Statistical modelling of repeated and multivariate survival data

Wintrebert, C.M.A.

CHAPTER 5

Estimation of the Correlation Between Processes With Frailties: Cardiac, Cerebral and Peripheral Atherosclerosis

Abstract

This chapter concerns and models data from patients having suffered from atherosclerosis. Atherosclerosis is often thought to be a systemic disease, which may appear in coronary-, in cerebral-, or in peripheral-vascular areas.
In our data, patients entered into the study after having had one of the three types of atherosclerotic events. These patients have been followed and the number and the type of new atherosclerotic events were registered. The important point of this chapter is the assumption that the processes leading to coronary, cerebral and peripheral vascular atherosclerosis are different. Our aim is to estimate the correlations between the three processes. In order to achieve this aim, patients who are at risk of having a second or later atherosclerotic event are needed. We based our study leading to this chapter on such patients coming from the Caprie-trial. To be able to estimate these correlations different frailty models were developed and applied. We began with the standard shared frailty models and followed with more complex multivariate frailty models.
In the Caprie-trial patients were included after they experienced an atherosclerotic event, and we therefore only considered the conditional likelihood of recurrent events conditional on age and type of the first event.
We found that a model with separate, but correlated, frailty-parameters for the three different atherosclerotic processes yielded a higher log likelihood than a model with only a shared frailty parameter, but the difference was not statistically significant.

This chapter has been submitted for publication as: C. M. A. Wintrebert, A. H. Zwinderman and J. C. van Houwelingen . Estimation of the Correlation Between Processes With Frailties: Cardiac, Cerebral and Peripheral Atherosclerosis.

5.1 Introduction

Atherosclerosis is often thought to be a systemic disease (Brand et al., 1998), which may appear in cardiac-, in cerebral-, or in peripheral-vascular areas. It is not well understood why it manifests itself in one of these areas in some persons, and in other areas in other persons. The risk profiles of patients with cardiovascular, cerebral vascular or peripheral vascular atherosclerosis are similar (Nicoloff et al., 2002), and therefore it might indeed very well be one singular process becoming manifest at a random location. The risk profiles of the three forms of vascular atherosclerosis are however (Roeters van Lennep et al., 2001) not exactly the same. Coronary vascular atherosclerosis presents itself -on average- at younger ages than cerebral or peripheral vascular atherosclerosis (Ascione et al., 2002; Hankey, 2003; Hutter et al., 2004; Nikolsky et al., 2004; Olijhoek et al., 2004). There is also a systematic difference between men and women with respect to prevalence and incidence of coronary, cerebral, and peripheral vascular atherosclerosis (Roeters van Lennep et al., 2001; Yusuf et al., 2004). It might therefore be useful to distinguish (at least) three different atherosclerotic processes, which are very likely correlated.

The starting point of the present chapter is therefore the assumption that the processes leading to coronary, cerebral, and peripheral vascular atherosclerosis are different, and the aim is to estimate the correlations between these three processes. If we would find these correlations to be very high, we may conclude that the three processes fall together.

The correlations between the three processes can be estimated in patients who are at risk for a second atherosclerotic event. Such patients were included in the Caprie-trial (CAPRIE Steering Committee, 1996); Caprie was a clinical trial evaluating efficacy of clopidogrel versus aspirin for reducing the risk of a recurrent atherosclerotic event.

Three strata were included in the trial: patients with a coronary vascular event, patients with a cerebral vascular event, and patients with a peripheral vascular event.

All patients were followed for on average 6.1 years and all coronary, cerebral, and peripheral vascular events were noted. The fact that the trial included all three types of events, and followed all patients not only for recurrence of the index event, but also for occurrence of other vascular events, allows estimation of the correlations between the three atherosclerotic processes.

Below we first describe the design and the data of the Caprie-trial, next we describe the statistical model used to estimate the correlations, and finally we present our results and concluding remarks.

5.2 Data

We considered data from 307 Dutch patients who were included in the Caprie trial (CAPRIE Steering Committee, 1996). These patients were included because they had

experienced either a cardiac atherosclerotic event (myocardial infarction $n = 97$), or a cerebral atherosclerotic event (stroke/transient ischemic attack (TIA) $n = 103$), or a peripheral atherosclerotic event (angioplasty, bypass surgery or amputation $n = 107$).

The data set included 196 men and 111 women and the age at index event varied between 30 and 89 (mean 63, standard deviation 12). Patients were followed for repeated events of any type, and follow-up varied between 4.9 and 16.6 years.

HDL cholesterol at baseline was 6.2 mmol/L (standard deviation SD 1.1), and was slightly lower in patients with cardiac events 1.09 (SD=0.30) versus 1.22 (SD=0.35): $p = 0.0005$. Among patients with peripheral events there were more smokers (50%) than among patients with cardiac (37%) or cerebral (35%) events ($p = 0.047$). Hypertension occurred far more often among patients with cerebral events (58%) than among patients with cardiac (36%) or peripheral events (30%) ($p < 0.001$).

5.3 Model

We consider three types of events: cerebrovascular (CVA: type 1), cardiac (MI: type 2) and peripheral (PAD: type 3). Let j be the type of event, $j = 1,2,3$. Consequently we will work with three strata, each stratum corresponding to one of the three types of index events.

Let T_k^j be the age of a patient when the k^{th} event of event type j occurs. For example: T_1^1 is the age of a patient at the first CVA event. It is important to note that in this chapter the time to event is the age of a patient when an event occurs and that we do not reset the clock.

The hazard of the k^{th} event type j at age T_k^j given age at first event of type j is denoted as $\lambda^j(t_k^j|Z_j)$ and is specified as a function of a baseline hazard, covariate effects and a frailty parameter (Hougaard, 2000):

$$\lambda^j(t_k^j|Z_j) = \lambda_0^j(t_k^j)\exp(Z_j + \beta_j X), \tag{5.1}$$

where X is a vector of covariate values, β_j the corresponding regression parameter, $\lambda_0^j(t_k^j)$ the baseline hazard, and Z_j the frailty associated with the event type j.

The frailty parameter Z_j represents the extent to which an individual is at risk for the associated event-type, and as such represents the specific process causing event-type j. This parameter cannot be observed, but must be inferred from the data of each individual.

It is essential to our aim to estimate correlations between the three atherosclerotic processes. We assumed that the three frailty parameters (Z_1, Z_2, Z_3) followed a normal distribution with zero mean and unspecified covariance matrix Σ. We will estimate the elements of this covariance matrix using the Caprie-data, and then transform it into a correlation matrix.

The central question is whether the three correlations equal unity, or are very close to unity. The first complication is that all patients experienced the index-event, defining the strata in the Caprie-trial, but only a minority had a second (or third) event during follow-up. This means that the failure time of the second event is censored for most patients.

A second complication concerns the fact that only patients with atherosclerotic events were included in the Caprie-trial. Thus patients were ascertained on the basis of having one event, excluding all individuals who were at serious risk for one or more events, but did not yet experience one. In the following we describe how we handled the censored observations, and the ascertainment issue, and how we estimated the parameters in the model given in equation (5.1).

We will describe the data as there being three processes for each individual studied. It will be assumed that the patients are independent. For events of type j the times of events for a patient are ordered: $0 < T_1^j < T_2^j < ....$ The number of events of type j for patient i is denoted as K_{ji}, and event times are observed except for the last which may be censored. Let d_k^j be the censoring indicator of the k^{th} event of type j of a patient.

The likelihood of the three series of event times in an individual who was ascertained with an event of type 1 at age t_1^1 is denoted as follows:

$$L_i^1 = Pr(T_2^1, ..., T_{K_{1i}}^1, T_1^2, ..., T_{K_{2i}}^2, T_1^3, ..., T_{K_{3i}}^3 \mid T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1) \qquad (5.2)$$

The total likelihood is obtained as a product over all the individuals.

We also considered the model in which $Z_{i1} = Z_{i2} = Z_{i3} = Z_i$ this is a single or shared-frailty model, and consequently we assume independence of all events of all types given this frailty. To study how the frailty distribution depends on $T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1$, we define first $U_i = \exp(Z_i)$ and then assume for convenience sake that this frailty U_i follows a gamma distribution gamma (δ, δ) such that the expectation equals one and the variance equals $1/\delta$.

Let U_i be the frailty of patient i; ascertainment of patient i at event of type 1 at age t_1^1 entails that the posterior distribution of U_i given $(T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1)$ is again a gamma distribution with expectation and variance as follows:

$$E[U_i \mid T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1] = \frac{\delta + 1}{\delta + \sum_{j=1}^3 \Lambda^j(t_1^1)}$$

$$var[U_i \mid T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1] = \frac{\delta + 1}{(\delta + \sum_{j=1}^3 \Lambda^j(t_1^1))^2},$$

where $\Lambda^j(t_1^1)$ is the cumulative hazard function of event type j at $T_j = t_1^1$. It becomes clear that when the first event occurs at an early age then $\Lambda^j(t_1^1)$ will be small, and thus the posterior expectation $E[U_i \mid T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1]$ will be larger than 1,

indicating that the patient is relatively more frail, and the reverse is the case when the first event occurs at a late age.

A model with a single frailty parameters induces constant correlation between events of the same type but also between events of different type. Our aim is to estimate the correlation between different event types and therefore we need three different frailty parameters. We will drop the assumed gamma distribution, because it is difficult to generalize it to three dimensions.

Let $g(Z_1, Z_2, Z_3 \mid T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1)$ be the posterior distribution of Z_1, Z_2, Z_3 given the ascertainment of individual i at age t_1^1 with event of type 1. Using Bayes rule, this is easily derived as

$$g(Z_1, Z_2, Z_3 \mid T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1) = \frac{Pr(T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1 | Z_1, Z_2, Z_3) f(Z_1, Z_2, Z_3)}{\int\int\int Pr(T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1 | Z_1, Z_2, Z_3) f(Z_1, Z_2, Z_3) dZ_1 dZ_2 dZ_3}, \tag{5.3}$$

where $f(Z_1, Z_2, Z_3)$ is the density function of the trivariate normal distribution with mean zero and covariance matrix Σ. Given (Z_1, Z_2, Z_3) the event times are independent, and therefore

$$Pr(T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1 | Z_1, Z_2, Z_3) = \lambda^1(t_1^1 | Z_1) \exp(-\sum_{j=1}^{3} \Lambda^j(t_1^1 | Z_j)). \tag{5.4}$$

Using equation (5.3), we may rewrite equation (5.2) as follows:

$$L_i^1 = \int\int\int Pr(T_2^1, ..., T_{K_{1i}}^1 | Z_1) Pr(T_1^2, ..., T_{K_{2i}}^2 | Z_2) Pr(T_1^3, ..., T_{K_{3i}}^3 \mid Z_3) * g(Z_1, Z_2, Z_3 \mid T_1^1 = t_1^1, T_1^2 > t_1^1, T_1^3 > t_1^1) dZ_1 dZ_2 dZ_3. \tag{5.5}$$

Using the model specified in equation (5.1) we can rewrite

$$Pr(T_2^1, ..., T_{K_{1i}}^1 | Z_1) = \prod_{k=2}^{K_{1i}-1} \lambda^1(t_k^1 | Z_1) [\lambda^1(t_{K_{1i}}^1 | Z_1)]^{d_{K_{1i}}^1} \exp(-\sum_{k=2}^{K_{1i}} \Lambda^1(t_k^1 | Z_1)), \tag{5.6}$$

where $\Lambda^j(t_k^j | \theta_j)$ is the conditional cumulative hazard function, and $d_k^j = 1$ if the k^{th} event of type j occurred, and zero otherwise. In the same way, we can derive

$$Pr(T_1^2, ..., T_{K_{2i}}^2 | Z_2) = \prod_{k=1}^{K_{2i}-1} \lambda^2(t_k^2 | Z_2) [\lambda^2(t_{K_{2i}}^2 | Z_2)]^{d_{K_{2i}}^2} \exp(-\sum_{k=1}^{K_{2i}} \Lambda^2(t_k^2 | Z_2)), \tag{5.7}$$

and $Pr(T_1^3, ..., T_{K_{3i}}^3 | Z_3)$.

For convenience sake we used a parametric survival function for $\lambda_0^j(t_k^j | Z_j)$, and we evaluated exponential, Weibull and lognormal functions. We especially need a para-

metric model for the baseline hazard because we have very little information on risk in early life, i.e. before the first index event.

The total likelihood is calculated by multiplying the contributions of all patients, and its logarithm is maximized using a Newton-Raphson algorithm in the S-plus computer package. The trivariate integral involved in the likelihood is approximated by Gauss-Hermite quadrature (Abramowitz and Stegun, 1965). In our experience the approximation is sufficiently accurate by using nine quadrature points per dimension. To allow quadrature, $Z = (Z_1, Z_2, Z_3)$ is transformed into independent components $\xi = (\xi_1, \xi_2, \xi_3)$ with variance equal to unity $Z = \Lambda\xi$:

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix},$$

where a, b, c, d, e and $f \in (-\infty, \infty)$ and are parameters to be estimated. The covariance matrix of Z is $\Sigma = \Lambda\Lambda^T$.

Notice that the shared-frailty model is a special case because we assume there that $Z_1 = Z_2 = Z_3$ or, equivalently, that

$$\Sigma = \begin{pmatrix} \sigma^2 & 1 & 1 \\ 1 & \sigma^2 & 1 \\ 1 & 1 & \sigma^2 \end{pmatrix}.$$

Other special cases of this multivariate frailty model are possible, for instance a model with perfect or zero correlations between Z_1, Z_2 and Z_3 but heterogeneous variances:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 & \sigma_1\sigma_3 \\ \sigma_1\sigma_2 & \sigma_2^2 & \sigma_3\sigma_2 \\ \sigma_1\sigma_3 & \sigma_3\sigma_2 & \sigma_3^2 \end{pmatrix}$$

or

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}.$$

5.4 Results

Our data set contains 307 Dutch patients who were included in the Caprie trial. These patients were included in the study because they suffered one of the three types of event. Table 5.1 shows how many events occurred in the three different strata and of which type: 145 new events in total occurred in 74 patients. 38 patients had 2 or more new events, 16 patients had 3 or more new events. Amongst the stratum of 103

patients who entered into the study with a cerebral atherosclerosis event, 28 recidives of cerebral events in 24 patients, 22 cardiac events in 16 patients and 7 peripheral events in 7 patients occurred. Amongst the stratum of 97 patients who entered into the study with a cardiac event, 11 cerebral events (9 patients) 25 recidives of cardiac events (14 patients) and 1 peripheral event (1 patient) occurred. Whereas amongst the stratum of 107 patients entered into the study with a peripheral event, 16 cerebral events (15 patients) 31 cardiac events (19 patients) and only 4 recidives of peripheral event (4 patients) occurred.

TABLE 5.1: *Number of events (cardiac, cerebral or peripheral) after entrance into the study with a cardiac, cerebral or peripheral atherosclerotic index event.*

	cerebral	cardiac	peripheral
cerebral index event	28	22	7
cardiac index event	11	25	1
peripheral index event	16	31	4

When we compare incidence of cardiac, cerebral and peripheral events within strata (excluding the index event), see Figures 5.1, 5.2 and 5.3, we found highly significant differences between event types, mainly suggesting that patients with a cardiac index event have a higher risk of cardiac events while patients with a cerebral index event have a higher risk of cerebral events. This suggests that - at least - cardiac and cerebral atherosclerosis are distinct disease entities but we argue that this is mainly due to the selection of specific patients on the basis of the index events. Indeed, the relative risks of the type of index event (cardiac, and peripheral versus cerebral) on new atherosclerotic events were not significantly different from unity (0.88 with 95% confidence interval $[0.58; 1.35]$), and 0.81 with 95% confidence interval $[0.55; 1.20]$, respectively) in a Cox-regression model with delayed-entry until age at first index event. In contrast, age at index event was highly significant ($RR = 0.81$ with 95% confidence interval $[0.78; 0.85]$). When fitting a simple shared-frailty model to these data, we found that the variance of the frailty distribution was highly significantly larger than zero. This indicates that there is at least some dependence between repeated events in the same patient.

To estimate the association between the cardiac, cerebral and peripheral atherosclerosis processes in a random person we estimated the parameters of the (multidimensional) frailty model. These estimates are given in Table 5.2. The model with 3 dependent frailties is significantly better than the model with 3 independent frailties ($p \leq 0.001$) and than the model assuming independence between events ($p \leq 0.001$), but not significantly better than the model with only one shared-frailty ($p = 0.75$).

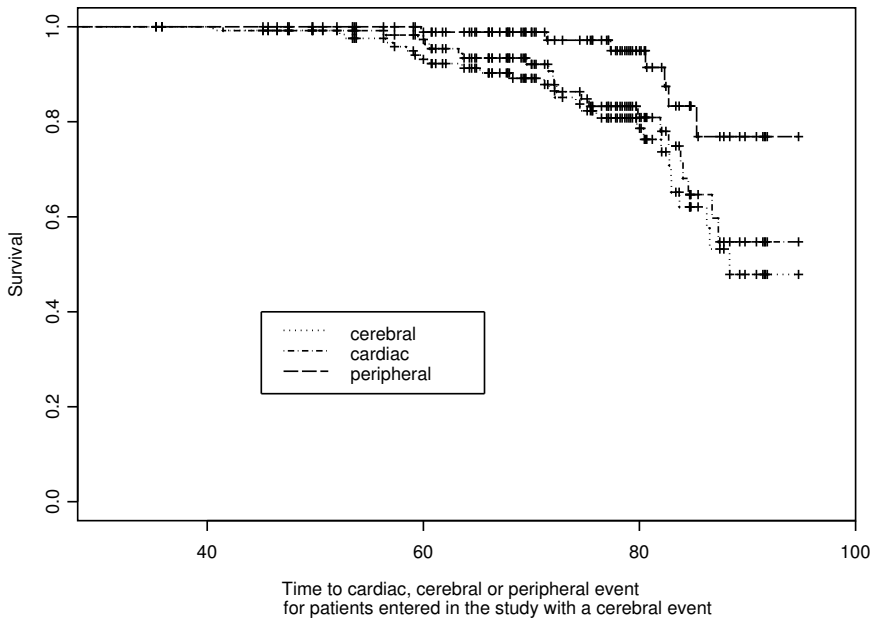

FIGURE 5.1: *Kaplan-Meier curves for the different event types for patients with a cerebral index event.*

TABLE 5.2: *Log likelihood and estimation of the different parameters of the models*

model	correlation	loglik	σ_1	σ_2	σ_3	ρ_{12}	ρ_{13}	ρ_{23}
independence	r[1]$= 0$	- 1027.1265	–	–	–	–	–	–
1 frailty	$r \geq 0$	- 1017.0529	2.21	2.21	2.21	1	1	1
3 ind. frailties	r_1[2]≥ 0 and r_2[3]$= 0$	- 1026.6539	0.29	0.35	0.01	0	0	0
3 dep. frailties	$r_1 \geq 0$ and $r_2 \geq 0$	- 1015.7292	2.76	1.89	2.14	1	1	1

[1] r = correlation between all the events
[2] r_1 = correlation between events of same type
[3] r_2 = correlation between events of different type

5.5 Concluding remarks

Our results indicate very strong relationships between the processes responsible for developing cardiac, cerebral, and peripheral atherosclerosis. The model with 3 correlated

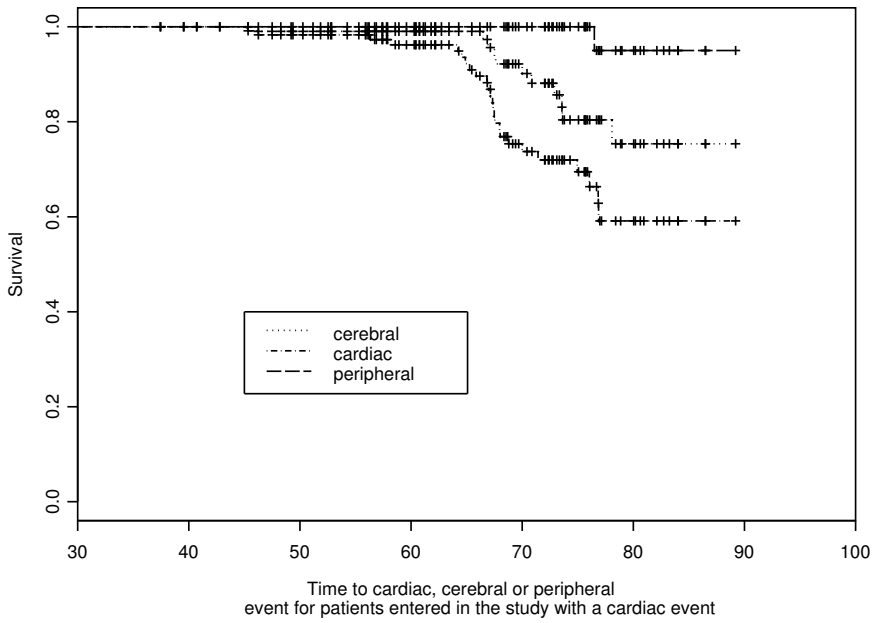

FIGURE 5.2: *Kaplan-Meier curves for the different event types for patients with a cardiac index event.*

frailty parameters for these three atherosclerotic processes fitted better than the shared frailty model, but this difference was not statistically significant. The point estimates of the correlations between the three correlated atherosclerotic processes was equal to unity, which is equal to the correlation implied by the shared frailty model. However, the standard deviations in the model with three correlated frailty parameters were not equal to each other, 2.757, 1.891 en 2.143, and this made the difference with the shared frailty model. Although it is difficult to distinguish the fit of the shared frailty model and the model with three correlated frailty parameters, both fit significantly better than the model with three independent frailty parameters, and also better than the independence model without frailty parameters. This clearly indicates that the risk for any atherosclerotic event differs between individuals, and this variation is responsible for the correlations that we observe.

Notice that the estimates of the standard deviations of the model with three independent frailty parameters were remarkably lower $(0.29, 0.35, 0.01)$ than that of the model with three dependent frailty parameters, or of the shared frailty model. The association

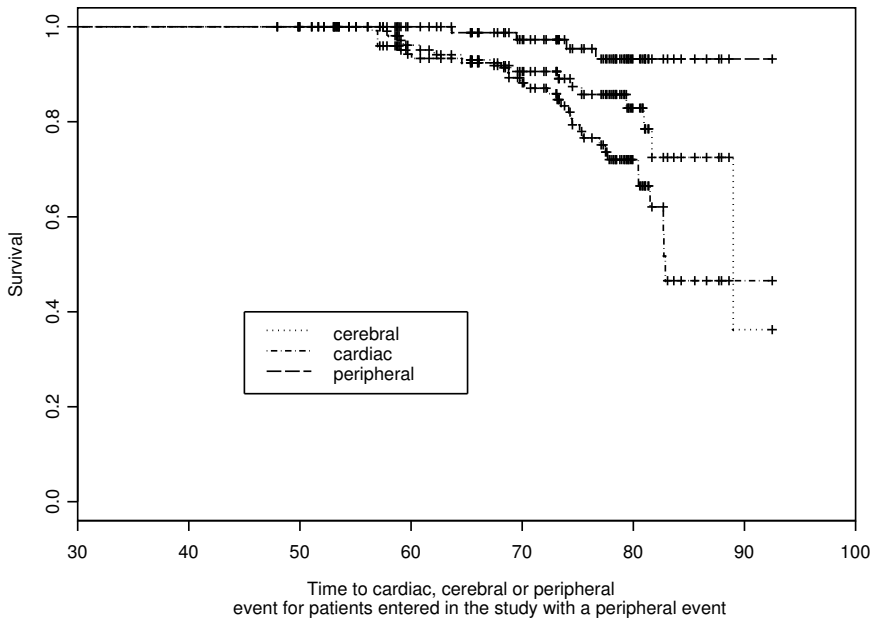

FIGURE 5.3: *Kaplan-Meier curves for the different event types for patients with a peripheral index event.*

between different type of events was clearly at least as strong as the association between repeated events of the same type, and when fixing the first to zero, the association between repeated events of the same type is also estimated much lower.

The model with three correlated frailty parameters is difficult to identify, obviously due to the small number of second, and third events in the patients in our sample. Our results need validation in much larger studies with longer follow-up. The ascertainment of patients with a first event complicates the model too, but such ascertainment occurs almost certainly in many follow-up studies of patients since patients present themselves with (suspected) atherosclerosis in at least one region (cardiac, cerebral or peripheral). Only in cohorts of individuals sampled at random such ascertainment does not occur.

In the estimation of the parameters of the three-correlated-frailties model standard algorithms for evaluating the three-double integral in the log likelihood were used, and we found that a nine-point quadrature algorithm was sufficiently precise, and this number can be easily enlarged when necessary. This approach is basically the same as was chosen by Vaida and Xu (2000) in their bivariate frailty model; we just generalized the

approach to three dimensions. Our models presume that the association between repeated events of the same type can be described by a simple exchangeable covariance structure. This is not necessary per sé, and generalization to other structures can be done in a similar fashion as in usual with random effects models. This will, however, complicate the efficiency of the estimation algorithm because a high-dimensional integral must then be evaluated. Perhaps MCMC techniques are useful in that case, but with the present data we felt to have too little information to evaluate more complicated association structures. We also had to little data to evaluate the fit of different parametric baseline functions, which was especially due to the inclusion in the trial only of patients after their first atherosclerotic event. That meant that we had no information at all about the shape of the cumulative hazard function before that event. We therefore assumed that the hazard was constant after this event.

Clinically, it is most important that our results suggest that the risk of a next event (any event) depends mainly on the age of patient at the first event, and not on the type of event. This is in contrast with the results of Cotter et al. (2003), but they did not consider the ascertainment issue. Cotter et al. found that patients with previous peripheral or cerebral atherosclerotic events had far more risk for cardiac events. This is similar to what we find, but we observed the reverse too.