

USING SURVIVAL TO PREDICT SURVIVAL

Jelle Goeman

Master's thesis in mathematics
written at the
Department of Medical Statistics
Leiden University Medical Center

Supervisors:
dr S. le Cessie
Department of Medical Statistics
Leiden University Medical Center
and
prof. dr S. A. van de Geer
Mathematical Institute
Leiden University

July 23, 2001

Contents

1	Introduction	3
2	Modeling	5
2.1	Modeling relative survival	5
2.2	Modeling frailty	6
2.3	Combining relative survival and frailty	6
2.4	Difficulties of the model	7
3	Population and individual	8
3.1	The population function	8
3.2	Properties of the population function	9
3.3	Examples	13
4	The effect of previous tumors	16
4.1	The effect function	16
4.2	Properties of the effect function	17
4.3	Proportional excess hazards	18
4.4	Examples	18
5	Application to the data	21
5.1	Introducing covariates	21
5.2	Using the approximation to the effect function	22
5.3	Using the full effect function	23
6	The weakened survivor	24
6.1	An extension of the model	24
6.2	Population and effect functions	25
6.3	Application	25
6.4	The limits of modeling	26
7	The extra hazard effect	28
7.1	Comparison of estimates	28
7.2	Comparison of predictions	30
8	The Puerto Rico example	32
9	Conclusion	35
A	Proofs of the lemma's	36
	References	40

Abstract

Information on disease history and comorbidity of patients can often be of great value to predict survival in cancer research. In this paper a model is presented that accommodates such information by combining relative survival and frailty. Relative survival is used to model the excess risk of dying from recent previous tumors. Individual frailty allows estimation of either overdispersion (past survivors live longer) or underdispersion (past survivors die sooner) by making use of each patient's survival before entry into the study. Results are shown to be independent of the chosen family of frailty distributions. The model is applied to data from the Leiden University Medical Center on patients with head/neck tumors.

1 Introduction

Patients have a history. Only rarely do they enter a clinical study as a pure *tabula rasa*. Some might have survived a similar disease in the past. Others may be suffering multiple afflictions at the same time. Such information is generally available about patients and can be of great value to predict patient survival or to estimate the ‘true’ hazard of the disease under study.

In this paper I will present a model to study how information on disease history and comorbidity can be used to predict survival. It forms part of a project by R. J. Baatenburg de Jong and S. le Cessie (2001) at the Leiden University Medical Center (L.U.M.C.) to build a prognostic model for the survival of patients with various cancers of the head and neck (lip, oral cavity, larynx, pharynx) using survival data on 1396 patients diagnosed at the L.U.M.C. between 1981 and 1998 and followed up to ten years. Prognostic factors include details on the stage characteristics of the tumor (M-, N- and T-values), the location of the tumor and age and sex of the patients. All these factors could be well fitted into a Cox proportional hazards model.

In addition to this, details of the patients’ cancer history were known. About 10% of the patients (139) had been diagnosed with some other primary tumor before being diagnosed with the head/neck tumor. The site of this previous cancer was known as well as its date of diagnosis. The interval between the diagnosis of the previous cancer and the diagnosis of the head/neck cancer under study varied between 1 day and 38 years.

It was expected that this information could be of great value for improving predictions. As there is no appealing way to do this within the Cox model, this meant abandoning this model and constructing a new one. Intuitively the following two effects were expected and had to be covered by the new model.

1. It was expected that a previous tumor would of course generate extra hazard in addition to the hazard generated by the head/neck tumor. A patient who has two tumors is more likely to die than a patient who has one. I will call this the *extra hazard effect*. The model should incorporate this effect.
2. It was also expected that there might be some sort of *selection effect*. Suppose a patient survives some serious tumor for a long time before entering the trial. This can be seen as evidence that this person is very tough and well able to survive tumors, because frailer people would not have survived long enough to even enter our study. Using the

model it should be possible to evaluate the existence and importance of such a selection effect.

In this paper a model will be formulated, investigated and applied that can deal with both expected effects of previous tumors. It will combine *relative survival* to deal with the extra hazard effect and *frailty* to deal with the selection effect. Finally the model will be extended to be able to include a 'negative' selection effect, called the *weakened survivor effect*, so that patients who have survived much in the past can also have shorter expected survival times than those who are 'fresh'. The new model will be applied to the data and compared to the original Cox model.

2 Modeling

In this section I will construct the model that incorporates both the extra hazard effect and the selection effect. Some difficulties in the resulting model will be addressed that will be solved later in this paper.

2.1 Modeling relative survival

First we need outside information on the hazard generated by the previous tumors, because there are too few patients in the dataset to be able to estimate this hazard. For this we used the 1-, 5- and 10-year relative survival rates that the Eindhoven Cancer Registry publishes for tumor sites (Coebergh e.a. 1995).

They derive these figures using relative survival models. In such models it is assumed that a tumor generates an *excess hazard* $h(t)$ which is added to the *background hazard* $h_0(t)$ that the patient would have experienced if he would not have had the disease. This $h_0(t)$ is assumed to be known from the tables of general population mortality. The total hazard a person experiences will be

$$\lambda(t) = h_0(t) + h(t - D),$$

where D is the date of diagnosis (and $h(t)$ is taken as zero when t is negative).

Note that this model assumes that at all times the event of dying of the tumor is independent of the event of dying from other causes. Under this assumption hazards may be added up.

There is a very straightforward way to extend the relative survival model to multiple tumors. We assume the events of dying from the two tumors to be independent as well. In that case we can write

$$\lambda(t) = h_0(t) + h_1(t - D_1) + h_2(t - D_2), \quad (1)$$

where $h_1(t)$ is the hazard generated by the previous tumor and $h_2(t)$ the hazard generated by the head/neck tumor we are interested in modeling. We can assume $h_1(t)$ known for each tumor from the data from the Eindhoven Cancer registry. To get its values for all t we must use some way of interpolating between the published 1- 5- and 10-year relative survival figures, e.g. linearly.

We see that using relative survival it is easy to incorporate the extra hazard effect of the previous tumor into the model. Each patient has a background hazard according to his or her age and sex, but those who have a previous tumor have their background hazard augmented by the hazard of this tumor.

2.2 Modeling frailty

The selection effect calls for a heterogeneity in the observed population. If some people are inherently ‘tougher’ or better able to survive tumors than others, a long survival of some tumor previous to entry in the study will be indicative of a long survival during the follow-up.

This heterogeneity can be modeled by assuming that each person is born with an individual *frailty*, a random effect that partly determines his or her hazard function. Each person’s frailty is considered principally unobservable and is often thought of as the combination of all risk factors not already included in the model.

In a commonly used model (Vaupel e.a. 1979) the hazard rate of a person at age t with frailty Z is given by

$$h(t|Z) = Zh(t) \tag{2}$$

The frailty Z is a stochastic variable defined on \mathbb{R}^+ (with $Z > 0$ almost surely), which describes the frailty in the population at birth. It is assumed that a person’s frailty does not change with time. Note that if $EZ < \infty$ we can put $EZ = 1$ without loss of generality, because all constants can be moved into $h(t)$.

2.3 Combining relative survival and frailty

As we are only interested in heterogeneity in the ability to survive tumors, we can combine the models (1) and (2) by assuming one frailty towards tumors. This frailty will then only influence the relative survival hazards $h_1(t)$ and $h_2(t)$. This way we get our final model

$$\lambda(t|Z) = \mu_0(t) + Z [\mu_1(t - D_1) + \mu_2(t - D_2)] \tag{3}$$

Here $\mu_0(t)$ is the hazard rate due to general causes and remains unchanged ($\mu_0(t) = h_0(t)$). However $\mu_1(t)$ is now the excess hazard due to the previous tumor *of a person with frailty $Z = 1$* and $\mu_2(t)$ the excess hazard due to the head/neck tumor *of a person with frailty $Z = 1$* . A similar model was studied for different purposes by Zahl (1997).

With the model (3) it becomes possible to model the desired selection effect. As patients experience tumor hazard, those with highest frailty will generally die first. As a result the average frailty in the population will decrease. So a patient who enters a study with a long history of tumor hazard will have a lower expected frailty than another patient with a shorter history. This lower expected frailty will predict a longer survival.

Note that if $\text{Var}(Z) = 0$ everyone will almost surely have frailty $Z = 1$ and the model (3) will reduce to the model (1). Model (3) is therefore an extension of model (1).

2.4 Difficulties of the model

There is a difficulty with individual frailty models such as the one I proposed in the previous paragraph. As each person will die only once, it can often not be determined whether this was due to personal frailty or to general hazard. Therefore it is often impossible to make an inference about frailty and adding an individual frailty to a given model usually does not allow any additional inference at all. Because of this, frailty models are now almost exclusively used when there are multiple events per individual or where one frailty is shared by many people (Hougaard 1984).

Individual frailty models can however become useful when additional assumptions can be made or when more information is somehow available. In this case we have the following additional information on the hazard function.

- The general population hazard $\mu_0(t)$ is known.
- Each patient in the study survived at least until his own date of diagnosis with the head/neck tumor, so that the time of death $T > D_2$.
- There is information on the hazard rate $\mu_1(t)$ of the previous tumor for the individual with $Z = 1$. We do not know this function as in the model (1) because we cannot observe frailty. However we know the function $h_1(t)$, which is the hazard due to the previous tumor as observed in the population.

In addition it will sometimes be necessary to assume a parametric distribution for Z . I will however show that if the variance of Z is small all distributions will yield similar results.

In this paper I will show that together these pieces of information are enough to make an inference on the presence and importance of a selection effect in the data on the head/neck tumors.

3 Population and individual

When modeling with frailty there is always a difference between the hazard an individual experiences and the hazard that is observable in the population. We will call the terms $\mu_1(t)$ and $\mu_2(t)$ the *individual hazard*. It is the hazard experienced by an individual with frailty $Z = 1$. If someone's frailty is known, his hazard can be found by multiplying his frailty with the individual hazard function. The model (3) is formulated in terms of this individual hazard.

We can never, however, observe individual hazard. We always observe population hazards. That is we observe deaths without knowing about frailty. In this section the connection between these two kinds of hazard will be explored, extending results by Hougaard (1984).

To simplify we write $\mu(t) = \mu_1(t - D_1) + \mu_2(t - D_2)$, so that model (3) becomes

$$\lambda(t|Z) = \mu_0(t) + Z\mu(t) \quad (4)$$

Model (4) written in this way can be seen as a more general model to study the connection between individual and population hazards. We will return to the original model later.

3.1 The population function

Most of the analysis in this paper will be in terms of cumulative hazard. Therefore it is convenient to rewrite the model (4) in terms of cumulative hazard. Let $M(t) = \int_0^t \mu(s)ds$ be the cumulative individual hazard, $M_0(t) = \int_0^t \mu_0(x)dx$ the cumulative background hazard and $S_0(t) = e^{-M_0(t)}$ the background survival probability. Now (4) becomes

$$\Lambda(t|Z) = M_0(t) + ZM(t).$$

In terms of survival this becomes

$$S(t|Z) = S_0(t)e^{-ZM(t)},$$

from which equation we can integrate Z out to get the population survival function. This is

$$S(t) = \mathbb{E} \left(S_0(t)e^{-ZM(t)} \right) = S_0(t)L[M(t)],$$

where $L(s) = \mathbb{E} [e^{-sZ}]$ is the Laplace transform of the density $f(z)$ of Z . Now taking logarithms yields the cumulative hazard in the population

$$\Lambda(t) = M_0(t) + g[M(t)], \quad (5)$$

where $g(s) = -\log L(s)$.

This function $g(s)$ will be called the *population function* and will play an important role in the rest of this paper. It is the function which ‘translates’ an individual cumulative hazard function into a population cumulative hazard function.

The population function can also be used to track the decrease in average population frailty. The density of frailty among survivors at time t is given by Bayes’ rule as

$$f(z|T \geq t) = \frac{P(T \geq t|Z = z)f(z)}{P(T \geq t)} = \frac{S(t|z)f(z)}{S(t)} = \frac{e^{-zM(t)}f(z)}{L[M(t)]}, \quad (6)$$

where $f(z)$ is the density of Z . The expectation of the density (6) is (Hougaard 1984)

$$E_{t+}(Z) = E(Z|T \geq t) = \frac{E(Ze^{-ZM(t)})}{E(e^{-ZM(t)})} = \frac{-L'[M(t)]}{L[M(t)]} = g'[M(t)], \quad (7)$$

where the final two equations hold whenever $L(s)$ and $g(s)$ are differentiable (see lemma’s 2 and 3).

The equation (7) has an important role in understanding the difference between individual and population hazards. If we differentiate (5) to get non-cumulative hazards we can use (7) to see that

$$\lambda(t) = \mu_0(t) + E_{t+}(Z)\mu(t).$$

This means that at each time t the population hazard is the individual hazard of a person with average frailty at that time t . Population hazard therefore decreases with time relative to individual hazard because average frailty in the population decreases with time.

3.2 Properties of the population function

In this paragraph I will deduce some properties of the population function $g(s)$ which hold for all distributions of Z . These properties will take the form of five lemma’s, which I will outline here. Proofs of these lemma’s can be found in appendix A. The most important result is formulated as a theorem and proved at the end of this paragraph.

Equation (5) tells us how to find the population hazard from the individual hazard when the distribution of Z is known. Lemma 1 says that conversely the individual hazard function can always be recovered from the population hazard.

Lemma 1 *The population function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is invertible.*

The result (7) (and more results later) require differentiability of the population function. A sufficient condition for this is given by lemma 2.

Lemma 2 *If $E(Z^n) < \infty$ the population function $g(s)$, $s \geq 0$, is n times continuously differentiable.*

For the following three lemma's we need to define a new family of stochastic variables, called *survivor frailty*. First note that the density of frailty at time t , as given in (6), depends only on $M(t)$, the total amount of individual hazard survived up to t .

This means that we can define the stochastic variable Z_m for $m \in \mathbb{R}^+$, called the m -survivor frailty. This is the frailty of patients who had a frailty distributed as Z at $t = 0$ and survived a cumulative cancer hazard of size m since. Thinking in terms of survivor frailty is useful because if a person enters the study after surviving a hazard of size m , his frailty will be distributed as Z_m .

By definition $Z_0 = Z$ and the density of Z_m is given by (6) with $M(t)$ replaced by m . Using this we can calculate the Laplace transform for the survivor frailty Z_m . This is given by

$$L_m(s) = E \left(e^{-sZ} \frac{e^{-mZ}}{L(m)} \right) = \frac{L(m+s)}{L(m)}$$

and we can define the m -survivor population function as

$$g_m(s) = -\log L_m(s) = g(m+s) - g(m).$$

The following rule governs the relationship between the population functions of different survivor frailty's.

$$\begin{aligned} g_m(t+s) &= g(m+t+s) - g(m) \\ &= g(m+t+s) - g(m+t) + g(m+t) - g(m) \\ &= g_{m+t}(s) + g_m(t) \end{aligned} \tag{8}$$

Finally note that Z_m is a frailty, so that lemma's 1 and 2 hold for $g_m(s)$ as well.

Using survivor frailty and survivor population functions it becomes easy to prove more things about $g(s)$. The first interesting lemma is about the relation between the derivatives of the population function and the moments of the survivor frailty.

Lemma 3 *If $m > 0$ or $EZ < \infty$, $g'(m) = EZ_m < \infty$ and $-g''(m) = \text{Var}(Z_m) \leq \infty$.*

The first part of lemma 3 is of course equivalent to (7). The lemma also shows that the population function is always differentiable, except perhaps in the origin, thus improving on the conditions of lemma 2.

The following lemma 4 can be seen as a proof that expected frailty $E_{t+}(Z) = g'[M(t)]$ in the population decreases with time if there is hazard, because a concave function has a decreasing derivative.

Lemma 4 *The function $g(s)$, $s \geq 0$ is concave.*

The lemma also shows that if $EZ < \infty$, $E(Z_m) = g'(m) \leq g'(0) = EZ$ and therefore $g(s) \leq sEZ$.

Lemma 5 proves the intuitive notion that (if the model is true) those who have experienced more hazard before entry into the study never stand a worse chance of survival than those who have experienced less. If $m < n$ the n -survivor hazard is never higher than the m -survivor hazard, so that the survival curve is never lower.

Lemma 5 *For $0 \leq m < n$ and $s \geq 0$, $g_m(s) \geq g_n(s)$*

From lemma 3 it is clear that if $EZ = 1$ and $\text{Var}(Z) = \sigma^2$, the Taylor approximation to $g(s)$ around $s = 0$ is

$$g(s) = s - \frac{1}{2}\sigma^2 s^2 + o(s^2)$$

Much more can be said about the error in this approximation using to the next theorem. This theorem states that under a mild condition, for all distributions of Z , the population function will become approximately quadratic as the variance of Z becomes smaller. The importance of this theorem for application will become clear in paragraph 4.3.

Theorem 6 *Let Z belong to some parametric family $(Z_\sigma)_{\sigma \in \mathbb{R}^+}$ depending on a parameter σ , so that $E(Z_\sigma) = 1$ and $\text{Var}(Z_\sigma) = \sigma^2$ for all σ . If the family is chosen in such a way that*

$$\lim_{\sigma \rightarrow 0} \frac{E(Z_\sigma - 1)^3}{\sigma^2} = 0 \tag{9}$$

it holds for all s that if $\sigma \rightarrow 0$

$$g(s) = s - \frac{1}{2}\sigma^2 s^2 + o(\sigma^2)$$

Proof: Let $Z_\sigma = 1 + \sigma X_\sigma$. Then $E(X_\sigma) = 0$ and $\text{Var}(X_\sigma) = 1$. Then $E(Z_\sigma - 1)^3 = \sigma^3 E(X_\sigma^3)$ so that the condition (9) can be rewritten as

$$\lim_{\sigma \rightarrow 0} \sigma E(X_\sigma^3) = 0. \quad (10)$$

Next we write

$$\begin{aligned} L(s) &= E(e^{-sZ_\sigma}) \\ &= E\left(e^{-s(1+\sigma X_\sigma)}\right) \\ &= e^{-s} E(e^{-\sigma s X_\sigma}) \\ &= e^{-s} E\left(1 - \sigma s X_\sigma + \frac{1}{2} \sigma^2 s^2 X_\sigma^2 + R_2(-\sigma s X_\sigma)\right), \\ &= e^{-s} \left(1 + \frac{1}{2} \sigma^2 s^2 + E(R_2(-\sigma s X_\sigma))\right) \end{aligned} \quad (11)$$

where $R_2(x) = \frac{1}{2} \int_0^x (x-t)^2 e^t dt$ is the error term of the second order Taylor approximation to e^x .

If $\sigma \rightarrow 0$

$$\begin{aligned} \frac{1}{\sigma^2} E(R_2(-\sigma s X_\sigma)) &= E\left(\frac{1}{2\sigma^2} \int_0^{-\sigma s X_\sigma} (-\sigma s X_\sigma - t)^2 e^t dt\right) \\ &= E\left(\frac{1}{2} \int_0^1 -\sigma s^3 X_\sigma^3 (1-u)^2 e^{-\sigma s X_\sigma u} du\right) \\ &= \frac{s^3}{2} \int_0^1 -\sigma E(X_\sigma^3 e^{-\sigma s X_\sigma u}) (1-u)^2 du \rightarrow 0 \end{aligned}$$

by (10). Therefore the error term $E(R_2(-\sigma s X_\sigma)) = o(\sigma^2)$.

Now (11) becomes

$$L(s) = e^{-s} \left(1 + \frac{1}{2} \sigma^2 s^2 + o(\sigma^2)\right)$$

and using that $-\log(1+x) = x + o(x)$ the assertion follows. \square

Corollary: Under the same conditions the inverse of $g(s)$ (which exists by lemma 1) has an approximation

$$g^{-1}(s) = s + \frac{1}{2} \sigma^2 s^2 + o(\sigma^2).$$

\square

The requirement (9) asks that the skewness of the distribution does not go to infinity as the variance goes to zero. In commonly used frailty distributions (e.g. the gamma) skewness goes to zero as $\sigma \rightarrow 0$. In this case

$$\lim_{\sigma \rightarrow 0} \frac{\mathbb{E}(Z_\sigma - 1)^3}{\sigma^3} = 0,$$

and the assertions of theorem 6 and its corollary hold with an error term of order $o(\sigma^3)$.

3.3 Examples

I will illustrate the theory with examples of specific distributions. These will be taken from the exponential family introduced by Hougaard (1984), as for this family the Laplace-transforms have a simple mathematical expression. As said before it can be assumed without loss of generality that either $\mathbb{E}Z = 1$ or $\mathbb{E}Z = \infty$.

An important distribution to be considered is when $Z = 1$ almost surely. In that case the model (3) collapses into model (1) and there is no selection effect. The population function that goes with this distribution is the identity function $g(s) = s$. If there is no selection effect population hazard and individual hazard are identical.

Example 1 For a gamma distributed frailty Z with $\mathbb{E}Z = 1$ and $\text{Var}(Z) = \sigma^2$

$$g(s) = \frac{1}{\sigma^2} \log(1 + \sigma^2 s).$$

See figure 1. All functions have $g'(0) = 1$, but the difference between the individual hazard s and the population hazard $g(s)$ increases with the standard deviation σ . The line designated $\sigma = 0$ is in fact the limit $\sigma \rightarrow 0$ and corresponds to $Z = 1$ almost surely.

Example 2 For an inverse Gaussian distributed Z with $\mathbb{E}Z = 1$ and $\text{Var}(Z) = \sigma^2$

$$g(s) = \frac{\sqrt{1 + 2\sigma^2 s} - 1}{\sigma^2}.$$

See figure 2, where the same patterns arise as for the gamma distribution, except that the curves flatten out less slowly for high σ .

Example 3 For an alpha-stable distribution with $\mathbb{E}Z = \infty$ and parameter $0 < \alpha < 1$,

$$g(s) = s^\alpha.$$

Figure 1: The population function $g(s)$ for a gamma distributed frailty.

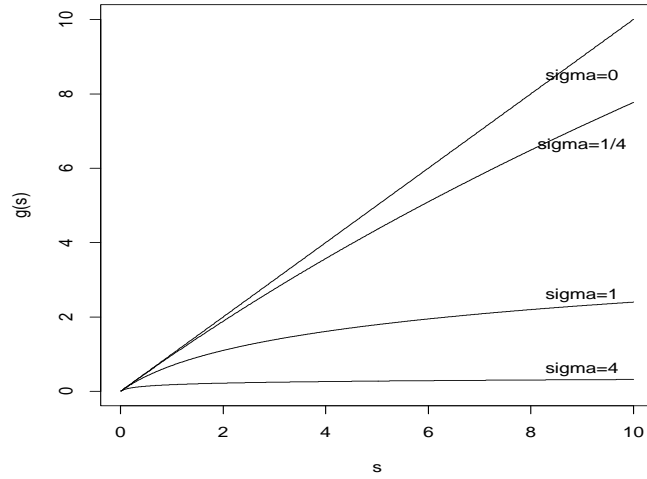
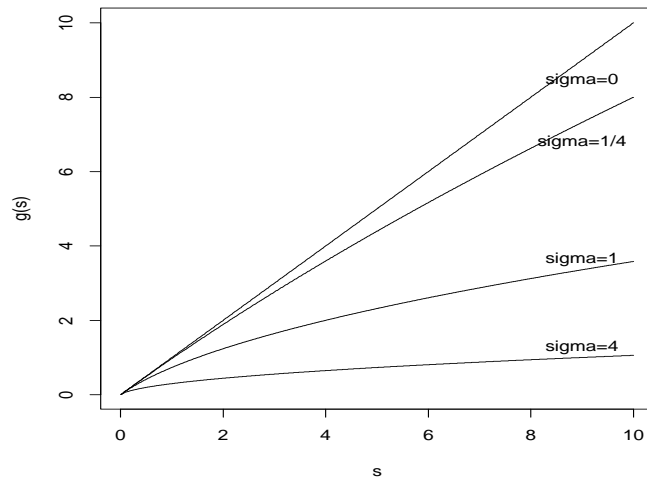
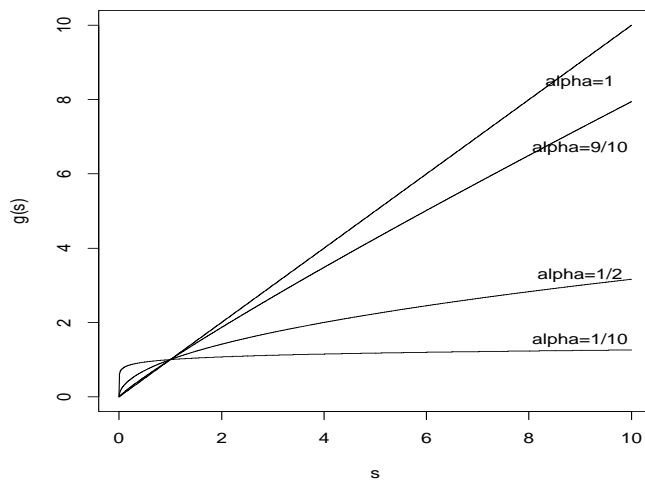


Figure 2: The population function $g(s)$ for an inverse Gaussian distributed frailty.



See figure 3. The main difference between the alpha-stable and the previous two distributions is the infinite expectation of the alpha-stable. This means that the derivative $g'(0)$ is infinite as well. The case $\alpha = 1$ is again a limit, corresponding to $Z = 1$ almost surely.

Figure 3: The population function $g(s)$ for an alpha-stable distributed frailty.



4 The effect of previous tumors

Now it is time to return to the model (3). This model was formulated in terms of unobservable individual hazards. Using the theory on population functions of section 3 it becomes possible to rewrite this model in terms of observable population hazards.

4.1 The effect function

We restart with model (3) and rewrite it in terms of cumulative hazards to get

$$\Lambda(t|Z) = M_0(t) + Z[M_1(t) + M_2(t)]$$

where $M_1(t) = \int_0^t \mu_1(x - D_1)dx$ and $M_2(t) = \int_0^t \mu_1(x - D_2)dx$.

We can assume that the patients that the Eindhoven Cancer Registry observes do not also have a head/neck tumor. Then $M_2(t) \equiv 0$, so that we can say using (5) that the excess cumulative hazard due to the previous tumor that the Registry observes, is given by

$$\int_0^t h_1(x - D_1)dx = H_1(t) = g[M_1(t)]. \quad (12)$$

On the other hand, for those patients who do not have a previous tumor $M_1(t) \equiv 0$. They experience an excess hazard which is purely due to the head/neck tumor and is equal to

$$\int_0^t h_2(x - D_2)dx = H_2(t) = g[M_2(t)], \quad (13)$$

also because of (5).

For those patients who have both a previous tumor and a head/neck tumor we can write $M(t) = M_1(t) + M_2(t)$ and use (5) again to see that their total cumulative excess hazard due to both tumors is given by

$$H(t) = g[M(t)] = g[M_1(t) + M_2(t)]. \quad (14)$$

By lemma 1 the function $g(s)$ is invertible so that we can rewrite (12) and (13) to $M_1(t) = g^{-1}[H_1(t)]$ and $M_2(t) = g^{-1}[H_2(t)]$, respectively. Now (14) becomes

$$H(t) = g(g^{-1}[H_1(t)] + g^{-1}[H_2(t)]) \quad (15)$$

Now we reset the time scale. Let $\tau = t - D_2$ be the time since entry into the study. Then define $G_1(\tau) = H_1(D_2 + \tau) - H_1(D_2)$, the hazard from the previous tumor since entry; $G = H_1(D_2)$, the ‘truncated’ hazard the patient

survived before entry; $G_2(\tau) = H_2(D_2 + \tau)$, the hazard of the head/neck tumor since entry and finally $G_0(\tau) = M_0(D_2 + \tau) - M_0(D_2)$ the background hazard since entry.

Now using (15) and the newly defined time scale we can write the model in terms of population hazards. The cumulative population hazard since entry becomes

$$\begin{aligned} \Lambda(D_2 + \tau) - \Lambda(D_2) &= H_0(D_2 + \tau) - H_0(D_2) + H(D_2 + \tau) - H(D_2) \\ &= G_0(\tau) + g(g^{-1}[G + G_2(\tau)] + g^{-1}[G_2(\tau)]) - G \\ &= G_0(\tau) + G_1(\tau) + \chi(G + G_1(\tau), G_2(\tau)), \end{aligned} \quad (16)$$

where $\chi(t, s) = g[g^{-1}(t) + g^{-1}(s)] - t$. The total hazard in equation (16) consists of two additive parts. The first is a known background hazard $G_0(\tau) + G_1(\tau)$, which gives the hazard this patient would have experienced if he would not have had a head/neck tumor. The second term is basically the head/neck hazard $G_2(\tau)$, transformed by the previous tumorhazard $G + G_1(\tau)$ via a function $\chi(t, s)$.

This $\chi(t, s)$ is called the *effect function*. It describes the transformation of a cumulative hazard s as a consequence of a previously or simultaneously survived hazard t . For example, suppose that up to some moment τ the excess hazard due to the head/neck tumor of people without a previous tumor is observed as s . Then for a person who has survived a previous tumorhazard of size t up to time τ this cumulative excess hazard will be observed as $\chi(t, s)$. Note that of course the shape of the effect function $\chi(t, s)$ depends on the distribution of Z .

4.2 Properties of the effect function

I will not deduce as many properties of $\chi(t, s)$ as I have done of $g(s)$. Only the following are important.

Firstly $\chi(t, s)$ is continuous because $g(s)$ and $g^{-1}(s)$ are. $\chi(0, s) = s$ and $\chi(t, 0) = 0$. As we can write the effect function as a survivor population function:

$$\chi(t, s) = g[g^{-1}(t) + g^{-1}(s)] - t = g_{g^{-1}(t)}(g^{-1}(s)),$$

it is an increasing function of s and by lemma 5 a decreasing function of t . By the same lemma $0 \leq \chi(t, s) \leq \chi(0, s) = s$.

The effect function has derivatives

$$\frac{\partial}{\partial s} \chi(t, s) = \frac{g'[g^{-1}(t) + g^{-1}(s)]}{g'[g^{-1}(s)]}$$

$$\frac{\partial}{\partial t}\chi(t, s) = \frac{g' [g^{-1}(t) + g^{-1}(s)]}{g'[g^{-1}(t)]} - 1, \quad (17)$$

which exist for $s > 0$ and $t > 0$ by lemma 3. Furthermore by the same lemma (if $t > 0$; trivial if $t = 0$): $\lim_{s \rightarrow 0} \frac{\partial}{\partial s}\chi(t, s) = g'[g^{-1}(t)] \leq EZ$. On the other side $\lim_{s \rightarrow \infty} \frac{\partial}{\partial s}\chi(t, s) = 1$.

Finally using the theorem 6 and its corollary we can find that when $\sigma \rightarrow 0$ in a family $(Z_\sigma)_{\sigma \in \mathbb{R}^+}$ as in the theorem, then for all t and s

$$\begin{aligned} \chi(t, s) &= t + \frac{1}{2}\sigma^2 t^2 + s + \frac{1}{2}\sigma^2 s^2 - \frac{1}{2}\sigma^2 \left(t + \frac{1}{2}\sigma^2 t^2 + s + \frac{1}{2}\sigma^2 s^2 \right)^2 - t + o(\sigma^2) \\ &= s - \sigma^2 st + o(\sigma^2) \\ &= se^{-\sigma^2 t} + o(\sigma^2). \end{aligned} \quad (18)$$

4.3 Proportional excess hazards

The approximation (18) will prove to be extremely useful. Substituted into the model equation (16) it leads to the equation

$$\Lambda(D_2 + \tau) - \Lambda(D_2) = G_0(\tau) + G_1(\tau) + G_2(\tau)e^{-\sigma^2(G+G_1(\tau))} + o(\sigma^2). \quad (19)$$

This is the key result of this paper. The approximation (19) to the model is now independent of the specific distribution of Z . The presence of a selection effect depends only on the variance of Z , so that it becomes unnecessary to postulate any distribution for Z . Apart from the variance σ^2 , which can be estimated, the model is formulated exclusively in terms of population hazard. Using (19) therefore the model can now be studied outside of the context of frailty models.

In fact (19) gives us a simple relative survival model in which the hazard $G + G_1(\tau)$ becomes a time-dependent covariate *with proportional hazards* to the head/neck tumor hazard $G_2(\tau)$. The approximated model is a model with proportional excess hazards, which is a well-known and frequently studied sort of model. See for example Sasieni (1996), Estève e.a. (1990) and software by Hédelin (1995).

4.4 Examples

I will illustrate the effect function with the same examples as I used with the population function. I use $t = 2$ fixed. All functions (except when $\text{Var}(Z) = 0$) converge at varying speeds to the line $s - t$. Varying t is not so interesting as it only changes the line to which the functions converge.

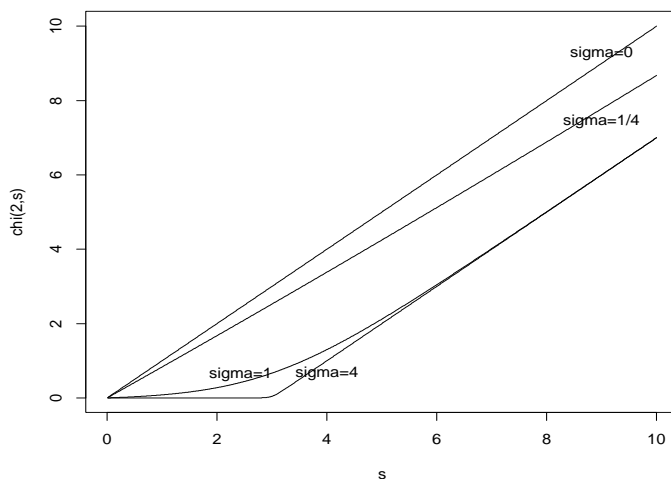
Notice how the effect function of the gamma and inverse Gaussian is almost a straight line through the origin when σ^2 is small. This corresponds to the approximation (18).

Example 1 (continued) For the gamma distribution $g^{-1}(s) = \frac{1}{\sigma^2} (e^{\sigma^2 s} - 1)$ so that

$$\chi(t, s) = \frac{1}{\sigma^2} \log (e^{\sigma^2 s} + e^{\sigma^2 t} - 1) - t.$$

See figure 4.

Figure 4: The effect function $\chi(t, s)$ for a gamma distributed frailty and $t = 2$.



Example 2 (continued) For the inverse Gaussian distribution $g^{-1}(s) = s + \frac{1}{2}\sigma^2 s^2$ so that

$$\chi(t, s) = \left(\left(s + t + \frac{1}{\sigma^2} \right)^2 - 2st \right)^{1/2} - \frac{1}{\sigma^2} - t.$$

See figure 5.

Example 3 (continued) For the alpha stable distribution $g^{-1}(s) = s^{1/\alpha}$ so that

$$\chi(t, s) = \left(t^{1/\alpha} + s^{1/\alpha} \right)^\alpha - t.$$

See figure 6.

Figure 5: The effect function $\chi(t, s)$ for an inverse Gaussian distributed frailty and $t = 2$.

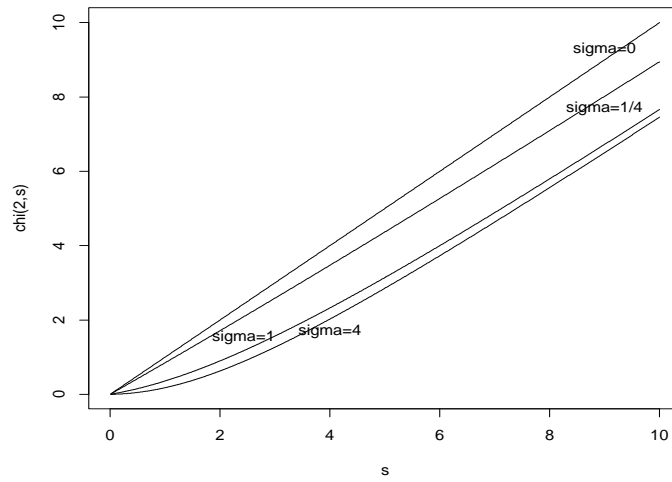
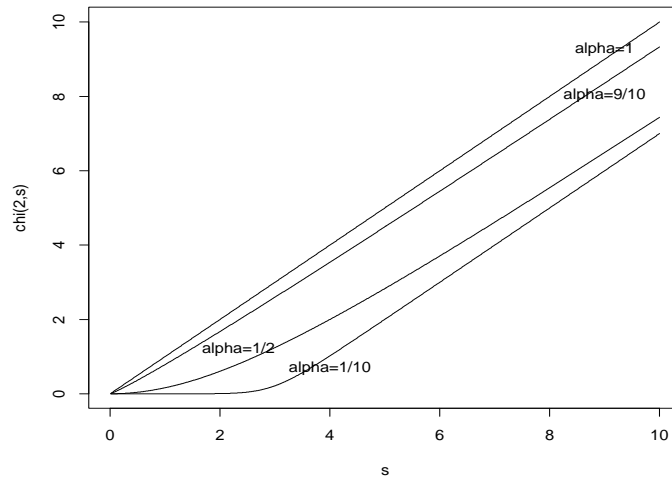


Figure 6: The effect function $\chi(t, s)$ for an alpha stable distributed frailty and $t = 2$.



5 Application to the data

We can now apply the theory developed in the previous paragraphs to the data on patients with head/neck tumors mentioned in the introduction. This application consists of two parts: determining the existence (testing) and establishing the importance (estimation) of the selection effect. First, however, we must introduce the covariates to our model that we have so far ignored.

5.1 Introducing covariates

We have a vector of covariates \mathbf{X} for each patient. There are two alternative ways of introducing this information into the model. We can assume proportionality of hazards at the level of the individual or of the population. I will consider both options and choose one.

In the original Cox-model mentioned in the introduction which did not use the information on previous tumors the effect of the covariates was modeled using proportional hazards on the population hazard $\Lambda(t)$. We can use a similar approach by replacing the relative population hazard $G_2(\tau)$ in (16) by $e^{\beta \cdot \mathbf{X}} B(\tau)$, with unknown parameters β and $B(\tau)$ an unknown baseline hazard. If we do this, we assume proportional hazards *at the population level*. This has the advantage of adhering closely to common practice and of being most easily applied.

However, assuming proportional hazards at the population level results in the effect of the covariates on the individual hazard *increasing with time* if there is a selection effect. Intuitively it would therefore be more appropriate to assume hazards to be proportional *at the individual level*. This would mean replacing $M_2(t)$ in (3) by $e^{\beta \cdot \mathbf{X}} B(t)$. If there is heterogeneity in frailty the effect of covariates on the population level now decreases with time.

Unfortunately, modeling in this latter way will introduce an interaction between the effect of the covariates and the effect of the previous tumor. Heterogeneity in frailty now corresponds to two things. Firstly it points to a selection effect due to previous survival as it was meant to. Secondly, however, it indicates that the effect of the covariates decreases with time. Therefore, if we find that some heterogeneity in frailty does exist, this might be only because the effect of the covariates decreases with time, and not because there is a selection effect. The two effects become entangled and cannot be estimated separately anymore. As it is only the selection effect we would like to estimate, it is better not to become confused in this way.

It is therefore preferable to model covariate effects as proportional at

the population level, because to do otherwise would make the existence of a selection effect impossible to determine.

5.2 Using the approximation to the effect function

It is most convenient to use the approximation of paragraph 4.3 for the effect function. This has the advantage of leading to a relatively simple and well-known model with proportional excess hazards. Furthermore using it does not require the assumption that the distribution of Z belongs to some family. Nonetheless being an approximation it should be used with care. There are two applications.

The first application is for estimation. If it is known by a priori convictions that σ^2 is small, (18) can be used to show that all distributions of Z will tend to the same effect function. In that case it is just as well to use the approximated model (19). There should be strong reasons to think that σ^2 is indeed small, however, because when σ^2 is not small errors will be large.

In the case of the head/neck tumor the great variety of previous tumors made it highly unlikely that there were many common factors involved in the ability to survive each of them. Therefore it was considered safe to assume that σ^2 was small.

For estimation in the proportional excess hazards model (19) different approaches exist. There is non-parametric approach developed by Sasieni (1996) and a simpler approach assuming a piecewise constant hazard rate by Estève (1990). Here we follow the second approach. As its associated software (Hédelin 1995) did not allow for time-dependent covariates, I adjusted the Estève model in S-plus to find a maximum likelihood estimate for σ^2 :

$$\hat{\sigma}^2 \approx 0.$$

This estimate is of course on the boundary of the domain of σ^2 . We could relax the restriction that $\sigma^2 \geq 0$, which, although unthinkable in the original model, is perfectly possible in the approximation (19). In that case we find even $\hat{\sigma}^2 \approx -0.244$. In section 6 I will investigate the meaning of this estimate and propose an extension of the model that will support this negative value.

The second application of the approximated model (19) is for testing the hypothesis $H_0 : \sigma^2 = 0$ that there is no selection effect against the alternative $H_A : \sigma^2 > 0$ that there is one. We can use this to assess whether a suspected selection effect exists.

Tests which could well be used here are a score test (compare Le Cessie and Van Houwelingen 1995) or a likelihood ratio test. As these tests are

based on (the derivative of the) likelihood (with respect to σ^2) at $\sigma^2 = 0$, we can substitute the approximation (18) for the effect function without introducing error. For the resulting proportional excess hazards model (19) these tests are treated by Sasieni (1996) for a non-parametric baseline hazard and by Estève (1990) for a parametric baseline hazard.

For the data on the head/neck tumors it was unnecessary to calculate the value of the test statistics, because the maximum likelihood estimate of σ^2 was zero. This means that the hypothesis H_0 can never be rejected by a score test or likelihood ratio test against the given alternative H_A that $\sigma^2 > 0$.

5.3 Using the full effect function

An different approach is needed if the selection effect that is to be evaluated is possibly large. In that case the approximated model (19) cannot be used and some parametric family of distributions must be chosen for Z .

This choice is not unproblematic and should be done with great care, as the family chosen greatly influences estimates of the characteristics of Z . Compare for example figures 4 and 5 and notice that in figure 6 all distributions (with $\alpha < 1$) have infinite variance. However if the same family is used for estimation and for prediction these differences should cause no great problems.

A straightforward way to proceed would be to parameterize the unknown baseline hazard function $B(\tau)$ as well, thus creating a full parametric model. Unfortunately the resulting likelihood function will be quite complex. Maximum likelihood estimates can therefore only be found numerically and it is not easy to construct confidence intervals. I have not pursued this route further.

6 The weakened survivor

In the data on the head/neck tumors we found a negative value as a maximum likelihood estimate for σ^2 . This is an indication that the effect found in the dataset is the opposite of the effect that was expected. People who have survived much hazard prior to entry in the study are not better able to survive new hazard, but instead they seem to do worse.

This is impossible in terms of the model outlined so far. This can be seen because by lemma 5 people who have survived more hazard in the past stand a better chance of survival in the future. Frailty always introduces *overdispersion* so that past and future survival are always positively correlated. This means we have to extend our model to capture *underdispersion*, the effect that past and future survival are negatively correlated.

If people who survived much in the past have a worse chance of survival, a clinician might explain this by what I will call a *weakened survivor effect*: after surviving a potentially lethal disease such as cancer many people never recover the state of health they had before the onset of the disease. This could be caused by the weakening effects of the disease itself and also by the aggressive treatments needed to overcome the disease. If survivors are weakened by survival, they are more prone to die by future hazard than those who have not experienced previous disease.

The weakened survivor effect is therefore the opposite of the selection effect. I will now integrate this effect into the existing model.

6.1 An extension of the model

The weakened survivor effect can be modeled with much the same theory as the selection effect.

We start with the same model as (3), but we now remove the assumption that the frailty is constant in time. In stead we write frailty as

$$Z(t) = Z_0 + \frac{1}{2}\alpha M(t) \tag{20}$$

Here $M(t) = M_1(t) + M_2(t)$ is the total cumulative tumor hazard up to t and Z_0 is a random frailty as in the previous model. After a random start the frailty increases as a patient experiences tumor hazard, at an unknown rate $\alpha \geq 0$.

The model (20) captures the weakened survivor effect because frailty increases with hazard. However the average frailty in the population might still decrease due to a selection effect, especially if α is small. Note that $\alpha = 0$ returns us to the previous model for just the selection effect. Taking

$Z_0 = 1$ almost surely ($\text{Var}(Z_0) = 0$) will model only the weakened survivor effect.

Note: There are many alternatives to the model (20). For example one can take

$$Z(t) = Z_0 \left(1 + \frac{1}{2} \alpha M(t) \right)$$

or

$$Z(t) = Z_0 + \frac{1}{2} \alpha Z_1 M(t)$$

where Z_0 and Z_1 are independent random effects. All these models lead to the same approximation (22).

6.2 Population and effect functions

A population function and an effect function can be derived for the new frailty $Z(t)$. We can write

$$\begin{aligned} L(s) &= \mathbf{E} \left(e^{-s(Z_0 + \frac{1}{2} \alpha s)} \right) \\ &= e^{-\frac{1}{2} \alpha s^2} \mathbf{E} \left(e^{-s Z_0} \right) \\ &= e^{-\frac{1}{2} \alpha s^2} L_0(s) \end{aligned}$$

so that $g(s) = g_0(s) + \frac{1}{2} \alpha s^2$, where $L_0(s)$ and $g_0(s)$ are the Laplace transform and the population function for Z_0 .

The function $g(s)$ will of course not have all the properties derived in paragraph 3.2 under the assumption of time-constant frailty. However $g(s)$ is invertible as a function $\mathbb{R}^+ \rightarrow \mathbb{R}^+$ because $g_0(s)$ is. So the effect function can be derived in the same way as in paragraph 4.1 and is still given by $\chi(t, s) = g[g^{-1}(t) + g^{-1}(s)] - t$, although without the properties of paragraph 4.2.

Using theorem 6 for $g_0(s)$ gives us (with $\sigma^2 = \text{Var}(Z_0)$)

$$g(s) = s + \frac{1}{2} (\alpha - \sigma^2) s^2 + o(\sigma^2)$$

From this it follows that $g^{-1}(s) = s + \frac{1}{2} (\alpha - \sigma^2) s^2 + o(\alpha + \sigma^2)$ and analogously to (18) we can derive an approximation to the effect function

$$\chi(t, s) = s e^{(\alpha - \sigma^2)t} + o(\alpha + \sigma^2) \quad (21)$$

6.3 Application

The approximation (21) can be used in the same way as the approximation (18) as described in paragraph 5.2. If we write $\gamma = \alpha - \sigma^2$, the model can

be written

$$\Lambda(D_2 + \tau) - \Lambda(D_2) = G_0(\tau) + G_1(\tau) + G_2(\tau)e^{\gamma(G+G_1(\tau))} + o(\alpha + \sigma^2) \quad (22)$$

where γ can be both negative and positive. This approximation might be used to estimate $\gamma = \alpha - \sigma^2$ if α and σ^2 are both known to be small. Note that it is not possible to estimate both α and σ^2 . Only their difference can be estimated.

A better application would be to use (22) to generalize the score and likelihood ratio tests mentioned in paragraph 5.2 to two-sided tests testing the hypothesis $H_0 : \alpha = 0$ and $\sigma^2 = 0$ that there is neither a selection effect nor a weakened survivor effect. Note that this test has very little power against the alternative that both effects are of the same ‘order’.

I calculated a likelihood ratio test to test the hypothesis H_0 . This can be done in the approximated model (22), using a piecewise constant hazard rate as in Estève (1990) and a parameterization

$$\alpha = \max(0, \gamma) \quad \sigma^2 = \max(0, -\gamma)$$

Note that this assumes that there is *either* a selection effect ($\gamma < 0$) *or* a weakened survivor effect ($\gamma > 0$). The likelihood ratio statistic was $\chi^2 = 1.79$ yielding a p -value of 0.18 on a χ^2_1 -distribution.

The weakened survivor effect that was found in the negative estimate of σ^2 was therefore not significant.

6.4 The limits of modeling

Extending the model in ways such as the one I just described quickly approaches the limits of what is possible with an individual frailty model. I have sketched the tightness of these limits in paragraph 2.4. Using the extra information on the hazard generated by the previous tumors supports only a limited amount of inferences.

For example the model can be used to estimate the importance of a selection effect *or* to estimate the importance of a weakened survivor effect. Trying to estimate both of them at the same time will already lead to difficulties. We have seen that is only possible to establish the difference, the ‘trade-off’ between the two effects, not to separate their individual effects. Only extra assumptions (for example on the unknown distribution of Z) or more extra information can support an estimate of the separate effects.

To make these limitations of the model clear, it is perhaps better to formulate models and hypotheses exclusively in terms of the population function. This is because only characteristics of the population function

can be estimated, not of the underlying models. The theory outlined in the previous paragraphs can then be used to support the model and interpret results.

For example the assumption that the ‘real’ population function becomes approximately quadratic as its second derivative $\gamma \rightarrow 0$, will lead to the familiar model

$$\Lambda(D_2 + \tau) - \Lambda(D_2) = G_0(\tau) + G_1(\tau) + G_2(\tau)e^{\gamma(G+G_1(\tau))} + o(\gamma)$$

which can be used to test the hypothesis that $g(s) = s$ against the alternatives that $g(s)$ is concave, meaning that a selection effect dominates, or convex, meaning that a weakened survivor effect dominates.

I believe thinking in these terms will make the issues clearer.

7 The extra hazard effect

Having established that there is no selection effect in the data, nor a significant weakened survivor effect, it is time to evaluate the extra hazard effect. First I will compare the estimates for the relative survival model to the original Cox-model which did not use the information on disease history.

In this section we assume that there is no selection effect and no weakened survivor effect. In that case we have just the relative survival model (1), without time-dependent covariates. This means that we can use the software *RELSURV* by Hédelin (1995). It is based on a model by Estève (1990) that uses a parametric baseline hazard that is piecewise constant.

RELSURV can also be used to calculate an ordinary Cox-model by setting the background hazard rate to zero. This way it is easy to compare the estimates of the relative survival and Cox models. I will show some of these comparisons to show the practical differences between an ordinary Cox model and a relative survival model incorporating the extra hazard effect.

7.1 Comparison of estimates

First I give a comparison of the estimates of the baseline hazard and the coefficients of the covariates. In both cases we write the head/neck tumor hazard as

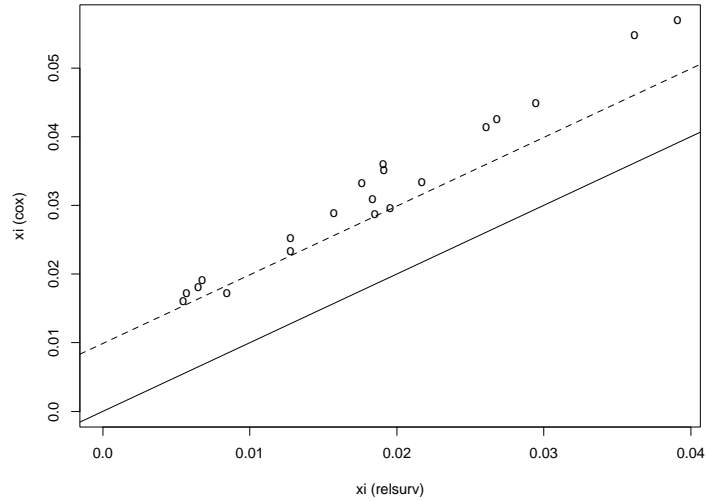
$$h_2(t - D_2) = e^{\beta \cdot \mathbf{X}} \sum_{k=1}^{20} \xi_k I_k(t - D_2) \quad (23)$$

where $I_k(t)$ is the indicator of the k -th three-month interval since diagnosis and ξ_k the constant baseline hazard for this interval. Note that in the Cox model the hazard (23) is the total hazard, while in the relative survival model it is only the head/neck hazard to which the background hazard is added.

In figure 7 we compare the estimates of the baseline hazards ξ_k , $k = 1, \dots, 20$ for the two models. We see that the ξ_k are all higher for the Cox-model than for the relative survival model. This is understandable because an extra background hazard is added in the relative survival model, which has to be incorporated in the baseline hazard in the Cox model. However subtraction of the average background hazard (dotted line) does not remove the difference completely.

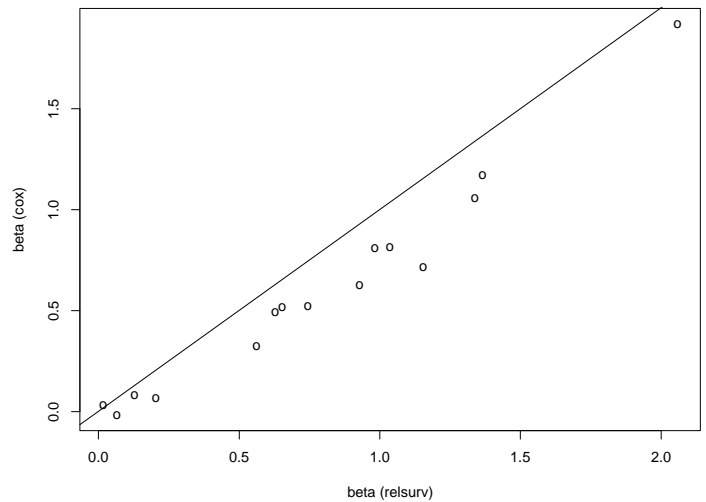
When we compare the coefficients of the covariates in the two models we also see a systematic difference. In figure 8 it is apparent that the estimates of the relative risks for the relative survival model are consistently higher than those of the Cox model. This too is an understandable consequence of

Figure 7: Comparison of Cox model and relative survival 1: baseline hazard ξ_k .



the presence of a baseline hazard. In the relative survival models the relative risks apply to the lower head/neck baseline hazard, while in the Cox model the relative risks apply to the higher total baseline hazard.

Figure 8: Comparison of Cox model and relative survival 2: coefficients β_j .

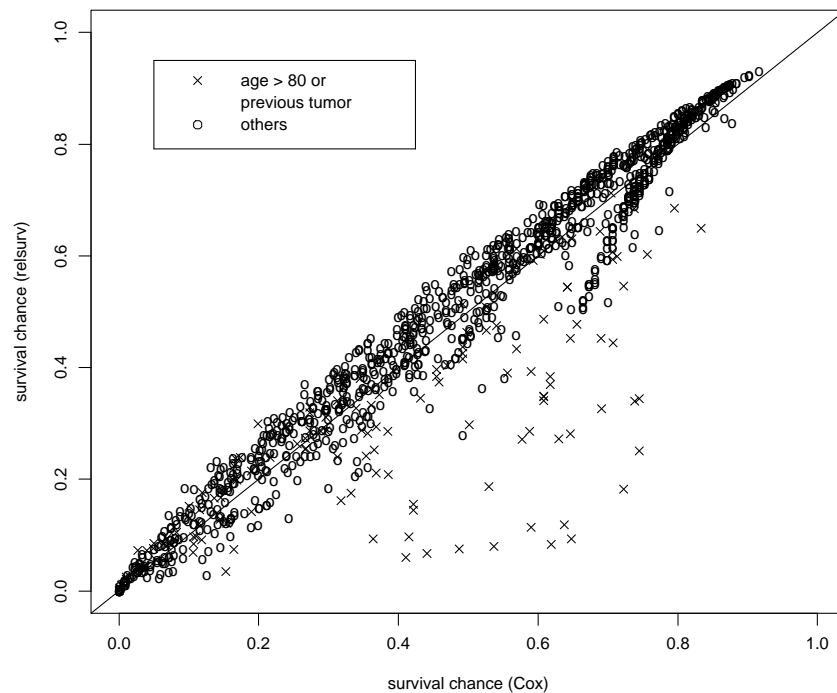


7.2 Comparison of predictions

What are the consequences of the differences found in the previous paragraph? I will compare predictions and deviance residuals.

First we can calculate the predicted five-year survival chance for each patient for the two models. This is given in figure 9. As could have been expected, the main differences between the predictions of the two models occur when patients have a high background hazard. This is the case when they are old or have a previous tumor. Most other patients have a slightly better prognosis in the relative survival model.

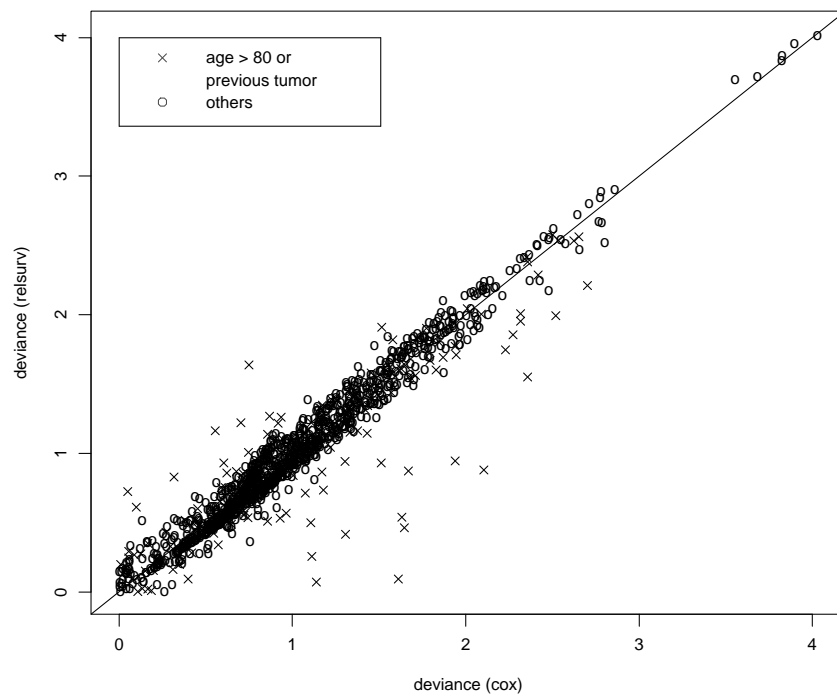
Figure 9: Comparison of Cox model and relative survival 3: Five year survival probabilities.



Finally we can calculate the deviance residuals to compare the predictions of the two models. The result given in figure 10. Again the greatest difference between the two models occurs when the patients have a high background hazard. Looking at the figure it seems that the relative survival model does indeed make a better prediction for most of these patients.

The total deviance is given by *RELSURV* as 2412.1 for the relative survival model and to 2697.4 for the Cox model. This can be used as a measure of how much better the prediction of the relative survival model is. Note that the difference is enormous even though the two models have the same number of parameters. Unfortunately as the models are not nested, there is no simple formal way to test one model against the other.

Figure 10: Comparison of Cox model and relative survival 4: deviance residuals (absolute values).



8 The Puerto Rico example

I will give another example of how the approximation (18) can be used to great effect.

In one of the first publications on individual frailty models by Vaupel e.a. (1979) some motivating examples are given why frailty models should be given serious consideration. These concern comparisons between populations.

‘[...] heterogeneity in frailty may explain some puzzling anomalies in period mortality differentials between populations. [...] For most pairs of populations—e.g., U.S. whites vs. blacks, for U.S. males vs. females, or for Americans vs. Swedes—mortality differentials tend to decline or even reverse with age. [...] One striking and surprising reversal concerns Puerto Rico which, among countries for which good mortality statistics are available, is the world’s leader in life expectancy at age 65 [...].’

The authors go on to show when such a reversal might take place, assuming a gamma distribution and proportional hazards.

Using the models and theory outlined in this paper much more can be said on this question. I will show that if a reversal occurs, an approximate minimum value may be constructed for the variance of the frailty.

Let us assume that there are two populations 1 and 2, each with a different individual cumulative hazard $M_1(t)$ and $M_2(t)$ at age t , but having the same distribution of frailty Z at birth, which has $EZ = 1$ and $\text{Var}(Z) = \sigma^2 < \infty$ and a population function $g(s)$.

Now define

$$\begin{aligned} G_i(t) &= g[M_i(t)] & i = 1, 2 \\ N_i(s, t) &= M_i(t) - M_i(s) & i = 1, 2 \\ H_i(s, t) &= G_i(t) - G_i(s) & i = 1, 2 \end{aligned}$$

In words, for population i , $G_i(t)$ is the cumulative population hazard up to age t ; $N_i(s, t)$ is the cumulative individual hazard between ages s and t ; and $H_i(s, t)$ is the cumulative population hazard between ages s and t .

Suppose it is known a priori that in population 1 is *disadvantaged* relative to population 2 in the sense that its individual hazard is larger (never smaller) at all ages. In terms of cumulative hazard this is equivalent to

$$N_1(s, t) \geq N_2(s, t) \quad \text{for all } (s, t) \quad (24)$$

For example population 1 is Puerto Rico and population 2 is the United States and it is assumed that health care in the U.S. is better at all ages. Or population 1 are patients with a more advanced stage of some disease than patients in population 2. Population 1 might also be patients with some risk factor and population 2 patients without this factor.

Note that (24) implies that the two survival curves *may not cross*. This is because (24) implies that $M_1(t) \geq M_2(t)$ for all t , so that $G_1(t) \geq G_2(t)$ for all t . For the two survival curves it follows that $S_1(t) \leq S_2(t)$ for all t . If the survival curves do cross, it is impossible that the one population is truly *disadvantaged* relative to the other.

Now we look at survival curves not since birth, but from some advanced age s . Suppose that given survival up to s the survival curves of the two populations do cross. In terms of cumulative hazard this means that there are s and t such that

$$H_1(s, t) < H_2(s, t) \quad (25)$$

This is the case for example if life expectancy at higher ages is better for the disadvantaged population, as mentioned in the Puerto Rico example. It is a reversal at higher ages of the relationship (24). This is the kind of reversal that Vaupel was interested in.

If the two conditions (24) and (25) are met, an approximate lower bound can be derived for the variance of the frailty Z . Thus it can be proved, if the model is true, that there is some heterogeneity of frailty in the two populations.

It holds for $i = 1, 2$ that

$$\begin{aligned} H_i(s, t) &= g[M_i(t)] - g[M_i(s)] \\ &= g[M_i(s) + N_i(s, t)] - g[M_i(s)] \\ &= \chi(G_i(s), g[N_i(s, t)]) \\ &\simeq g[N_i(s, t)]e^{-\sigma^2 G_i(t)} \end{aligned} \quad (26)$$

where the last equation uses the approximation (18) to the effect function. By (26) and (24)

$$e^{\sigma^2 G_1(t)} H_1(s, t) \simeq g[N_1(s, t)] \geq g[N_2(s, t)] \simeq e^{\sigma^2 G_2(t)} H_2(s, t). \quad (27)$$

Now $G_1(s) - G_2(s) \geq 0$ because of (24), so if $G_1(s) - G_2(s) \neq 0$, (27) can be rewritten as

$$\sigma^2 \gtrsim \frac{\log[H_2(s, t)] - \log[H_1(s, t)]}{G_1(s) - G_2(s)} \quad (28)$$

This gives us a lower bound to σ^2 , which is positive by (25). Because it is given in terms of the population functions G and H it is easily calculated.

When using the lower bound (28) one should keep in mind that it is based on an approximation of the effect function for $\sigma^2 \rightarrow 0$. This means that it may be incorrect for high values of σ^2 . However it can well be used to exclude low values of σ^2 (especially $\sigma^2 = 0$), because the approximation is quite correct for these low values. Therefore (28) can be used to prove that under the conditions (24) and (25) there must be some heterogeneity in frailty.

Unfortunately such a proof could not be given for the example by Vaupel e.a. on the comparison between Puerto Rico and the United States. In the data used by Vaupel (*U.N. Demographic Yearbook* 1976) it was apparent that the survival curves for Puerto Rico and the U.S. did in fact cross. According to the data out of every 100,000 there are more Puertoricans who reach the age of 85 than Americans. This can not be reconciled with the assumption (24) that Puerto Rico is disadvantaged relative to the United States. This might be due to problems in the data, especially because the U.N. figures are not cohort data.

9 Conclusion

In this paper I have constructed a simple and usable model, which can be used in prognostic models to predict survival times of patients. It improves predictions by using comorbidity and disease history of the patients. There are two important aspects.

1. The model incorporates an *extra hazard effect* of previous tumors. This effect occurs when a patient suffers from two or more primary tumors simultaneously. Such a patient experiences more hazard than other patients. The model outlined in this paper incorporates this excess hazard and uses it to achieve a better prediction.
2. The model can also test whether there is a *selection effect* (overdispersion), which means that people who have survived much in the past live longer than others, or a *weakened survivor effect* (underdispersion) which means the opposite happens. When one of these effects is found to exist, it can also be incorporated to achieve a better prediction.

The model was applied to data on patients with a head/neck tumor, some of which had experienced a previous tumor. Here it was shown that there were no significant selection or weakened survivor effects, but that incorporating the extra hazard effect did greatly improve prediction.

The theory developed for the model concerns the relationship between frailty and relative survival and is also applicable in other circumstances. This was shown by formalizing a classic reasoning used to show the presence of frailty effects.

A Proofs of the lemma's

I will prove the lemma's about the population function stated in paragraph 3.2.

Lemma 1 *The population function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is invertible.*

Proof: Note that e^{-sz} is a monotone decreasing function of s for $z > 0$ and that we assumed that $Z > 0$ almost surely. It follows that $g(s) = -\log \mathbb{E}(e^{-sZ})$ is a monotone increasing function of s . By the monotone convergence theorem, for all $s_0 \geq 0$

$$\lim_{s \rightarrow s_0} g(s) = \lim_{s \rightarrow s_0} -\log \mathbb{E}(e^{-sZ}) = -\log \mathbb{E}\left(\lim_{s \rightarrow s_0} e^{-sZ}\right) = g(s_0) \quad (29)$$

so that $g(s)$ is continuous. Taking $s_0 = 0$ and $s_0 = \infty$ in (29) gives $g(0) = 0$ and $\lim_{s \rightarrow \infty} g(s) = \infty$. □

Lemma 2 *If $\mathbb{E}(Z^n) < \infty$ the population function $g(s)$, $s \geq 0$, is n times continuously differentiable.*

Proof: Let $k \leq n$ and $s > 0$. Then

$$\left|L^{(k)}(s)\right| = \left|\mathbb{E}((-Z)^k e^{-sZ})\right| \leq \mathbb{E}(Z^k),$$

where $\mathbb{E}(Z^k) < \infty$ if $\mathbb{E}(Z^n) < \infty$. The result for $g(s)$ follows trivially. Continuity can be proved by the monotone convergence theorem as in (29). □

Lemma 3 *If $m > 0$ or $\mathbb{E}Z < \infty$, $g'(m) = \mathbb{E}Z_m < \infty$ and $-g''(m) = \text{Var}(Z_m) \leq \infty$.*

Proof: First we use Jensen's inequality. Because x^p is a convex function for $x \geq 0$ when $p \geq 1$, while $-x^p$ is convex for $x \geq 0$ when $0 < p \leq 1$, Jensen tells us that

$$\begin{aligned} L_m(ps) = \mathbb{E}([e^{-sZ_m}]^p) &\geq [\mathbb{E}(e^{-sZ_m})]^p = [L_m(s)]^p & p \geq 1 \\ L_m(ps) = \mathbb{E}([e^{-sZ_m}]^p) &\leq [\mathbb{E}(e^{-sZ_m})]^p = [L_m(s)]^p & 0 < p \leq 1, \end{aligned}$$

so that

$$\begin{aligned} g_m(ps) &\leq pg_m(s) & p \geq 1 \\ g_m(ps) &\geq pg_m(s) & 0 < p \leq 1. \end{aligned} \quad (30)$$

Now

$$\begin{aligned}
g'(m) &= \lim_{h \downarrow 0} \frac{g(m+h) - g(m)}{h} \\
&= \lim_{h \downarrow 0} \frac{g[m(1 + \frac{h}{m})] - g(m)}{h} \\
&\leq \lim_{h \downarrow 0} \frac{(1 + \frac{h}{m})g(m) - g(m)}{h} \\
&= \frac{g(m)}{m} < \infty
\end{aligned}$$

and analogous for the left-side derivative.

Furthermore because $L_m^{(k)} = \mathbb{E}[(-Z_m)^k e^{-sZ_m}]$, taking $s = 0$ gives us $L'_m(0) = -\mathbb{E}(Z_m)$ and $L''_m(0) = \mathbb{E}(Z_m^2)$. We can now calculate

$$g'(m) = g'_m(0) = \frac{-L'_m(0)}{L_m(0)} = \mathbb{E}(Z_m)$$

and

$$g''(m) = g''_m(0) = \frac{[L'_m(0)]^2 - L_m(0)L''_m(0)}{[L_m(0)]^2} = -\mathbb{E}(Z_m^2) + \mathbb{E}(Z_m)^2 = -\text{Var}(Z_m).$$

□

Lemma 4 *The function $g(s)$, $s \geq 0$ is concave.*

Note that if $\mathbb{E}(Z^2) < \infty$ this lemma follows from the lemma's 2 and 3, because $g''(m) = -\text{Var}(Z_m) \leq 0$. The proof that follows does not need this condition.

Proof: Choose $0 \leq m < s$, $0 \leq p \leq 1$ and $q = 1 - p$. Now we use (8) and (30) to write

$$\begin{aligned}
g(ps + qm) &= g(m + p(s - m)) \\
&= g(m) + g_m[p(s - m)] \\
&\geq g(m) + pg_m(s - m) \\
&= g(m) + p[g(s) - g(m)] \\
&= pg(s) + qg(m).
\end{aligned} \tag{31}$$

□

Lemma 5 *For $0 \leq m < n$ and $s \geq 0$, $g_m(s) \geq g_n(s)$.*

Proof: For $u < v < w$, taking $p = \frac{v-u}{w-u}$ and $q = \frac{w-v}{w-u}$, (31) becomes

$$[(w-v) + (v-u)]g(v) \geq (v-u)g(u) + (w-v)g(w),$$

so that lemma 4 is equivalent to

$$\Delta_{u,v} = \frac{g(v) - g(u)}{v-u} \geq \frac{g(w) - g(v)}{w-v} = \Delta_{v,w}. \quad (32)$$

Furthermore it follows from (30) that for $u < v < w$

$$\begin{aligned} \Delta_{u,v} &= \frac{g_u(v-u)}{v-u} \\ &= \frac{g_u\left(\frac{v-u}{w-u}(w-u)\right)}{v-u} \\ &\geq \frac{v-u}{w-u} \cdot \frac{g_u(w-u)}{v-u} \\ &= \Delta_{u,w}. \end{aligned} \quad (33)$$

Now we note that $g_t(s) = s\Delta_{t,t+s}$ and the assertion that $g_m(s) \geq g_n(s)$ for $m < n$ can be proved. We split it up in three cases

1. The case $m + s = n$ follows trivially from (32).
2. If $m + s < n$ the assertion

$$g_m(s) = s\Delta_{m,m+s} \geq s\Delta_{m+s,n} \geq s\Delta_{n,n+s} = g_n(s).$$

also follows from (32).

3. Finally, if $m + s > n$ the assertion

$$\begin{aligned} g_m(s) &= g_m(n-m) + g_n(m+s-n) \\ &= (n-m)\Delta_{m,n} + (m+s-n)\Delta_{n,m+s} \\ &\geq (n-m)\Delta_{n,n+s} + (m+s-n)\Delta_{n,n+s} \\ &= g_n(s) \end{aligned}$$

follows from (8), (32) and (33).

□

A much simpler proof of lemma 5 can again be given under more restrictive conditions. If $m > 0$ or $EZ < \infty$ then $E(Z_m) < \infty$ so that the covariance $\text{Cov}(e^{-sZ_m}, e^{-tZ_m})$ is defined (though possibly infinite). This covariance is positive if $t, s \geq 0$. Therefore

$$L_m(t+s) = E(e^{-tZ_m} \cdot e^{-sZ_m}) \geq E(e^{-tZ_m}) \cdot E(e^{-sZ_m}) = L_m(t)L_m(s),$$

whence $g_m(t + s) \leq g_m(t) + g_m(s)$. So

$$g_m(s) \geq g_m(t + s) - g_m(t) = g_{m+t}(s)$$

by (8). This is equivalent to lemma 5 (take $n = m + t$).

References

- Odd O. Aalen, 1988, ‘Heterogeneity in survival analysis’, *Statistics in medicine* 7: 1121–1137.
- R. J. Baatenburg de Jong and S. le Cessie, ‘Prediction of survival in patients with head and neck cancer. *Head and Neck* (accepted for publication).
- R. J. Baatenburg de Jong and S. le Cessie, 2001, ‘Reliable prediction of survival in patients with lip and oral cavity cancer’ in: A. K. Varma and J. L. N. Roodenburg (ed.), *Oral Oncology*, vol. 7., 423–429. (MacMillan Press)
- Saskia le Cessie and Hans C. van Houwelingen, 1995, ‘Testing the fit of a regression model via score tests in random effects models’, *Biometrics* 51: 600–614.
- J. W. W. Coebergh, L. H. van der Heijden and M. L. G. Janssen-Heijnen (ed.), 1995, *Cancer incidence and survival in the Southeast of the Netherlands 1955–1994: a report from the Eindhoven Cancer Registry* (Eindhoven: IKZ).
- *Demographic Yearbook 1976*, (New York: United Nations).
- J. Estève, E. Benhamou, M. Croasdale and L. Raymond, 1990, ‘Relative survival and the estimation of net survival: elements for further discussion’, *Statistics in medicine* 9: 529–538.
- Guy Hédelin, 1995, *RELSURV: a program for relative survival* (Strasbourg: Louis Pasteur University).
- Philip Hougaard, 1984, ‘Life table methods for heterogenous populations: distributions describing the heterogeneity’, *Biometrika* 71: 75–83.
- Peter D. Sasieni, 1996, ‘Proportional excess hazards’, *Biometrika* 83 (1) 127–141.
- James W. Vaupel, Kenneth G. Manton and Eric Stallard, 1976, ‘The impact of heterogeneity in individual frailty on the dynamics of mortality’ *Demography* 16 (3) 439–454.
- Per-Henrik Zahl, 1997, ‘Frailty modelling for the excess hazard’ *Statistics in medicine* 16: 1573–1585.