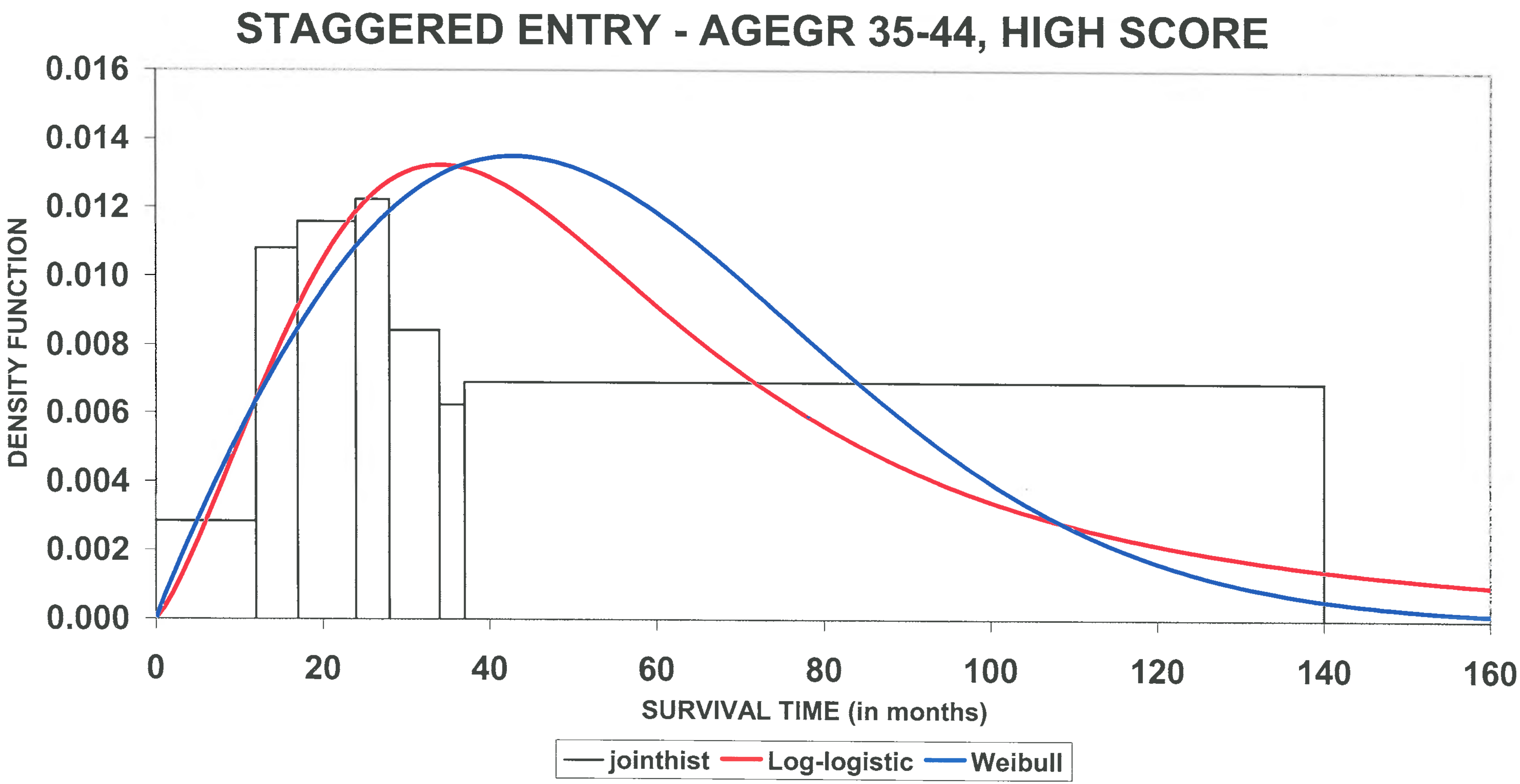

Figure 5.21: Joint histogram and fitted survival distributions for age group [35;45) and high score

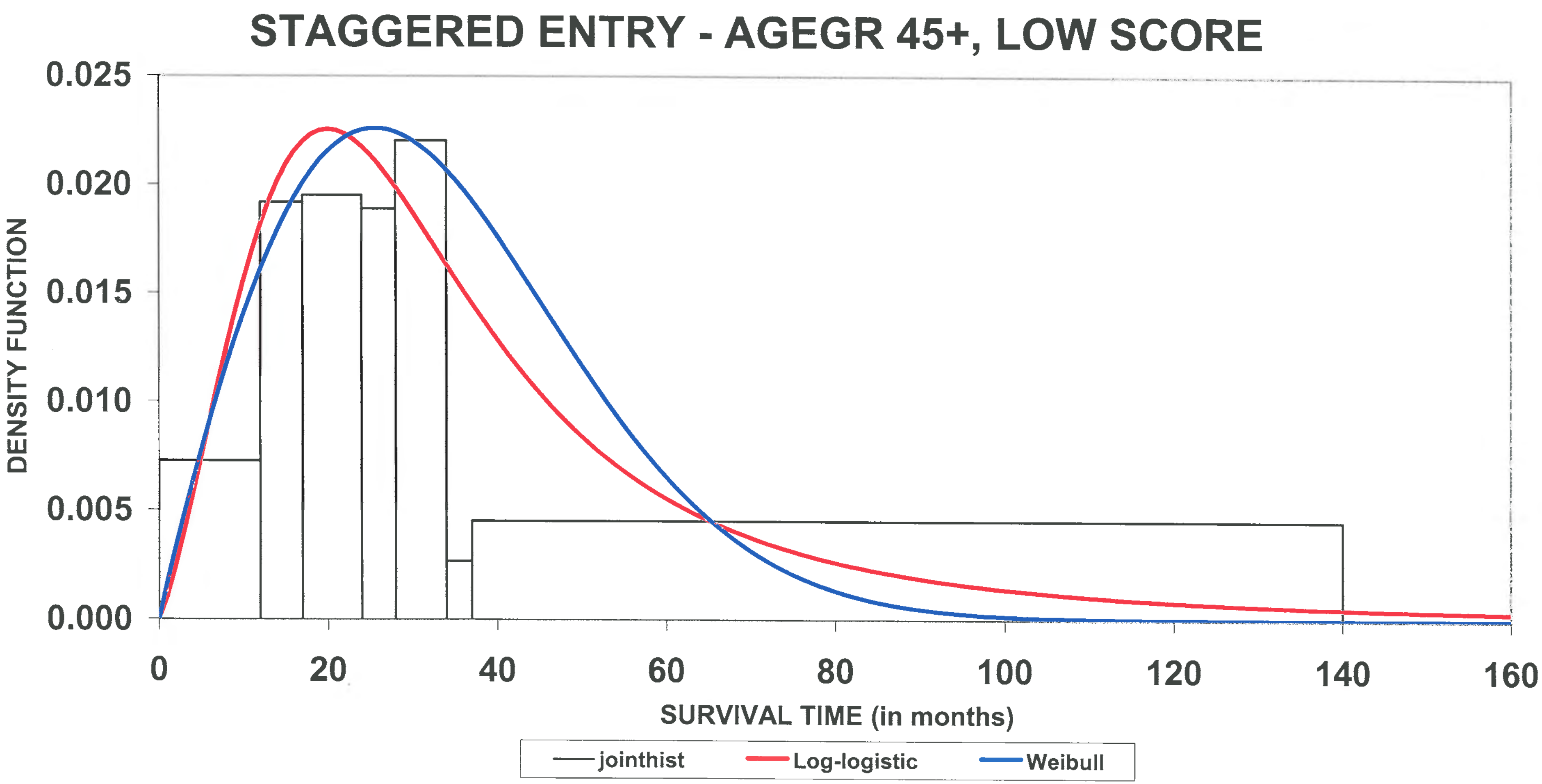

Figure 5.22: **Joint histogram and fitted survival distributions for age group [45+) and low score**

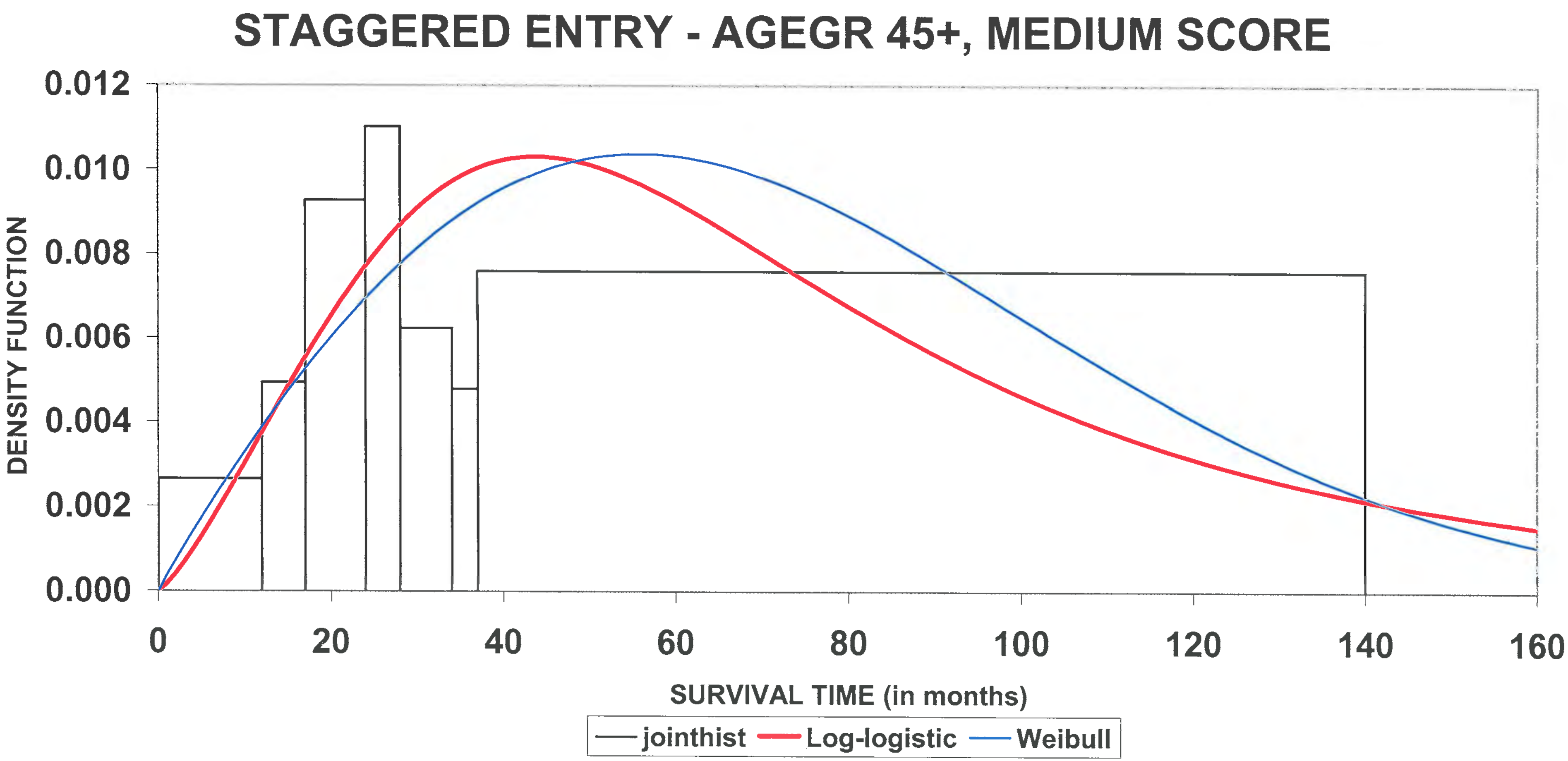

Figure 5.23: **Joint histogram and fitted survival distributions for age group [45+) and medium score**

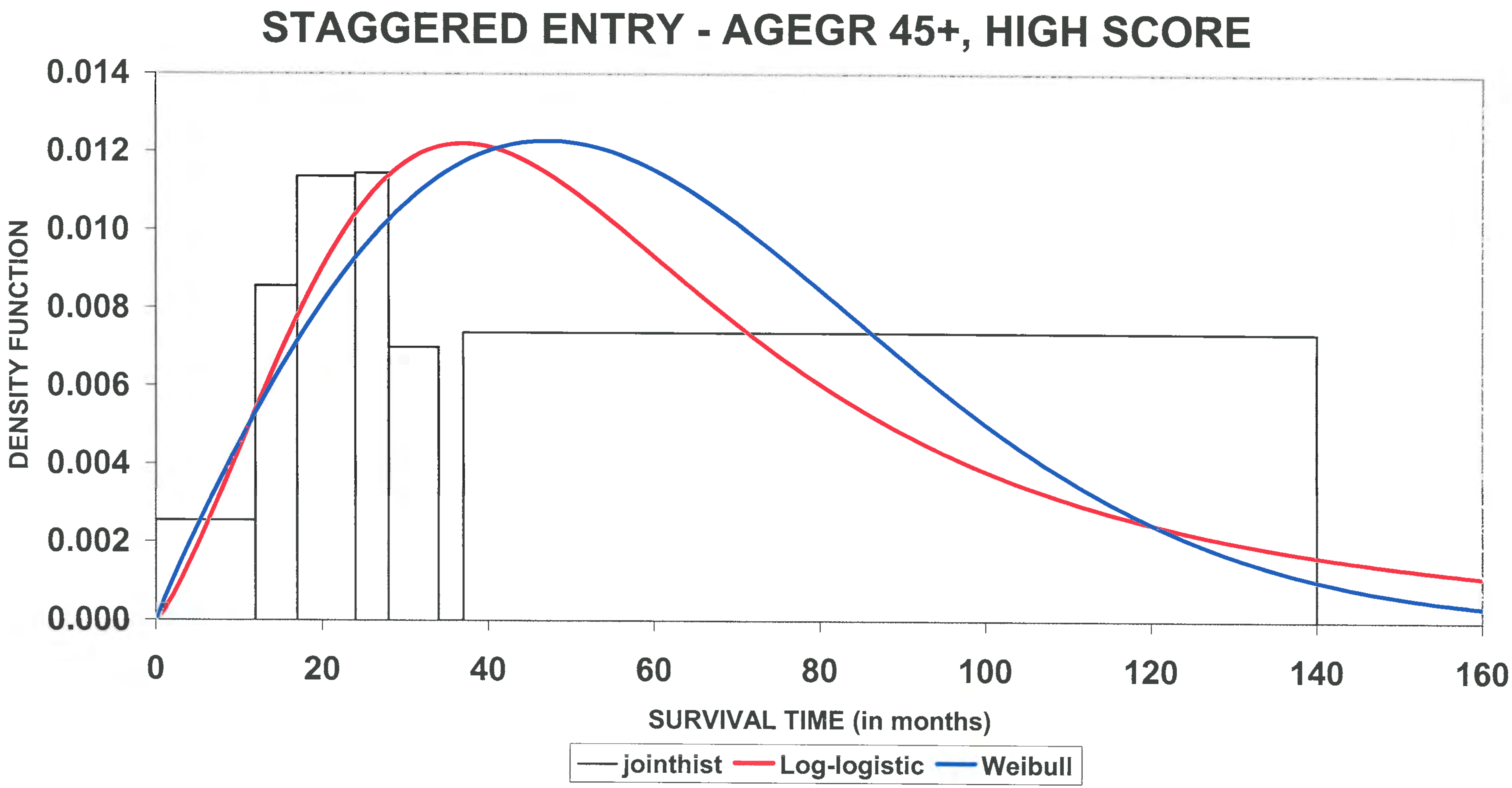

Figure 5.24: **Joint histogram and fitted survival distributions for age group [45+) and high score**

5.3.7 Relationship between the indices of the regression and logit model

Once the parameters of the baseline survival distribution and the nine age-score distributions have been estimated, estimated hazard and survivor functions, odds of a lapse, odds ratios and hazard ratios at time t can be calculated in a similar way as at the regression model with one risk factor (constant shape).

The odds ratio for age group [35;45) and a medium score is the relative odds of a lapse at time t of a policy, where the age of the policyholder is in [35;45) years and the policyholder has a medium score , compared to a policy with the baseline characteristics.

As an example , the odds ratio for a lapse of a policy at time t is calculated if the age of the policyholder is in the age group [18;35) years and the policyholder has a low score.

$$\boxed{\widehat{oddsratio}_{A_1B_1}(t) = \frac{\widehat{odds}_{A_1B_1}(t)}{\widehat{odds}_0(t)}}$$

where

$$\boxed{\widehat{odds}_{A_1B_1}(t) = \frac{1-\widehat{S}_{A_1B_1}(t)}{\widehat{S}_{A_1B_1}(t)} = \widehat{\lambda}_{A_1B_1} \cdot t^{\widehat{\alpha}}}$$

$$\Rightarrow \widehat{odds}_{A_1B_1}(12) = e^{-7.297757} \cdot 12^{2.249510} = 0.181240$$

$$\begin{aligned} \Rightarrow \widehat{oddsratio}_{A_1B_1}(12) &= \frac{0.181240}{0.051768} \\ &= 3.501004 \end{aligned}$$

This odds ratio of 3.5 is called an index and shows the effect of age group [18;35) and a low score on the baseline odds of a lapse at time t. This effect is multiplicative on the baseline odds of a lapse. Thus the effect of the combination of this age group and this score group is to increase the baseline odds of a lapse by a factor 3.5 .

The other eight indices for the log-logistic regression model can be calculated in a similar way.

The relationship between the indices of the nine age-score combinations, obtained from the log-logistic model, must be compared to the six 'indices', obtained from the **loglinear logit model** for the three age levels and the three score levels.

Recall that the loglinear logit model models

$$\ln(\text{odds of a lapse}) = \mu + \lambda_i{}^{AGE} + \lambda_j{}^{SCORE}$$

where

$$\begin{array}{lcl} \mu & = & \text{the overall mean effect, over all AGE levels and SCORE levels} \\ \lambda_i{}^{AGE} & = & \text{effect of the } \; i^{th} \; \text{ level of AGE} \\ \lambda_j{}^{SCORE} & = & \text{effect of the } \; j^{th} \; \text{ level of SCORE.} \end{array}$$

The odds of a lapse then can be modelled as

$$\begin{aligned} \text{odds of lapse} &= e^{\mu+\lambda_i{}^{AGE}+\lambda_j{}^{SCORE}} \\ &= e^{\mu} \cdot e^{\lambda_i{}^{AGE}} \cdot e^{\lambda_j{}^{SCORE}} \\ &= \text{geometric mean odds } \cdot \text{index}_{AGE_i} \cdot \text{index}_{SCORE_j} \end{aligned}$$

i=1,2,3 and j=1,2,3.

The six 'indices' obtained from the logit model for each age level and for each score level are given in Table 5.32.

Table 5.32: **Logit model indices for three age levels and three score levels obtained from the logit model**

Effect	n	Logit model Index t=12	t=24
Baseline odds	10077	0.0537	0.2694
Age [18;35)	3644	1.1558	1.1745
Age [35;45)	3425	0.9844	0.9981
Age [45+),	3008	0.8790	0.8530
Low score	2415	2.2622	2.5756
Medium score	4845	0.5757	0.5234
High score	2817	0.7678	0.7418

The odds of a lapse of a policy in the first year , with the policyholder in the age group [18;35) and a low score, is calculated from the logit model as the product of the baseline odds and the index of age group [18;35) of 1.1558 and the index of score group 'Low' of 2.2622 for the first year (t=12), this means

$$odds_{A_1B_1}(12) = 0.0537 \times 1.1558 \times 2.2622 = 0.1404$$

$$\Rightarrow P(\text{lapse of this policy}) = \frac{odds}{1 + odds} = \frac{0.1404}{1.1404} = 0.1231$$

Thus the odds ratio (relative odds of a lapse of this policy) is calculated by

$$oddsratio_{A_1B_1}(12) = \frac{odds_{A_1B_1}(12)}{baselineodds} = \frac{0.1404}{0.0537} = 2.614525$$

It is clear that this odds ratio can easily be found by multiplication of the two indices from the logit model, that is

$$oddsratio_{A_1B_1}(12) \doteq index_{A1}(12) \times index_{B1}(12) = 1.1558 \times 2.2622 = 2.614651$$

The odds ratio shows the effect of the combination of this age group and this score group on the baseline odds of a lapse. This effect is multiplicative on the baseline odds of a lapse. Thus the effect of the combination of this age group and this score group is to increase the baseline odds of a lapse by a factor 2.614 .

In the context of survival analysis, this odds ratio can be called an **index** for age group [18;35) and a low score. The odds ratios for the nine age-score combinations result in a set of indices, showing the effect of each combination of age group and score on the baseline odds of a lapse at time t.

The odds ratios (indices) of the nine age-score groups, estimated from the log-logistic or Weibull regression model, are compared to the odds ratios, obtained from the logit model, in Table 5.33.

Table 5.31: **Comparison of odds ratios (indices): log-logistic and Weibull regression models and logit model**

Effect	n	Log-logistic regression model					Weibull regression model					Logit model	
		Odds ratio					Odds ratio					Odds ratio	
		t=6	t=12	t=24	t=36	t=60	t=6	t=12	t=24	t=36	t=60	t=12	t=24
Baseline odds	10077	0.01	0.05	0.25	0.62	1.93	0.01	0.06	0.24	0.59	2.5	0.05	0.27
Age [18;35), Low score	833	3.5	3.5	3.5	3.5	3.5	3.08	3.22	3.83	5.25	17.6	2.6	3.0
Age [35;45), Low score	769	2.8	2.8	2.8	2.8	2.8	2.44	2.52	2.84	3.51	7.88	2.2	2.6
Age [45+), Low score	813	2.3	2.3	2.3	2.3	2.3	2.03	2.07	2.25	2.62	4.59	2.0	2.2
Age [18;35), Medium score	1758	0.6	0.6	0.6	0.6	0.6	0.67	0.66	0.65	0.62	0.53	0.7	0.6
Age [35;45), Medium score	1546	0.5	0.5	0.5	0.5	0.5	0.53	0.53	0.51	0.48	0.38	0.6	0.6
Age [45+), Medium score	1541	0.4	0.4	0.4	0.4	0.4	0.44	0.44	0.42	0.39	0.30	0.5	0.5
Age [18;35), High score	1053	0.9	0.9	0.9	0.9	0.9	0.93	0.93	0.92	0.91	0.88	0.9	0.9
Age [35;45), High score	1110	0.7	0.7	0.7	0.7	0.7	0.74	0.74	0.72	0.69	0.61	0.8	0.7
Age [45+), High score	654	0.6	0.6	0.6	0.6	0.6	0.62	0.61	0.59	0.56	0.47	0.7	0.6

It is clear from Table 5.33 that the odds ratios are constant over time at the log-logistic regression model, but the odds ratios do not remain constant over time at the Weibull regression model.

From Table 5.33 follows that one log-logistic regression model provides odds ratios (indices) for any time value, while a new logitmodel has to be built for a fixed time value, say t=12 months, conditional on a restricted experimental design where all the policies must have an exposure of at least one year when investigating the lapses of policies in the first year. There is no such restrictions in the more general experimental design for the log-logistic regression model where all the policies can be used in the analysis, even those policies with inception dates very close to the cut-off point.

The same argument holds for the Weibull regression model, except that the odds ratios do not remain constant over time.

5.3.8 Median lifetimes of the nine survival distributions

The median lifetimes (in months) of the nine survival distributions can also be estimated and compared with the baseline median. The medians are reported in Table 5.34. It is

Table 5.34: **Median lifetimes of the nine survival distributions**

	Regression model	
	Log-logistic	Weibull
Effect	median lifetime	median lifetime
Baseline	44.75	44.19
Age [18;35), Low score	25.64	24.92
Age [35;45), Low score	28.24	28.02
Age [45+), Low score	30.61	30.80
Age [18;35), Medium score	56.13	54.31
Age [35;45), Medium score	61.82	61.08
Age [45+), Medium score	67.02	67.13
Age [18;35), High score	47.36	45.88
Age [35;45), High score	52.16	51.59
Age [45+), High score	56.55	56.70

evident from Table 5.34 that the log-logistic and Weibull models deliver the same results. The estimated median values of the nine combinations of age and score levels suggest that the policy of a policyholder with a low score, coming from any age group, has a high risk to lapse. The policy of a policyholder in agegroup 45+ with a medium score has the lowest risk to lapse, lower than the combination 45+ and a high score.