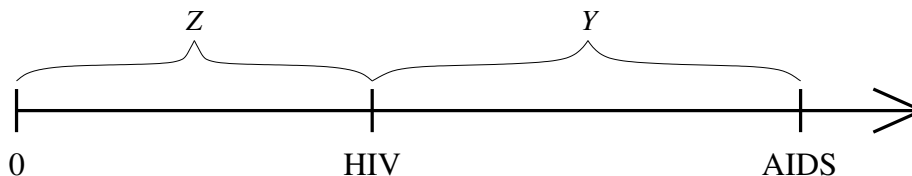


## Chapter 6

# Evaluation of the Data Set on Injecting Drug Users in Badalona

The data set on injecting drug users in Badalona and surroundings has been introduced in Section 2.2: it contains the data of 361 individuals, who have started using intravenous drugs between 1974 and 1997. They have been admitted to the detoxification unit of the Hospital Universitari Germans Trias i Pujol (Hospital Can Ruti) in Badalona from 1987 through 2000. Many of them (266, 73.7%) have been tested HIV-positive. Our main interest focuses on evaluating the possible association between the time from first injecting drug use until HIV infection,  $Z$ , and the subsequent AIDS incubation time,  $Y$ :



**Figure 6.1:** Pattern of disease stages

Given the complex censoring patterns in both  $Y$  and  $Z$  —see Table 2.8 on page 32—, in Chapter 3, we have developed the theoretical background of an accelerated failure time model with an interval-censored covariate. This model and the techniques of simultaneous maximization presented therein, enable us to find a possible answer to the epidemiological issue of interest. This will be shown within the present chapter.

Details of the used model together with the parameter estimates, obtained under the Weibull assumption for the response variable, are presented in Section 6.1. For this first data evaluation, all data are used, whereas in further sections, only the data of the injecting drug users, who started intra-venous drug use after 1985, are considered. This is due to the fact that HIV tests have not been available before 1985, which causes a bias in the results as will be shown in Section 6.2. The accelerated failure time model is then evaluated using both the Weibull and the log logistic

distribution. Furthermore, in Section 6.3 the age and the gender of the individuals are included as covariates in the model. Finally, in Section 6.4, we compare the estimated distribution function of  $Z$  under the used model with the Turnbull estimate, and close the chapter with a tentative goodness-of-fit method.

## 6.1 Use of the accelerated failure time model

The accelerated failure time model applied to the given data set uses the logarithm of the AIDS incubation period as response variable. For this reason, we include the natural logarithm of the time until HIV infection as the model's covariate. This is not necessary, but recommendable as the scales of response and covariate are kept equal. Hence, the log linear expression of the model is given by

$$\ln(Y) = \mu + \beta \ln(Z) + \sigma W. \quad (6.1)$$

The units of both  $Y$  and  $Z$  are months. Hence, the observed intervals  $[Z_l, Z_r]$ , measured from the (known) date of the first HIV exposure by injecting drug use until the dates of the corresponding HIV tests, as well as the subsequent AIDS incubation periods are grouped into units of months. All values are truncated (and not rounded), that is, for example, if the first seropositive observation is registered after  $n$  months and  $x$  days,  $Z_r$  is always equal to  $n$  independently of  $x$ .

It is possible that infection with HIV occurs as a consequence of the first intravenous drug use, that is  $Z = 0$ . Nevertheless, we consider  $Z = 1$  as the covariate's minimum value in order to avoid that  $Z$  can take on a zero value, for which the logarithm is not defined. We assume that this decision will hardly alter the estimation of the parameters.

### 6.1.1 Simultaneous maximization with AMPL

In Section 3.3, the likelihood function of the above model has been derived assuming noninformative censoring and given the following data:  $(U, Z_l, Z_r, \delta_1, \delta_2, \epsilon)$ . The variable  $U$ , according to its definition (3.6), is the elapsed time from  $Z_r$  until AIDS onset being this moment either exactly observed ( $\delta_1 = 1, \delta_2 = 0$ ), right-censored ( $\delta_1 = 0, \delta_2 = 1$ ), or left-censored ( $\delta_1 = \delta_2 = 0$ ). We also consider the 95 individuals with missing information respect to  $Y$ —see Table 2.8—, identified by  $\epsilon = 0$ ; otherwise  $\epsilon$  is equal to 1.

Consequently, given the independent observations  $(U_i, Z_{l_i}, Z_{r_i}, \delta_{1_i}, \delta_{2_i}, \epsilon_i)$ ,  $i = 1, \dots, n$ , the resulting likelihood function is a slightly modified expression of (3.22) on page 48:

$$L(\boldsymbol{\theta}, \boldsymbol{\omega}) = \prod_{i=1}^n \sum_{j=1}^m \alpha_{ij} f(u_i(s_j)|s_j)^{\epsilon_i \delta_{1_i}} S(u_i(s_j)|s_j)^{\epsilon_i \delta_{2_i}} (1 - S(u_i(s_j)|s_j))^{\epsilon_i (1-\delta_{1_i})(1-\delta_{2_i})} \omega_j, \quad (6.2)$$

where  $\boldsymbol{\theta} = (\mu, \beta, \sigma)'$ ,  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_m)'$ , and  $u_i(s_j) = u_i + z_{r_i} - s_j$ . That is, the duration of the AIDS incubation periods depends on the possible times  $s_j$  which have to be considered for a

given interval  $[z_l, z_r]$ .

As pointed out in Section 3.5.2, the maximization of likelihood function (6.2) requires to determine a maximum value of  $Z$ , if right-censored values of  $Z$  are present. This is the case with the given data set. Since the largest finite right-endpoint  $z_r^* = \max\{Z_{r_i} | Z_{r_i} < \infty, i = 1, \dots, n\}$  lies beyond the largest right-censored value  $z_l^* = \max\{Z_{l_i} | Z_{r_i} = \infty, i = 1, \dots, n\} = 202$ , we set  $s_m = z_r^* = 204$ . Any other choice  $s_m > 204$  would not change the estimation results obtained. Note that for the present analysis, we do not consider the possibility that injecting drug users remain seronegative at long term.

As described in Section 3.5, with the help of the mathematical programming language AMPL and the NEOS solver SNOPT, it is possible to maximize the likelihood function (6.2) simultaneously with respect to  $\theta$  and  $\omega$  once the distribution of  $Y$  is specified. Two distributions are considered: the Weibull and the log logistic distribution.

### 6.1.2 The Weibull regression model

In case of the Weibull regression model, the log likelihood function to be maximized is very similar to expression (3.27) adding a fourth summand that corresponds to the individuals with missing information concerning AIDS onset. The small difference lies in the use of the logarithm of the times until HIV infection,  $\ln(s_j)$ :

$$\begin{aligned}
 l(\theta, \omega) = & \sum_{i=1}^n \left[ \epsilon_i \delta_{1_i} \ln \left( \sum_{j=1}^m \frac{\alpha_{ij}}{\sigma} \exp \left( \frac{\ln(u_i(s_j)) - \mu - \beta \ln(s_j)}{\sigma} \right) - e^{\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta \ln(s_j))} \right) \omega_j \right) \\
 & + \epsilon_i \delta_{2_i} \ln \left( \sum_{j=1}^m \alpha_{ij} \exp \left( - e^{\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta \ln(s_j))} \right) \omega_j \right) \\
 & + \epsilon_i (1 - \delta_{1_i})(1 - \delta_{2_i}) \ln \left( \sum_{j=1}^m \alpha_{ij} \left( 1 - \exp \left( - e^{\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta \ln(s_j))} \right) \right) \omega_j \right) \\
 & + (1 - \epsilon_i) \ln \left( \sum_{j=1}^m \alpha_{ij} \omega_j \right) \right],
 \end{aligned} \tag{6.3}$$

where  $n = 361$  and  $m = 204$ . Hence, the objective function (6.3) comprises more than 200 variables — $\mu, \beta, \sigma$  as well as  $\omega$ —, and its maximization is subject to the constraints:

$$\begin{aligned}
 \sigma &> 0, \\
 \omega_j &\geq 0, \quad j = 1, \dots, m, \\
 \sum_{j=1}^m \omega_j &= 1.
 \end{aligned} \tag{6.4}$$

The AMPL script, that solves this optimization problem, consists of the three files attached in

Section C.1.1 on page 129. It invokes the NEOS solver SNOPT<sup>1</sup> (Gill, Murray, and Saunders 1999), which carries out simultaneous maximization with respect to all variables in less than 25 seconds on a Pentium III PC with 871 MHz. In a second step, the 95% confidence intervals of  $\theta$  are calculated using the MAPLE script on page 137.

The obtained estimation results are summarized in Table 6.1a, where ‘RR’ denotes relative risk and ‘AF’ stands for acceleration factor; both are estimated according to (3.33). For the computation of the confidence interval of the relative risk, we use the fact that the estimator,  $\widehat{RR} = \exp(-\hat{\beta}/\hat{\sigma})$ , is asymptotically normal distributed and derive its asymptotic variance (B.5) applying the delta method as shown in Section B.2 on page 125. In contrast with that, the confidence interval of the accelerating factor is a mere transformation of the confidence interval of  $\beta$ , given that  $AF = \exp(\beta)$ .

**Table 6.1a:** Estimation results and 95% confidence intervals for model (6.1)

$\hat{\mu}$	$\hat{\beta}$	$\hat{\sigma}$	$\widehat{RR}$	$\widehat{AF}$
4.822	0.137	0.475	0.749	1.147
[4.795, 4.848]	[0.129, 0.146]	[0.469, 0.481]	[0.739, 0.759]	[1.138, 1.157]

We see that  $\hat{\beta}$  is positive and that the corresponding confidence interval does not include zero. That implies that, on average, the longer an injecting drug user remains seronegative, the longer s/he remains AIDS-free once s/he is infected with HIV. The magnitude of this effect is reflected by the relative risk and the acceleration factor: increasing the logarithm of the time from first intravenous drug use until HIV infection by 1, the risk of AIDS onset decreases 25% and the median AIDS incubation time increases by the factor 1.15. These conclusions, based on model (6.1), are valid at a 95% significance level.

In some cases, the date of the first positive HIV test result nearly coincides with the date of their AIDS diagnosis. In part, this is due to the fact that these individuals have been unaware of their HIV status for a long time. Thus, when admitted to the unit, they have had a positive HIV test result followed by an AIDS diagnosis within a few days. Others could not be tested for HIV during many years, as the specific HIV tests have not been available before 1985. Since the probability of developing AIDS during the first two years after seroconversion is rather low (about 1% according to Brookmeyer and Gail (1994)), we have adjusted the value  $Z_r$  for minimum AIDS incubation periods of one and two years. For this purpose, the values of  $Z_r$  of 21 and 32 individuals, respectively, have been changed. The estimation results summarized in Table 6.1b show very little difference from the ones in the table above. For this reason, this data adjustment is not considered further on.

---

<sup>1</sup>Version 6.2 by June 2004

**Table 6.1b:** Estimation results and 95% confidence intervals for model (6.1)

Model	$\hat{\mu}$	$\hat{\beta}$	$\hat{\sigma}$	$\widehat{RR}$	$\widehat{AF}$
(6.1a) <sup>a</sup>	4.846	0.132	0.461	0.751	1.141
	[4.824, 4.867]	[0.126, 0.139]	[0.456, 0.467]	[0.743, 0.759]	[1.134, 1.149]
(6.1b) <sup>b</sup>	4.839	0.135	0.45	0.741	1.144
	[4.819, 4.859]	[0.129, 0.141]	[0.445, 0.456]	[0.734, 0.749]	[1.138, 1.151]

<sup>a</sup> assuming a minimum AIDS incubation period of 12 months

<sup>b</sup> assuming a minimum AIDS incubation period of 24 months

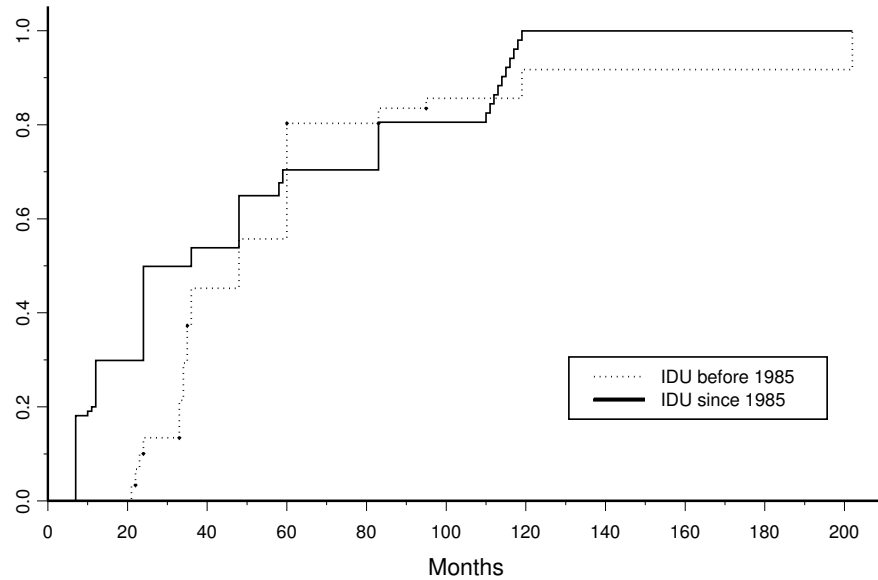
## 6.2 Stratification according to start of injecting drug use

As mentioned before, routine HIV tests have not been available before 1985. This fact might be a source of bias since about half of the 361 subjects of the study have started using intravenous drugs before that year as illustrated in Figure 2.2 on page 29. For example, injecting drugs users of the early eighties could be tested positively by an HIV test not before 60 months. Even though the applied methodology considers all possible values of  $Z$  given  $[Z_l, Z_r]$ , an introduced bias cannot be ruled out.

For this reason, the distribution function of time from first injecting drug use until HIV infection has been estimated by means of the Turnbull estimator separately for individuals starting intravenous drug use before and since 1985. The result is illustrated in the following Figure 6.2. The plot shows that the minimum value within the group of former injecting drug users is 21 months, but only six months in the other group. Also, the observed maxima differ substantially (202 versus 119 months). It is very probable that these differences are due to the fact that short times until HIV infection have not been identifiable before 1985.

To see, whether these differences have any influence on the estimation of the model parameters, model (6.1) has been adjusted separately for both groups. The results are shown in Table 6.2. According to the results obtained, the conclusions above are confirmed for the group of injecting drug users since 1985. The individuals of this group have had the possibility of immediate HIV tests; therefore, a systematic bias due to the lack of HIV tests before 1985 can be excluded. In contrast with that, within the group of former injecting drug users, hardly any effect of  $Z$  on the AIDS incubation periods can be observed ( $\hat{\beta} = 0.017$ ) and the model constant is even bigger than 5. This observation is most probably due to the possible bias mentioned.

For this reason, in the sequel, all data analysis are based on the reduced data set of 176 injecting drug users, who have started using intra-venous drugs after 1985.



**Figure 6.2:** Nonparametric estimation of  $F_Z$  by start of intravenous drug use

**Table 6.2:** Estimation results for Model (6.1) by start of injecting drug use

Period	$\hat{\mu}$	$\hat{\beta}$	$\hat{\sigma}$	$\widehat{RR}$	$\widehat{AF}$
Before 1985 ( $n = 185$ )	5.228 [5.19, 5.265]	0.017 [0.004, 0.031]	0.439 [0.429, 0.448]	0.961 [0.941, 0.981]	1.018 [1.004, 1.031]
Since 1985 ( $n = 176$ )	4.374 [4.318, 4.431]	0.266 [0.246, 0.286]	0.47 [0.454, 0.485]	0.568 [0.55, 0.586]	1.304 [1.278, 1.331]

### 6.2.1 The log logistic regression model

With the log logistic regression model, the log likelihood to be maximized has the following expression (6.5). It is similar to log likelihood (3.32) on page 54 accounting for subjects with missing information on AIDS onset and the fact that the model's covariate is the natural logarithm of time until HIV infection:

$$\begin{aligned}
l(\boldsymbol{\theta}, \boldsymbol{\omega}) = & \sum_{i=1}^n \left[ \epsilon_i \delta_{1_i} \ln \left( \sum_{j=1}^m \frac{\alpha_{ij}}{\sigma} \frac{\exp(\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta \ln(s_j)))}{(1 + \exp(\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta \ln(s_j))))^2} \omega_j \right) \right. \\
& + \epsilon_i \delta_{2_i} \ln \left( \sum_{j=1}^m \frac{\alpha_{ij}}{1 + \exp(\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta \ln(s_j)))} \omega_j \right) \\
& \left. + \epsilon_i (1 - \delta_{1_i})(1 - \delta_{2_i}) \ln \left( \sum_{j=1}^m \alpha_{ij} \left( 1 - \frac{1}{1 + \exp(\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta \ln(s_j)))} \right) \omega_j \right) \right] \\
& + (1 - \epsilon_i) \ln \left( \sum_{j=1}^m \alpha_{ij} \omega_j \right) \Big].
\end{aligned} \tag{6.5}$$

The estimation procedure applied to maximize (6.5) is the same as with the Weibull regression model: the NEOS solver SNOPT is used to carry out simultaneous maximization and the confidence intervals of  $\mu$ ,  $\beta$ , and  $\sigma$  are obtained using MAPLE. The corresponding results are summarized in Table 6.3 below, in which ‘OR’ denotes the odds ratio.

**Table 6.3:** Estimation results assuming a log logistic distribution

$\hat{\mu}$	$\hat{\beta}$	$\hat{\sigma}$	$\widehat{\text{OR}}$	$\widehat{\text{AF}}$
4.132	0.28	0.396	0.494	1.323
[4.064, 4.20]	[0.256, 0.303]	[0.383, 0.41]	[0.473, 0.515]	[1.292, 1.354]

As with the Weibull model, the positive value of  $\hat{\beta}$ , significantly at a 95% level, implies an association between times until HIV infection and the AIDS incubation periods. The relatively high value of  $\hat{\mu}$  reflects the fact that, independently of  $Z$ , the median AIDS incubation is equal to, at least, five years as shown in Table 6.4.

This Table 6.4 shows an illustration of the obtained results for both model assumptions in terms of the median AIDS incubation times given different values of  $Z$  and based on the estimated parameter values. For all chosen values of  $Z$ , the estimated median with the Weibull choice lies about five months above the values of the log logistic model. For each increase of the logarithm of  $Z$  approximately one, the median times until AIDS development are increased by factors roughly equal to the acceleration factors of 1.304 and 1.323, respectively.

**Table 6.4:** Illustration of estimation results

$Z$	$\ln(Z)$	$\text{Med}(Y Z)$	
		Weibull	Log logistic
1	0	66.8	62.3
3	1.099	89.5	84.7
8	2.079	116.1	111.4
22	3.091	152.0	147.7

The conditional median times have been calculated according to the following formulas:

$$\text{Med}(Y|Z) = \begin{cases} \exp(\hat{\mu} + \hat{\beta} \ln(Z)) \ln(2)^{\hat{\sigma}} & \text{Weibull model,} \\ \exp(\hat{\mu} + \hat{\beta} \ln(Z)) & \text{Log logistic model.} \end{cases}$$

### 6.3 Inclusion of further covariates

Two more covariates have been considered in the present analysis: the gender and the age of the individuals. Including these variables into model (6.1), the log linear expression of the extended model is given by

$$\ln(Y) = \mu + \beta \ln(Z) + \gamma X_1 + \kappa X_2 + \sigma W, \quad (6.6)$$

where  $X_1$  denotes the gender of the individuals (0: male, 1: female), and  $X_2$  their age at first injecting drug use (0: < 20 years, 1:  $\geq$  20 years). As mentioned previously, only the data of the 176 injecting drug users since 1985 are considered. Adjusting model (6.6) to these data assuming a Weibull distributed response, the estimates summarized in Table 6.5 are obtained.

**Table 6.5:** Estimation results and 95% confidence intervals for model (6.6)

$\hat{\mu}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\kappa}$	$\hat{\sigma}$
4.478	0.278	0.249	-0.409	0.432
[4.401, 4.555]	[0.253, 0.304]	[0.209, 0.289]	[-0.439, -0.378]	[0.417, 0.447]

Concerning gender, the positive value of  $\hat{\gamma}$  implies a lower risk of developing AIDS and, hence, longer AIDS incubation periods, for female injecting drug users compared with males given the same values in the two other covariates. In case of the dichotomized covariate age, we observe that older HIV-infected injecting drug users are at higher risk of developing AIDS. Also with this model, adjusting for gender and age, the value  $\hat{\beta}$  is significantly larger than zero, that is the same conclusions hold respect to time until HIV infection and the subsequent AIDS incubation periods.

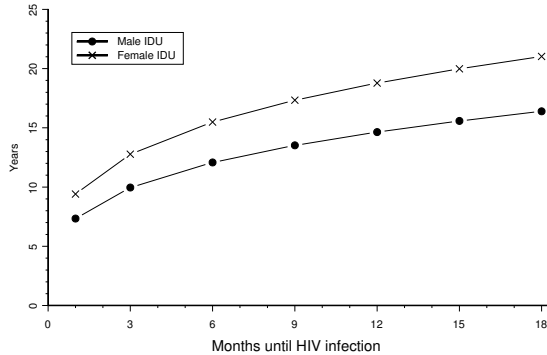
A summary of these estimates in terms of relative risks and acceleration factors is given in the Table 6.6: adjusting for the other covariates, male injecting drug users are nearly at double risk of developing AIDS, older ones even at a higher risk compared with female and younger injecting drug users, respectively.

Finally, the conditional median AIDS incubation periods based on the parameter estimates for several values of  $Z$  are illustrated in Figures 6.3(a) and 6.3(b). The left figure shows the estimated median times for male and female injecting drug users given an age less than 20 at first intravenous drug use, the right figure the corresponding estimates for individuals older than 20 years. In both figures, the curves of the women lies above the ones of the men.

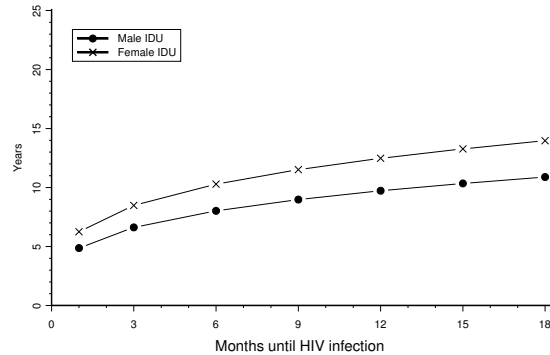


**Table 6.6:** Relative risks and acceleration factors for model (6.6)

Time till HIV infection		Gender		Age	
$\widehat{RR}$	$\widehat{AF}$	$\widehat{RR}$	$\widehat{AF}$	$\widehat{RR}$	$\widehat{AF}$
0.525	1.32	0.565	1.282	2.578	0.664
[0.503, 0.547]	[1.288, 1.355]	[0.526, 0.598]	[1.233, 1.335]	[2.455, 2.695]	[0.645, 0.685]



(a) Young IDU



(b) Old IDU

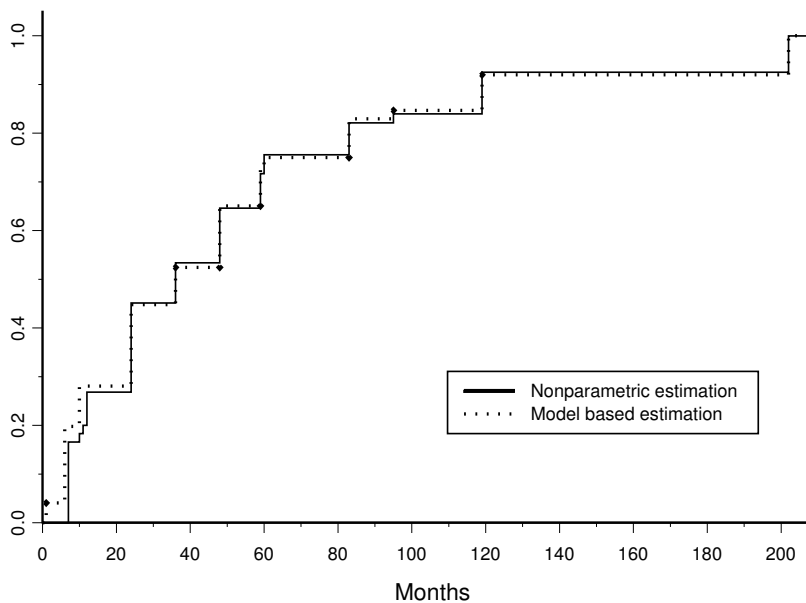
**Figure 6.3:** Estimated median AIDS incubation periods based on model (6.6)

## 6.4 Estimation of the covariate's distribution function

Besides the parameter estimates, the simultaneous maximization of the respective likelihood functions has furnished an estimation of  $\omega$ , equivalent to the estimation of  $F_Z$ , the distribution function of the covariate. This estimate based on model (6.1) under the Weibull assumption has been compared with the nonparametric maximum likelihood estimator (NPMLE) determined by the Turnbull estimator.

A visual comparison of both estimates is shown in Figure 6.4, that shows just slight differences between both estimates. The model based estimate identifies two more values of  $Z$ , on which positive probability mass is put, namely  $Z = 1$  and  $Z = 10$ . Apart from that, both curves differ in less than 0.01. Hence, the empirical quantiles for any  $p \geq 0.25$  are the same.

Very similar observations have been made when comparing the NPMLE  $\widehat{F}_Z$  with the model based estimates of the covariate's distribution function under the log logistic assumption or in case of model (6.6). That is, substantial differences between  $\widehat{\omega}$  based on either model (6.1) or model (6.6) have not been not detected.



**Figure 6.4:** Estimation of distribution function of time until HIV infection

## 6.5 Tentative goodness-of-fit

Herein, we present a possible plot to judge the goodness-of-fit of model (6.1) that is based on the Cox-Snell residuals of parametric survival models (Klein and Moeschberger 1997, Chap. 12).

### 6.5.1 Cox-Snell residuals in the accelerated failure time model

The Cox-Snell residuals of the accelerated failure time model provide a check of the overall fit of the model. They are defined as the estimated cumulative hazard function of  $Y$  given  $Z$ :  $r_i = \hat{\Lambda}(y_i|z_i)$ . If the model fits well to the data, these residuals follow approximately an exponential distribution with parameter equal to 1. With the Weibull and the log logistic models, the Cox-Snell residuals are, hence, given by:

$$r_i = \begin{cases} \exp\left(\frac{\ln(y_i) - \hat{\mu} - \hat{\beta}\ln(z_i)}{\hat{\sigma}}\right) & \text{Weibull model,} \\ \ln\left(1 + \exp\left(\frac{\ln(y_i) - \hat{\mu} - \hat{\beta}\ln(z_i)}{\hat{\sigma}}\right)\right) & \text{Log logistic model.} \end{cases} \quad (6.7)$$

Plotting the residuals corresponding to the uncensored survival times versus the Nelson-Aalen estimator of their cumulative hazard function results roughly in a straight line with slope 1 if the model fit is appropriate.

### 6.5.2 Application to the data set on injecting drugs users

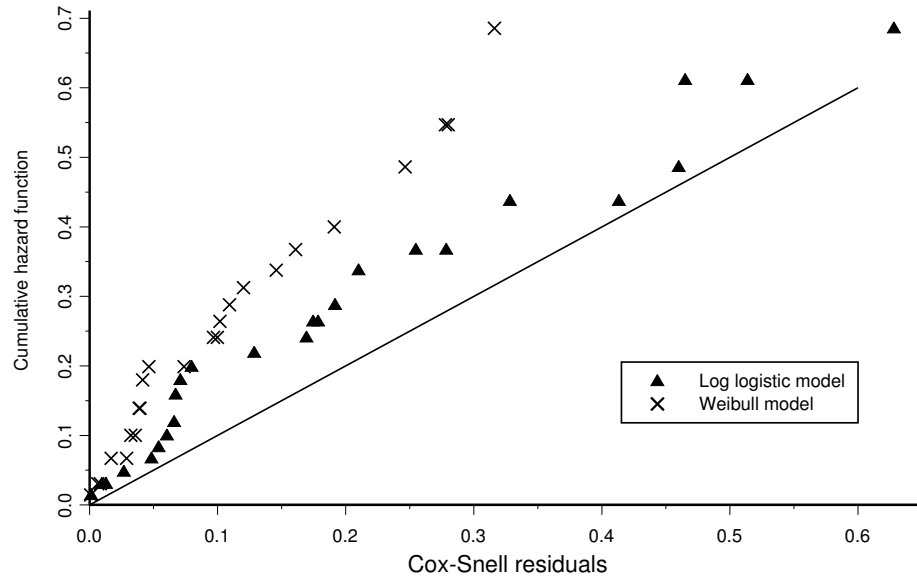
In case of the data set on injecting drug users, the Cox-Snell residuals (6.7) cannot be calculated given the fact that the covariate  $Z$  is not observed exactly and that the response variable is the subsequent time to  $Z$ . Hence, our proposal is to use a modification of the Cox-Snell residuals based on imputed values for  $Z$ . For the imputation, we choose the conditional mean of  $Z$  given  $[Z_l, Z_r]$  and the Turnbull estimate of  $\omega$ :

$$\hat{Z}_i = E_{\hat{F}_Z}(Z_i | Z_{l_i}, Z_{r_i}) = \sum_{j=1}^m \alpha_{ij} s_j \hat{\omega}_j / \sum_{l=1}^m \alpha_{il} \hat{\omega}_l, \quad i = 1, \dots, n. \quad (6.8)$$

Given these imputed values of time until HIV infection, AIDS incubation periods are easily estimated as the time from  $\hat{Z}_i$  until the observed value of AIDS diagnosis, regardless of whether this time is exactly observed, right-, or left-censored. With these values, the expression of the modified Cox-Snell residuals is the same as in (6.7) replacing  $z_i$  and  $y_i$  by its respective estimates based on the imputation (6.8).

For the plot of the residuals versus the cumulative hazard function, we apply the Turnbull estimator to the residuals, from which the cumulative hazard function is obtained by using the formula:  $\hat{\Lambda} = -\log(\hat{S})$ . This is due to the fact that left-censored AIDS incubation times have been present. The resulting plot for model (6.1) containing the residuals of both the Weibull and log logistic assumption is shown in Figure 6.5. We can see, that neither of the two shows a straight line.

This figure is a first attempt to judge the goodness-of-fit of a accelerated failure time model with a doubly censored response variable and an interval-censored covariate. However, more work is needed to check whether the used modification of the Cox-Snell residuals really behave as these residuals in common parametric survival models. If that was the case, Figure 6.5 would indicate a poor fit of the model. The introduced variability by imputing  $\hat{Z}$  and estimating  $\hat{Y}$  given  $\hat{Z}$  must most probably be taken into account. The residuals presented by Topp and Gómez (2004) for linear regression models with interval-censored covariates could be an alternative. This topic and some further aspects of interest concerning the evaluation of the data on injecting drugs users will be briefly addressed in Section 8.1.



**Figure 6.5:** Adapted Cox-Snell residuals of model (6.1) with data from IDU since 1985