ESTIMATING THE RISK OF RECURRENT OR MULTIPLE EVENTS IN LONGITUDINAL STUDIES

PIETRO RAVANI


CENTRE FOR NFLD. STUDIES
MEMORIAL UNIVERSITY
OF NEWFOUNDLAND

Memorial University of Newfoundland

Estimating the Risk of Recurrent or Multiple Events in Longitudinal Studies

by

Pietro Ravani

Candidate for Doctor of Philosophy Ph.D. - Clinical Epidemiology

Research and Graduate Studies

Faculty of Medicine

St John's NL – Spring 2009

Abstract

Background. Longitudinal studies usually evaluate risk by modelling time to first event using standard Cox's regression. However this method fails to utilize further outcome information after the first event. In patients with Chronic Kidney Disease repeated events of the same type (recurrent events) and or different type (multiple events) occur frequently. Risk estimation based on partial information may be inaccurate and imprecise. Yet, the analytical tool must take into account the lack of independence of the failure times.

Methods. To determine whether other methods of analysis are more informative and powerful than standard Cox's regression I re-evaluated data from previous research I had undertaken in Chronic Kidney Disease patients. Data from a multi-centre dialysis access study of incident hemodialysis patients were used as an example of recurrent failure events. Data from a cohort of pre-dialysis patients were used as an example of multiple competing events (dialysis start and death). Correlation in the data was accounted for using either robust variance methods or incorporating frailty effects into the model. Different approaches were used more or less free from distributional assumptions, including generalized models for counts, and using the robust version of the Cox's model as reference estimation method.

Results. The work shows that standard survival techniques that disregarded further information after the first event have limitations, in terms of power (precision of each estimate and number of estimated effects) and possibly of accuracy (bias). For example, the hazard ratio (HR) of primary failure of the first arterio-venous access for dialysis was 1.96 (95% Confidence Intervals 0.93 to 4.1) in presence of both history of heart failure and nephrology follow-up shorter than 3 months before dialysis start (effect of the interaction controlling for the main effects and other covariates). The estimate was more precise in the corresponding extended Cox's model for recurrent events (HR 2.02, 95% CI 1.11 to 3.65). Similar fits were obtained using variance corrected parametric models. However, all these variance corrected models did not take into account any random effects. This may have induced underestimation of the true effects if the frailty models were true (HR from the frailty Weibull model 3.5, 95% CI from 1.34 to 9). Improvement of model efficiency and more flexibility in model building were observed also using competing risk models for multiple events.

Conclusion. Analytical techniques for repeated events exist that make more efficient the use of longitudinal data while accounting for their correlation. These methods help address research questions about risk (and some also survival time) considering the entire course of a disease process or multiple possible outcomes, and have implications on design, implementation and costs of clinical research.

Acknowledgement

This thesis work represents the highest achievement in my career as a student, and a crucial step in my career as a researcher, clinician and teacher. There are some people without whom this accomplishment could not be achieved. Brendan Barrett was far more than a supervisor for me. He likes numbers and philosophical questions at least as much as I do, but always helped me to get them grounded in real world situations. Most importantly, he showed me the importance of balancing time for family, gardening, cooking, recreation, music, research, studying, teaching and clinical service. Amazing. Pat Parfrey showed me the importance of "getting things done", and how leadership can serve all members of a research group and make team work successful. Gerry Mugford and Sean Murphy coordinated the clinical epidemiology program and made Canadian bureaucracy less painful when I discovered it was not less complicated than the Italian. Bryan Curtis and Adeera Levin deserve a special recognition. They decided that Newfoundland had to be the place where I would spend the most amazing time of my life, and get to know the most genuine and generous people I have ever met. I will never forget the Christmas times my family enjoyed with Brendan, Mary, John, Eilish, Pat and Benvon, and their children. I am finally thankful to my new colleagues Brenda Hemmelgarn, Braden Manns and Nairne Scott-Douglas who acknowledged that this thesis work had to be a priority since I joined them in Calgary.

My wife, Anna, and my children, Marta, Attilio and Lorenzo, are the real reason this

thesis had to be completed. Anna provided patience, love, support, and nutrition to sustain me through my graduate degrees. She is my inspiration. My darling children often fell asleep in my arms while I was holding stats, epidemiology or nephrology books in my hands. And finally my parents. My father, Attilio, taught me the importance of "accuracy" and "precision" in restoring fine wood antiques, and how masterpieces can only be build on patience and sacrifice. And my mom, Alice, who fulfilled her dreams of completing high school by studying after work hours while I was in junior high: she instilled in me the passion for science, arts and culture, thirst for knowledge and truth, and love for hope and faith in the Reason I live my life.

TABLE OF CONTENTS **Page**

List of Tables

List of Figures

Figure 14: Observed primary and secondary recurrent event rate over time bands in months

List of symbols and abbreviations

AV	Arterio-Venous
HF	Heart Failure
LR	Late Referral to the nephrologist
CV	Cardio-Vascular
CKD	Chronic Kidney Disease
λ	Rate or hazard (instantaneous rate)
H	Cumulative hazard function
S	Survivorship function (survivor probability function)
RR	Risks Ratio
IRR	Incidence Rate Ratio
TR	Times Ratio
HR	Hazards Ratio
PH	Proportional Hazards
AFT	Accelerated Failure Time
AP	Acceleration Parameter (1/TR)
ρ	Correlation coefficient
σ^2	Variance of the residuals

θ	Variance of the random effects
ψ	Variance of the random errors when σ^2 is decomposed
SE	Standard error of the regression coefficient (fixed effect)
β	Regression coefficient
$X\beta$	Linear predictor (LP)
p	ancillary parameter (exponentiated coefficient)

Statistical notations

Standard Poisson model

Rates (λ) are estimated from event counts (d) over person-time (n):

$\lambda = d / n$

Before considering any covariate, the null Poisson model estimates the overall or unconditional rate λ as $\exp(\beta_0^*)$, corresponding to the intercept (overall mean) of a null linear model; thus $\beta_0^* = \log(\lambda)$. However, rates may vary by level of the exposure x: $\lambda_1 = d_1 / n_1$ when $x = 1$; and $\lambda_0 = d_0 / n_0$ when $x = 0$. The conditional Poisson model (conditional on "x"), is $\log(\lambda_i) = \log[E(d_i \mid x_i)] = \log(n_i) + \beta_0 + \beta_x x_i$, where $E(d_i \mid x_i)$ is the expected count given the covariate value, $\beta_0 = \log(\lambda_0)$ and $(\beta_x x_i)$ is the log(IRR) or natural log of the incidence rate ratio (λ_1 / λ_0).

In fact, $\log(IRR) = \log(\lambda_1) - \log(\lambda_0) = \log(d_1) - \log(n_1) - \log(\lambda_0)$.

Proportional hazard regression

Instantaneous rates are modeled as a function of the covariate vector X:

$\lambda(t) = \lambda_0(t)\exp(X\beta)$, where λ_0 indicates the baseline hazard (when all covariates are zero)

If no distribution is specified for $\lambda_0(t)$, the model is semi-parametric (Cox's regression); vice versa additional parameters are estimated that fully specify the hazard (parametric models).

Accelerated failure time models

$Log(t) = -X\beta + log(\tau)$, where the observed log-time $log(\tau)$ can follow several parametric distributions (exponential, Weibull, log-logistic, log-normal, gamma, for example). This model is time proportional, i.e. $exp(\beta)$ = time ratio (TR).

Frailty models

These models (within the Poisson or survival family) add a variance component to the variability in the data unexplained by fixed effects. The random effect is estimated as an additional term of the linear predictor (coefficient of the random factor defining groups and / or subjects).

1) Chapter 1: Problem statement and thesis outline

Chapter overview

This first chapter introduces the issues of repeated events in longitudinal studies, and provides a general overview of the thesis work.

1. Introduction

i. Problem to be addressed

Longitudinal designs are often used to measure observation time and record event occurrence during the study period. The risk of a failure event (or the probability of a successful outcome such as disease remission) can be estimated from such measurements using different analytical techniques. One common complication of longitudinal designs is that the event of interest may recur or compete with other events. This is the case, for example, when repeated failure events may recur in the same subject (e.g., repeated urinary tract infections or repeated flares of an inflammatory disease) or when competing risks are correlated (e.g., dialysis and cardiovascular events in subjects with Chronic Kidney Disease). In these circumstances, data are not independent because they come from the same individuals, and standard regression techniques can not be used unless only partial information is considered such as one single observation per subject.

In the recent years I have carried out analyses of several longitudinal studies of patients with kidney disease. Before enrolling in the PhD program I extensively applied standard survival methods for risk estimation (1 – 4). However, as I will illustrate in this thesis, in nephrology, as in other areas of clinical research, correlation of risk data is often the rule rather than the exception. Examples of where this may arise in nephrology include repeated episodes of acute kidney injury in subjects with cardiovascular disease undergoing angiographic procedures, repeated peritonitis in peritoneal dialysis patients,

rejection or infection episodes in kidney transplant patients, and dialysis catheter infections or dysfunction and arterio-venous access thrombosis in hemodialysis patients. The presence of correlation in such data is related to the existence of repeated or multiple observations from the same individuals, which requires more sophisticated statistical approaches. Further examples are offered in the second section of this introductory chapter.

During my PhD program, I completed other studies addressing a variety of questions relevant to understanding kidney diseases, including competing risks (5 – 8). For the purpose of this thesis I have reviewed some of the dialysis access data I already published using standard techniques for single observations per subject (3), and compared them with more appropriate analytical approaches to repeated event processes. I also expanded on competing risks in a Chronic Kidney Disease longitudinal cohort study (8).

ii. Work outline

The second chapter of the thesis introduces standard regression methods for risk estimation, and some challenges posed by longitudinal designs. Some of the content of this and subsequent chapters (chapter 3 and 4) was published as a series of 5 papers in Nephrology Dialysis and Transplantation of which I was the first and corresponding author (9 – 13). Standard methods are appropriate for the analysis of independent events, such as the first failure episode of a recurrent process (e.g., repeated hospitalizations) or

one event only from multiple possible failure episodes of different type (e.g., dialysis and myocardial infarction). These standard methods can be used to model count outcomes or survival times, using non parametric, semi-parametric and fully parametric regression models.

The subsequent chapters introduce the general approach to correlated data, the possible sources of correlation, and the theory and implications of the main analytical methods for survival and count data (chapters 3 and 4). Chapters 5 and 6 provide examples of modelling using previously published data. Results are presented and discussed along with the strength and weaknesses of each possible approach, and its consistency with the study question and design. Each of these analyses illustrates the methodological challenges in analysing longitudinal data and the solutions that may be considered to deal with them. On the basis of my findings I conclude this thesis work with some comments and recommendations for researchers facing similar methodological challenges as to how best deal with these problems (chapter 7).

iii. Study data discussed in this thesis work

To introduce standard semi-parametric and parametric regression models (chapter 2) I used examples from an Italian multi-centre study on hemodialysis vascular accesses (3). Study definitions, methods and results are reported at the beginning of the chapter. I designed the study and published the data as principal and corresponding author. As

mentioned above, vascular access thrombosis may recur in the same individual several times as one patient may have several vascular accesses. However, at the time of the study I was unaware of the existence of techniques capable to account for the correlation of repeated observations per subject, and only the analysis of the first access was reported in the paper. In subsequent years I studied this correlation problem and found ways to use Cox's regression for both repeated failure events of the same type (such as dialysis access thrombosis) and of different types (such as dialysis and death in chronic kidney disease patients). More recent publications are based on the analysis of time varying exposures and competing risks using Cox's regression (5 – 8). I designed and analysed the data from all of these studies, being the first author in three of them. One of these studies evaluates the impact of Vitamin D levels on the risk of renal and patient death and it is used as an example of competing risks in chapter 5 (8). Also for this study, definitions, methods and results of the main publication are described in chapter 5. For semi-parametric (chapter 5) and parametric (chapter 6) analyses of repeated events of the same type, I used the vascular access data as examples, including information not included in the original published report (3). The same data are used also to run count models. This may represent a limitation, as censored survival data may contain excess "zero counts" when using event history data as opposed to aggregated data.

iv. Statistical packages used for the analyses

The statistical packages I used for this thesis work are STATA and R.

STATA: StataCorp. 2007. Stata Statistical Software: Release 10. College Station, TX: StataCorp LP; web site: www.stata.com

R: R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org. Website for software download: http://cran.r-project.org/

Both packages offer an easy programming interface based on text files. STATA is more friendly for management of large data sets. R (which is a "dialect" of S-Plus) allows more flexibility of use and generates better plots. For example, at the time of this thesis work one of the limitation of STATA is the impossibility to run stratified frailty models (either semi-parametric or parametric survival models). One of the greatest advantages of R is the free source, including several updates per year. Both have free online educational resources including discussion group and technical support.

What I learnt after my initial experience with other packages and from this thesis work, is that serious analysis cannot be done with only one statistical package or avoiding some minimal programming work. Both STATA and R are excellent tools and together offer most of the necessary support for risk estimation in longitudinal studies.

2. Correlated data

i. Study designs generating correlated data

Much medical research can be simplified as the study of an exposure-disease relationship (9, 10). In some study designs the association between exposure and disease is assessed more than once on the same subject or group of people. Multiple measures can be taken under different experimental conditions in cross-sectional studies or longitudinally over time. For example, in a *cross-sectional* study of endothelial function, brachial artery flow rates can be measured in the same subject under exposure to different vasoactive substances. Researchers are interested in differences in flow rates by level of the exposure (i.e., the effect of one or more vasoactive drugs versus control or reference standard). Unbiased estimates of such effect(s) can be obtained removing from the overall (unconditional) variation of the response both the variability between subjects due to measured subject level characteristics affecting the outcome, and the within subject variability due to the dependencies of the repeated measures taken on the same subject. The residual variability (random error variability) is what remains to be explained after the model has been fitted. Failure to account for the "extra-variability" in the data may result in biased estimates of the effects of interest and their standard errors. This variance component approach to the analysis of correlated data can be extended also to risk data.

Longitudinal studies typically monitor participants over time and both predictors (e.g., blood pressure) and outcomes (e.g., left ventricular mass index) are measured in different

occasions in the same subject. In some designs, observations can fall into groups (clustered data). This occurs when single measurements are taken on a paired organ (e.g., the eye or the kidney) or on different members of the same hospital, region or family. More complex designs may lead to a combination of clustering and repeated / longitudinal measuring. For example, a longitudinal multi-centre study of non fatal cardiovascular (CV) events in patients with Chronic Kidney Disease (CKD) will generate outcome data recurring in the same subject and in the same centre. Table 1 shows examples of how data generated by these study designs can be set up for analysis. In all these longitudinal designs part of the variability of the response under study is due to unobserved cluster or unmeasured subject level characteristics that can and should be taken into account.

ii. Origin of correlation

Outcome data generated from such clustered and / or repeated / longitudinal study designs are correlated because it is possible to identify patterns of association within individuals or clusters. In other words, multiple measurements on the same subjects or single assessments of paired organs or members of the same hospitals, region or family generate values that are closer than those obtained from different individuals, organs, hospital, regions or families. In fact different organs of the same subject and different individuals of the same community share biologic experiences, environmental exposures and genetic background. Factors underlying such correlation are often unknown or

Table 1: Examples of correlated (panel) data sets: Each study participant can be assessed once or in several occasions, in random sequence (repeated measures) or over time (longitudinal data). Each row in the panels represents an observation, with single measurements per subject (left) or updated values (right) of both predictors (X) and outcomes (e.g., mean arterial pressure values – Y). Predictors can be time invariant (such as gender, age at baseline or presence of diabetes – Xti) or time varying, i.e. assume different values (e.g., glucose – Xtv). In either case, observations can belong or not to clusters (CP / CA – present / absent), such as families, schools, or hospitals (lower panel). In addition, there can be multilevel data, when clusters are nested in super-clusters, such as patients (Pt) in physicians (MD) in hospitals (H). In these situations, level 1 is the most detailed level (the single observation); level 2 the epidemiological unit (the patient); and level 3 or higher the next level of hierarchy (membership level).

	Single measurements					**Repeated / longitudinal measures**					
	H	MD	Pt	X	Y	H	MD	Pt	Xti	Xtv	Y
CA	.	.	1	0	123	.	.	1	0	90	123
	.	.	2	1	120	.	.	1	0	93	125
	.	.	3	0	118	.	.	2	1	110	120
	.	.	4	1	100	.	.	2	1	105	122
CP	1	1	1	0	123	1	1	1	0	90	123
	1	1	2	1	120	1	1	1	0	93	125
	1	2	3	0	118	1	1	2	1	110	120
	1	2	4	1	100	1	1	2	1	105	122
	1	3	5	0	123	1	2	3	0	90	123
	1	3	6	1	120	1	2	3	0	93	125
	1	4	7	0	118	1	2	4	1	110	120
	1	4	8	1	100	1	2	4	1	105	122

unmeasured, and may confound the observed relationship of interest. Furthermore, the degree of correlation in the data can vary and its sign is not necessarily positive. Indeed for some outcomes such as infection disease recurrence, previous experience may induce negative correlation. In all cases, once a measurement value has been obtained further values within the same individual or cluster can be more accurately "guessed". In other words, the within individual / cluster variance differs from the between individual / cluster variance. This implies that the corresponding errors – or deviates from the mean – are no longer due to chance alone causing excess variability in the data or heterogeneity (Figure 1). In these situations traditional regression methods are inappropriate as they assume independent errors (11, 12).

iii. Analytical approaches

Two major analytical approaches exist for the analysis of correlated data: random effect modelling and variance corrected methods (13). The main assumption underlying these approaches is that the responses are correlated within subject / cluster, but independent between subject / cluster. These two approaches can be applied to outcome of different types, i.e. generalized linear (continuous, binary and count outcomes) as well as time-to-event (survival) data. The present thesis work will consider both generalized linear models for count outcomes and survival analysis techniques because risks in longitudinal studies can be estimated from either rates of event count or survival times depending on the way the response has been measured. However, it is important to note that all models

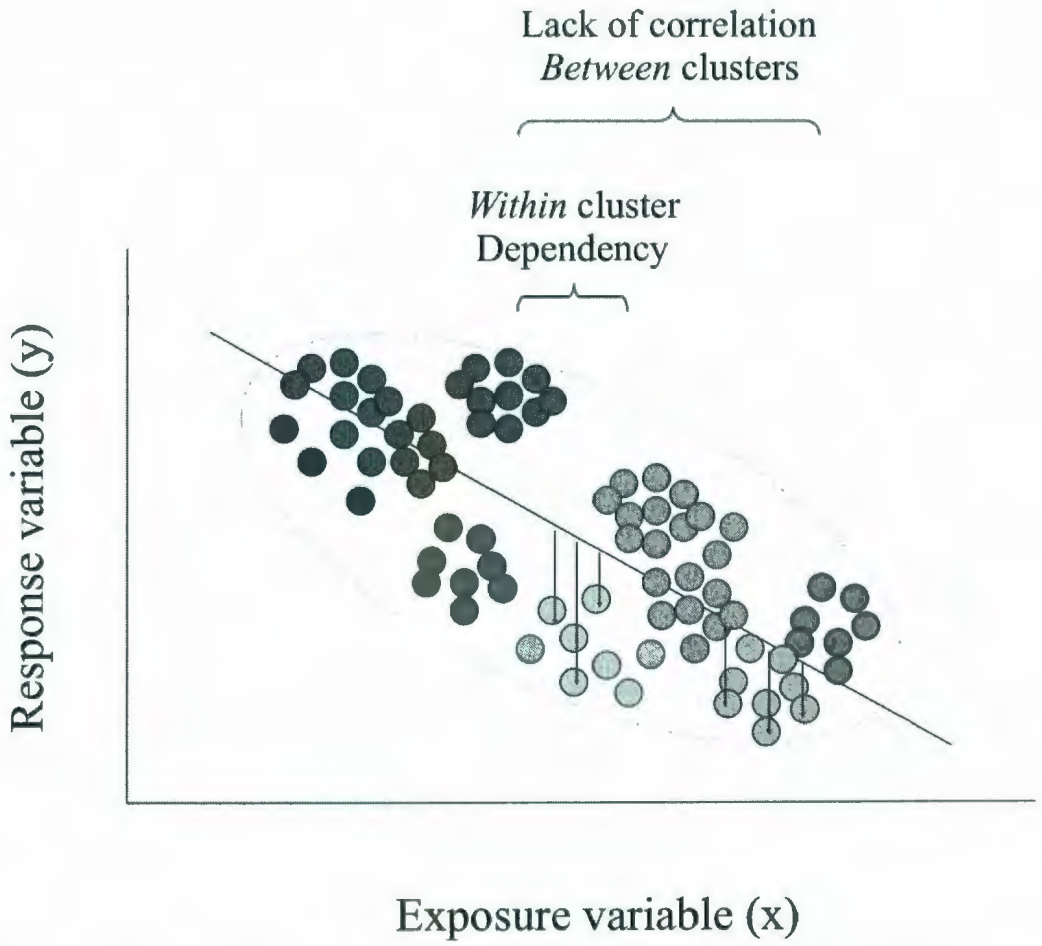


Figure 1: Repeated observations within the same individual or single measurements obtained from clusters (groups of dots) increase the heterogeneity of the study sample (larger oval delimited by a dashed line). Residuals (deviates of each observation point from the mean – continuous straight line) are dependent within individuals or clusters and independent between individuals or clusters. Failure to acknowledge these dependencies and group heterogeneity may result in biased estimates of the model coefficients (in the example the intercept and the slope of the line describing the average change in y as x changes).

for rates and some survival models are "parametric models". For example, the Poisson process is characterized by a slow and constant event rate with event count variance equal to the mean. Parametric survival models assume that the observed failure times have been randomly drawn from some specific distributions. Since these assumptions are often violated or some requirements not met, the Cox's model has become the most popular approach to estimate risks in longitudinal studies. The Cox's model is in fact a "semi-parametric" model because the effects (relative risks) are estimated without making any assumption about the distribution of the failure times. When the observed risk distribution is well described by a specific parametric model (and the underlying biological and clinical process are consistent with how the risk is supposed to vary over time according to that distribution), that model will make better use of the data in terms of efficiency (precision).

The following sections will summarize the main principles of traditional risk estimation methods in longitudinal studies using examples from previous publications (1 – 4). Alternative approaches for repeated or multiple events are subsequently introduced and applied to the same data or data from other studies (5 – 8).

2) Chapter 2: Standard modelling approaches

Chapter overview

Chapter 2 introduces general principles of risk estimation in longitudinal studies, using either count data or survival data. The data used as example are from the vascular access study, which is briefly summarized at the beginning of the chapter (3). Concepts of risk and rates are also explained as well as ways to estimate risk using semi-parametric methods and fully parametric methods.

1. The Italian multi-centre hemodialysis access study

i. Study description

The Italian Multi-centre Hemodialysis Access study data consists of 535 incident hemodialysis patients receiving an arterio-venous (AV) access for the first time (3). The study was designed to study the association between timing of referral to the nephrologist and the risk of AV access failure, controlling for baseline characteristics and presence of comorbid conditions. During a 6 year follow-up these patients received 633 AV fistulae and 67 grafts (700 AV accesses). The survival data from these individuals record up to 4 recurrence times: 404 individuals received one AV access, 101 received 2 accesses, 26 received 3 accesses and 4 received 4 accesses.

ii. Arteriovenous accesses for hemodialysis

An AV fistula is a communication between an artery and a vein usually created in the arm (at the wrist or at the elbow crease). The new vessel keeps some of the desirable properties of the vein and the artery: 1) it is superficial with elastic walls which make it suitable to repeated needling, and 2) it carries high blood flow rates. These are both necessary conditions for maintenance hemodialysis. Alternatively, an AV graft is a communication between artery and veins through a prosthetic bridge made of synthetic (as in this study) or natural material (human or bovine veins). Also a graft is repeatedly cannulated and carries high blood flow rates. However, it contains a foreign material and its use has been associated with increased risk of death from infection and cardiovascular

diseases. Once created an AV access can fail for several reasons leading to its thrombosis (primary failure). If possible, revision interventions (either surgical or radiological) are undertaken to salvage the malfunctioning or failing AV access. When they are successful the access patency is restored and its (secondary) survival extended (14 – 21).

iii. Study outcomes

In the access study the primary AV access creation was successful in 313 cases. During the study period 310 primary (unassisted) AV access failures occurred: 222 failures of the first AV access created, 72 of the second, 13 of the third and 3 of the fourth. Three hundred and fifty nine subjects used their first AV access for the duration of the study including those who underwent at least one salvage procedure. There were 245 secondary (final) access failures (i.e., failures despite revisions): 176 subjects lost their first AV access, 55 lost 2 AV accesses, 11 lost 3 and 3 lost 4 AV accesses (Figure 2).

iv. Analysis of the first event only

When an event can recur, such as the thrombosis of an AV access, or multiple events can occur such as myocardial infarction or death, longitudinal studies can generate complex data. One approach to the analysis of such data is to observe each individual until the first event that occurs, disregarding further events after the first. In the case of events of different types, studies often estimate the risk of combined end-points considering only

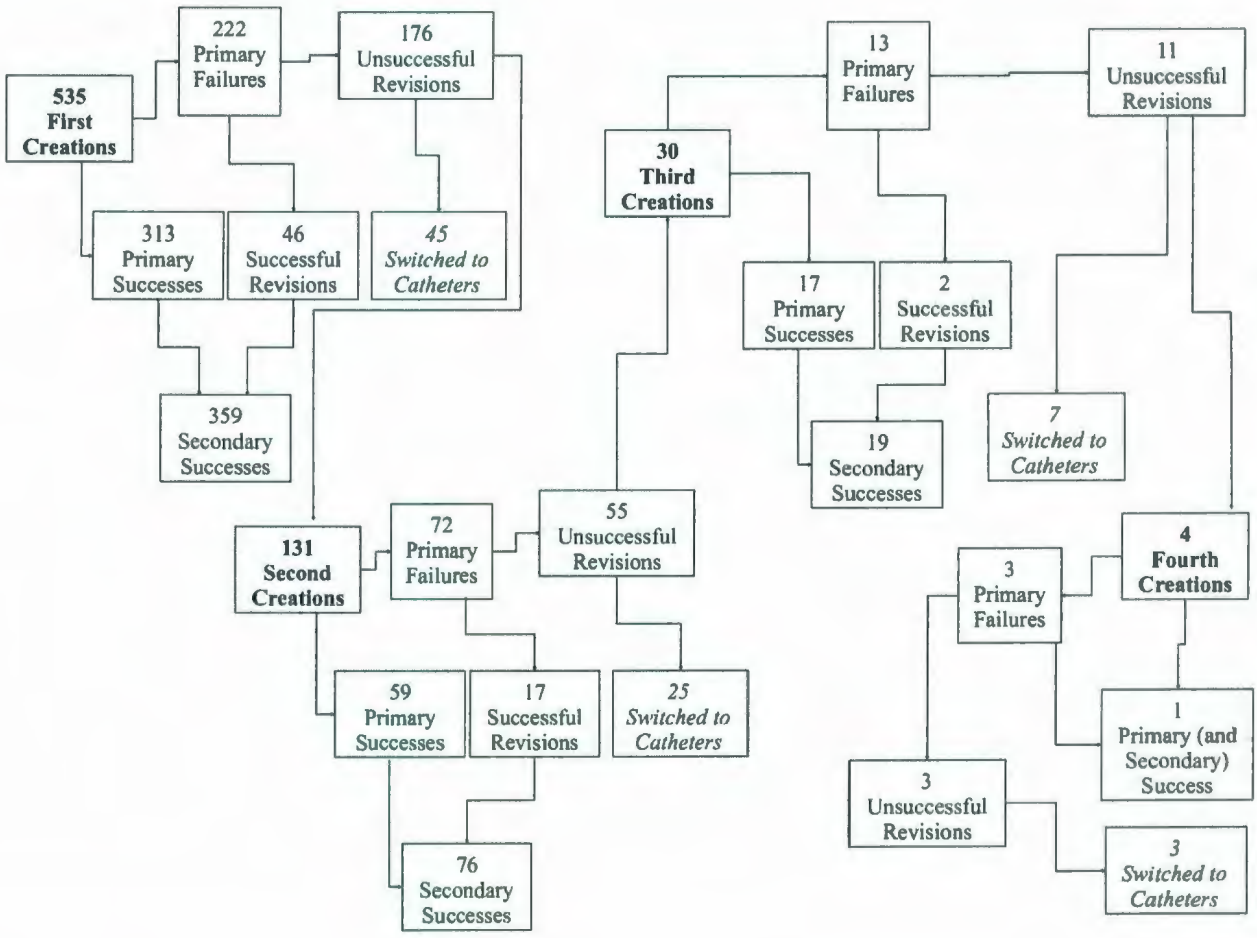


Figure 2: Outcomes of the arterio-venous accesses from the access study (3). Up to 4 attempts were made in 535 subjects. The possible outcomes of each intervention are primary success or primary failure. Failed AV accesses may or may not be salvaged surgically or with interventional radiology procedures. The final (secondary) lost of an AV access may be followed by a further attempt or switch to central venous catheter.

the first of two or more possible events (composite end-point). The common approach to the analysis of the fistula data reflects this tendency and is based on the standard Cox's model (1 – 4). As a result such studies estimate the risk of failure of the first AV fistula only disregarding further study data. For example, in the Italian AV study those few subjects (N = 22) who received an AV graft as first access were excluded for power reasons. An important consequence of this approach is that the study question addressed is: "Which factors affect the risk of failure of *the first AV fistula* in incident hemodialysis patients?". Standard approaches for the analysis of survival or count data can be used to answer this question. However, the information regarding small groups (e.g., AV grafts) and further accesses created in the same patient is ignored by such methods.

2. Estimating the risk of independent events from count and survival data

Risk estimation in longitudinal studies can be carried out using regression modelling for count outcomes or time-to-event analysis of survival data. Count outcome regression methods can be used both for aggregated as well as individual data. In fact, counts can be thought of as aggregated versions or summaries of more detailed data on occurrences of some kind of event (event-history data). Survival analysis is applicable only if times have been measured for each individual. Individual and aggregated data set up examples from the access study are shown in Tables 2 and 3.

Table 2: Individual data set up example (access survival data): multiple records per subject

ID	N	LR	HF	VAS	AVG	P_f	P_start	P_stop	S_f	S_start	S_stop
1	1	No	No	No	No	Yes	02/01/96	04/02/96	No	02/01/96	05/20/98
2	1	No	No	No	No	Yes	03/15/97	03/16/97	Yes	03/15/97	03/16/97
2	2	No	No	No	No	Yes	03/20/97	03/21/97	Yes	03/20/97	03/21/97
2	3	No	No	No	No	No	03/22/97	03/24/97	No	03/22/97	03/24/97
.	.	.	.	.	.	.	.	.	.	.	.
10	1	Yes	No	Yes	No	Yes	06/21/98	06/28/98	Yes	06/21/98	06/28/98
10	2	Yes	No	Yes	No	Yes	07/05/98	07/07/98	Yes	07/05/98	07/07/98
10	3	Yes	No	Yes	Yes	Yes	07/12/98	07/14/98	Yes	07/12/98	07/14/98
10	4	Yes	No	Yes	Yes	Yes	07/22/98	07/23/98	Yes	07/22/98	07/23/98
.	.	.	.	.	.	.	.	.	.	.	.

The table shows 8 of the 700 records from the access survival data. Id is the patient identifier and N is the arteriovenous access number: patient 1 had 1 access only, patient 2 had 3, and patient 10 had 4 accesses created during the study period. The independent variables are all binary (i.e., yes / no): LR = late referral (referral to the nephrologist less than 3 months of the dialysis start); HF = history of heart failure; VASC: presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events); and G = arterio-venous graft (as opposed to AV fistula). P_f and S_f are the binary event variables defining the failure status at follow-up end (primary and secondary failure). P_Start, P_Stop, S_Start and S_Stop are the dates defining the beginning (time zeros) and end of the primary and secondary survival times.

Table 3: Aggregated data set up example (access count data) including multiple accesses per subjects

LR	HF	VASC	AVG	P_MOS	P_CNT	S_MOS	S_CNT
No	No	No	No	4755.95	94	5718.45	75
No	No	No	Yes	250.87	10	384.88	6
No	No	Yes	No	1990.25	60	2415.62	49
.	.	.	.	.	.	.	.
Yes	Yes	Yes	No	158.35	14	215.42	12
Yes	Yes	Yes	Yes	14.24	3	14.24	3
.	.	.	.	.	.	.	.

The table shows 5 of the 15 possible combinations of the selected independent variables for the count outcomes (primary event count – P_CNT, and secondary event count – S_CNT) and exposure times in months (primary survival time – P_MOS and secondary survival time – S_MOS). These independent variables are all binary (i.e., yes / no): LR = late referral (referral to the nephrologist less than 3 months of the dialysis start); HF = history of heart failure; VASC: presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events); and AVG = arterio-venous graft (as opposed to AV fistula).

Table 4-7 show the results of two regression models including the following predictors (all nominal binary variables coded "yes" or "no"): late referral to the nephrologist (resulting in less than 3 months exposure to specialist care prior to dialysis commencement), history of heart failure, presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events) and arterio-venous graft (as opposed to AV fistula). Other variables were not significant predictors of event occurrence (or recurrence) in any model (including centre) and are not considered further for simplicity. One important finding of the study was that the maturation time (i.e., the time span between creation and first use in those 414 subjects whose AV fistulae were cannulated at least once) and the use of catheters were possible intermediate variables in the pathway linking late referral and the risk of primary and secondary failure respectively. Since these covariates were measured only for the first access, they will also be omitted to avoid exclusion of subjects from the analysis and allow comparison of different statistical approaches.

The two models shown in Table 4-7 are standard models of independent events estimating the risk of failure of the *first* AV access created in each study participant. Each model passed standard checking tests, including diagnostics, assumption verifications, and goodness of fit, sensitivity and residual analyses, which will not be discussed here.

Table 4: Cox's models of the first access primary survival times in the access study

MODEL 1: Observations 535; event number 222 (time at risk 8633.775 months)

Co-variate	**HR**	**P>|z|**	**95% Confidence Interval**
LR	1.316	0.096	(0.953 to 1.817)
HF	1.138	0.546	(0.748 to 1.731)
HXL	1.957	0.064	(0.961 to 3.985)
VASC	1.347	0.033	(1.025 to 1.771)
AVG	1.181	0.613	(0.620 to 2.248)

MODEL 2: Observations 513; event number 212 (time at risk 8415.458 months)

Co-variate	**HR**	**P>|z|**	**95% Confidence Interval**
LR	1.508	0.006	(1.123 to 2.024)
CVD	1.487	0.004	(1.135 to 1.948)

Legend: HR = hazard ratio; P>|z| two sided P value of the Wald test on the coefficients; LR = late referral (referral to the nephrologist less than 3 months of the dialysis start); HF = history of heart failure; HXL interaction term between HF and LR; VASC: presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events); CVD: presence of any cardiovascular disease (HF and / or VASC); and AVG = arterio-venous graft (as opposed to AV fistula). In the second model subjects receiving an AVG were excluded (N = 22).

Table 5: Cox's models of the first access secondary survival times in the access study

MODEL 1: Observations 535; event number 176 (time at risk 10531.63 months)

Co-variate	**HR**	**P>|z|**	**95% Confidence Interval**
LR	1.418	0.060	(0.985 to 2.043)
HF	1.339	0.213	(0.846 to 2.118)
HXL	1.564	0.260	(0.718 to 3.404)
VASC	1.408	0.029	(1.035 to 1.916)
AVG	0.840	0.677	(0.369 to 1.910)

MODEL 2: Observations 513; event number 170 (time at risk 10137.84 months)

Co-variate	**HR**	**P>|z|**	**95% Confidence Interval**
LR	1.594	0.005	(1.150 to 2.210)
CVD	1.674	0.001	(1.237 to 2.267)

Legend: HR = hazard ratio; P>|z| two sided P value of the Wald test on the coefficients; LR = late referral (referral to the nephrologist less than 3 months of the dialysis start); HF = history of heart failure; HXL interaction term between HF and LR; VASC: presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events); CVD: presence of any cardiovascular disease (HF and / or VASC); and AVG = arterio-venous graft (as opposed to AV fistula). In the second model subjects receiving an AVG were excluded (N = 22).

Table 6: Poisson models of the first access primary failure counts in the access study

MODEL 1: Observations 12; event number 222 (8633.775 person-months)

Co-variate	**IRR**	**P>|z|**	**95% Confidence Interval**
LR	1.535	0.009	(1.112 to 2.119)
HF	1.116	0.608	(0.735 to 1.695)
HXL	2.180	0.032	(1.068 to 4.448)
VASC	1.526	0.003	(1.160 to 2.007)
AVG	1.640	0.131	(0.862 to 3.120)

MODEL 2: Observations 8; event number 212 (8415.458 person-months)

Co-variate	**HR**	**P>|z|**	**95% Confidence Interval**
LR	1.839	0.016	(1.120 to 3.018)
CVD	1.641	0.033	(1.041 to 2.587)

Legend: IRR = incidence rate ratio; P>|z| two sided P value of the Wald test on the coefficients; LR = late referral (referral to the nephrologist less than 3 months of the dialysis start); HF = history of heart failure; HXL interaction term between HF and LR; VASC: presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events); CVD: presence of any cardiovascular disease (HF and / or VASC); and AVG = arterio-venous graft (as opposed to AV fistula). In the second model subjects receiving an AVG were excluded (N = 22).

Table 7: Poisson models of the first access secondary failure counts in the access study

MODEL 1: Observations 12; event number 176 (10531.63 person-months)

Co-variate	IRR	P>|z|	95% Confidence Interval
LR	1.597	0.012	(1.111 to 2.297)
HF	1.373	0.174	(0.869 to 2.169)
HXL	1.585	0.244	(0.730 to 3.444)
VASC	1.499	0.010	(1.102 to 2.038)
AVG	0.879	0.759	(0.387 to 1.999)

MODEL 2: Observations 8; event number 170 (10137.84 person-months)

Co-variate	HR	P>|z|	95% Confidence Interval
LR	1.832	0.005	(1.201 to 2.794)
CVD	1.760	0.005	(1.189 to 2.606)

Legend: IRR = incidence rate ratio; P>|z| two sided P value of the Wald test on the coefficients; LR = late referral (referral to the nephrologist less than 3 months of the dialysis start); HF = history of heart failure; HXL interaction term between HF and LR; VASC: presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events); CVD: presence of any cardiovascular disease (HF and / or VASC); and AVG = arterio-venous graft (as opposed to AV fistula). In the second model subjects receiving an AVG were excluded (N = 22).

The results of primary survival analysis (Table 4) indicate that some of the effects of interest were of borderline statistical significance when detailed information on specific components of cardiovascular disease was used. Insufficient data were available to study the risk associated with AV grafts (model 1). Both late referral and presence of any cardiovascular disease independently predicted worse outcomes, but their statistical interaction was not significant (model 2). Relative risks were even higher for secondary failure models (Table 5), but again specific components of cardiovascular disease did not reach statistical significance. Poisson analyses of aggregated data show different results (Table 6 and 7): a) the risk of primary failure increases over-multiplicatively if subjects with history of heart failure are referred late; b) there does not seem to be any interaction on the risk of secondary survival; and c) relative risks associated with late referral and presence of cardiovascular disease may be greater than those estimated in time-to-event models. These differences between traditional survival analysis and count analysis using the two most popular models, i.e. the Cox's model and the Poisson model, need to be explained before discussing how their extensions can be applied to analyse recurrent or multiple events properly.

3. Estimating rates from count data

i. Models for count responses

A popular approach to estimate risks in longitudinal studies is to consider the outcome as

a count variable of independent events. Poisson regression, for example, is appropriate when the underlying risk for such events is small and constant, but the number of individuals is large, and thus the total number of events is considerable. The outcome variable is counted over a period of time-at-risk – the principal co-variate in the model (exposure time), which is recorded for each observation or aggregated data. Standard methods for count outcomes assume that events are independent, that the risk experience of the subjects when they are under observation reflects their true risk of failure (i.e., when they leave the study their risk does not change), that the follow-up duration is based on disease severity and that the risks for subjects recruited early and late in the study are similar. These are standard validity issues to consider in all longitudinal designs (9, 10).

ii. Risks and rates

The analytical methods for both count and survival time outcomes do not model risks but rates. As opposed to risks (event count / persons during a specified period of time) which are dimensionless and range from 0 to 1, rates (λ = event count / person-time) have the dimension of 1 / time and range from 0 to $+\infty$. Risks can be estimated directly in short studies where subject follow-up is approximately complete (e.g., studies of contrast induced nephropathy). Rates are estimated in longer studies because as the study duration increases, fewer subjects have complete follow-up. Rates treat one time unit as equivalent to another, regardless of which individual they come from (e.g., one person observed for 10 years and another for 20 years would contribute for a total of 30 person-

years of follow-up or 30 persons per unit time). Depending on the chosen time unit, the same rate can have different numerical values and can exceed 1 (100%). For example, if 8 cases occur among 36 subjects in 1 month, then the same rate can be expressed as 0.22 cases per person-month or 2.66 cases per person-year (9, 10). Finally, rate estimates are unaffected by the precision of the measurement time scale. If 3 deaths are observed in 10 subjects and their exposure times are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 months (bands of 1/12 = 0.0833 years duration), the event rate is 3/(55*0.0833) = 0.655 per year = 655 deaths per 1000 person-years. Had the data been updated every day, then the length of the band would have been 1 day and there would have been 55*(365.25/12) = 1674 bands (units of n) of 0.002737 year duration. However, the rate would have been the same, 3/(1674*0.002737) = 0.655 per year.

Despite these differences incidence rates can be used to estimate risks. In fact if the underlying risk is constant and small (e.g., less than 0.2) it can be estimated as the product of the estimated rate and the observation time. For example, if 1000 subjects are followed for 10 years and experience a mortality rate of 0.01 per person-year (0.01 $year^{-1}$), the risk can be estimated as 0.01*10 or 0.1 over 10 years (each individual has a probability of 10% to die in 10 years). However, as deaths occur over time the same mortality applies to a steadily smaller population at risk. Since this population shrinking is neglected in the calculation, the risk approximation of the incidence rate does not work well for high risks or very long observation times. Fortunately, risks of interest to

clinical epidemiologists are usually small and studies not too long. Similarly incidence rate (IR) ratios are interpretable as risk (R) ratios since $R_1 / R_1 = (IR_1 * time) / (IR_0 * time) = IR_1 / IR_0$. An example will clarify why this is important to Poisson regression. Suppose that "d" AV access failure (independent) events are observed during "n" person-years, where d is small as compared to n. For example, considering only the first AV access attempted in the access study, 222 primary failure events occurred to 535 individuals over 8633.774 access-months of follow-up. The observed incidence of thrombosis was $\lambda_1 = d_1/n_1 = 68 / 1673.183 = 0.0406411$ month^{-1} or 40.64 per 1000 access-months in patients referred late to the nephrologist [1] and $\lambda_0 = d_0/n_0 = 154 / 6960.591 = 0.0221246$ month^{-1} or 22.12 per 1000 access-months in patients followed for longer than 3 months before dialysis start [0]. The incidence rate ratio of the two groups is $IRR = \lambda_1 / \lambda_0 = 1.83$. Poisson regression can be used to estimate the IRR associated with "one unit change" of the predictor (11, 12). Poisson regression (and its extensions) can be used both for aggregated count data and for individual times data. When the outcome of interest is a binary event, the output of Poisson regression is the same as the output of a parametric survival model called "exponential model". In fact, when the event rate is constant over time the distribution of the survival times is an exponential function of time, because $S(t) = e(-\lambda * t)$, i.e. survival probabilities are an exponential function of this constant hazard multiplied by time.

4. Estimating rates from survival data

i. Measuring survival data

The outcome variable of survival analysis is called survival time, although it may be applied to the time 'survived' from complete remission to disease relapse or progression as equally as to the time from diagnosis to death. Outcome measurement implies precise definition of the event of interest and when the period of observation starts and finishes. For example, in the fistula data, time was recorded from the creation date of the AV access (time zero), and the observation continued for each subject until either a failure event occurred (including recurrent thrombosis episodes in multiple AV accesses), the study ended, the patient died or was transplanted, or further observation became impossible.

A critical aspect of survival data is that the true time to event remains unknown for some individuals who may not have had the event of interest at the end of the follow-up. This phenomenon is called censoring and may arise because a patient (a) has not (yet) experienced the outcome event by the study close date; (b) is lost to follow-up during the study period (e.g., due to transfer to another centre or for consent withdrawal); or (c) experiences another (competing) event that makes further follow-up impossible (e.g., heart transplantation, a new health problem or even a car accident). Censored observations are those who survived at least as long as they remained in the study but for whom their actual event-free survival times are not known exactly. Such right-censored survival times underestimate the true (but unknown) time to event. If the event occurred

in all individuals, other methods of analysis would be applicable. However the presence of censoring and the distribution of the failure times make survival analysis necessary for time-to-event data.

The analytical tool used to study survival data assumes that if censoring occurs it occurs randomly and is unrelated to the reason for failure (independent censoring principle). In practical terms, this means that censoring must carry no prognostic information about the subsequent survival experience. This must be guaranteed by the study design and implementation as it cannot be controlled for during the analysis. This "uninformative" assumption would be violated if subjects were highly likely to leave the study just prior to failure or dropout rates between groups were differential. In addition to the independent censoring principle, the other key requirements of any risk study also need to be satisfied for the validity of survival analysis: follow-up duration based on disease severity (sufficient to capture enough events), homogeneous cohort effect on survival (similar survival probabilities for subjects recruited early and late in the study), and independence of the failure times for standard approaches (absence of correlation in the data).

ii. Functions used to study survival data

Survival data are generally described and modelled in terms of three related functions, namely the survivor, the hazard and the cumulative hazard functions. They are different

functions of the linear predictor (the regressors and their coefficients) meant to summarize the information on the outcome components described above (time zero, end date and censor status) in one response variable (11, 12). The *survival probability* (cumulative survival probability or survivor function) is the probability (from 1 at t = 0 to 0 as time goes to infinity) that an individual survives from time zero up to a specified future time t (observation end). Survival probabilities at different times provide essential summary information from time to event data. Figure 3 shows the estimated primary survival probabilities of all 700 AV accesses in the fistula data (left panel). For example, a survivor function of 0.48 at 3 years informs that 48% of the AV accesses (observed from t = 0) are event free at 3 years (risk of 0.52 at 3 years). The *hazard* is the instantaneous probability that an individual who is under observation at time t has an event at that time. So the hazard is a rate, i.e. a probability over a time interval, though very small. Put another way, it gives the instantaneous potential for the event to occur, given that the subject has survived up to that instant (conditional rate). In contrast to the survivor function, which can only decrease over time, the hazard function can remain constant or vary with different shapes over time (Figure 3, right panel). The hazard is like a speed, with the risk of failure over time instead of distance covered over time, and may assume different values over time (from 0 to $+\infty$) independent of the average value calculated in an interval. In the access study the hazard of both primary and (to a lesser extent) secondary failure was greater initially, and tended to decrease over time. There is a defined relationship between survival and hazard and they are both related to a third

quantity called *cumulative hazard* (Figure 4). The cumulative hazard at t is the integral of the hazard (area under the hazard function between times 0 and t). To understand the concept it is useful to go back to the speed example. If an AV access faced a hazard rate of failure of 0.1 thromboses per hour (a speed of 0.1 mph), then the cumulative hazard is such that were that rate to continue for two days (the speed constantly at 0.1 mph) 4.8 failures were expected to occur (4.8 miles travelled) in 2 days. Since an integral is indeed just a sum, a cumulative hazard is not unlike the total number of times the AV access "would fail" over the interval period (cumulative force of mortality).

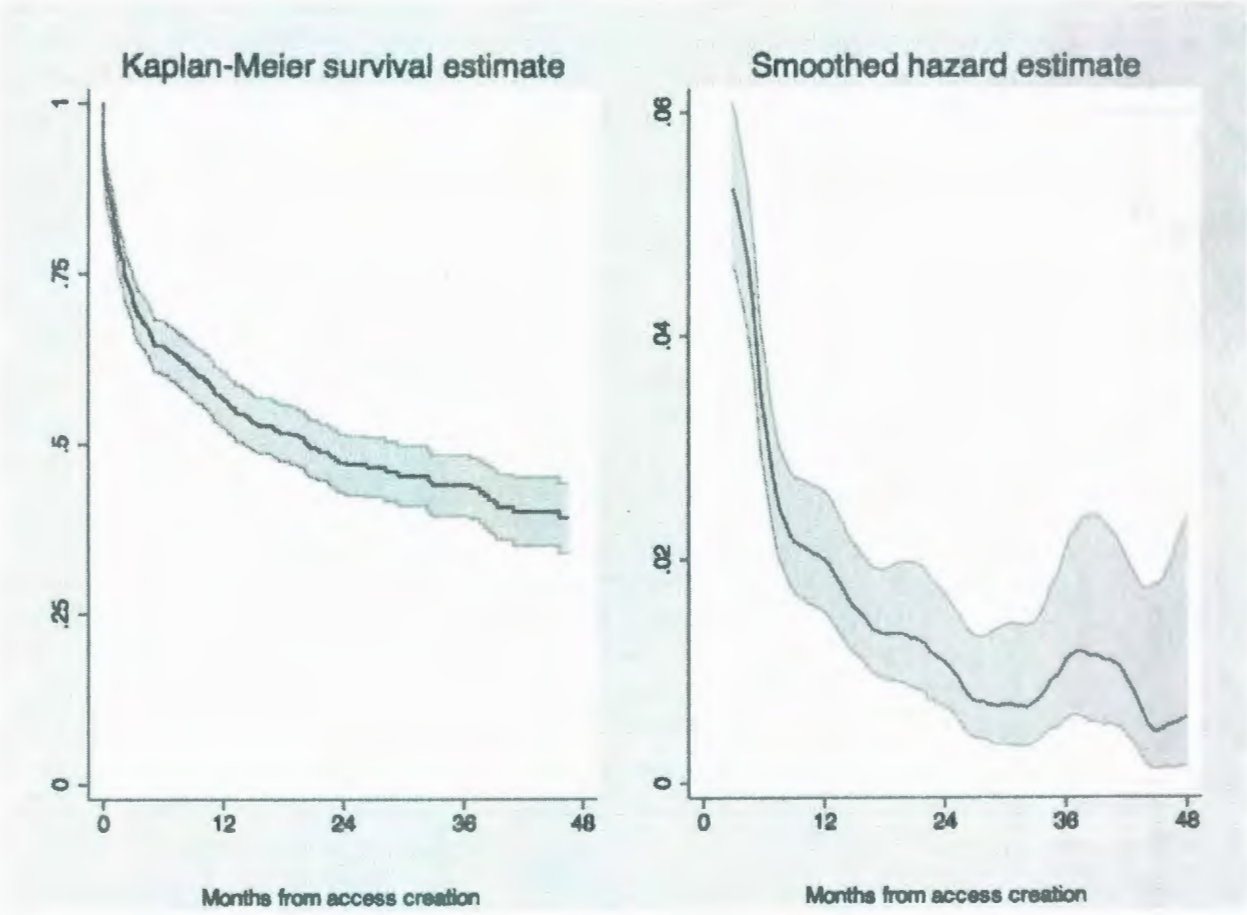


Figure 3: Primary survival data from the access study (3). All 700 AV accesses attempted in 535 incident hemodialysis patients are included. The left panel shows the overall survival probabilities over time since creation (Kaplan-Meier estimator and 95% confidence intervals). The right panel shows the hazard (or instantaneous risk) of failure over time (with 95% CI). It can be seen that the hazard declines sharply during the first months, and remains lower after one year (fluctuations may be due to sample shrinkage). Estimates are less precise (wider confidence bands) as fewer AV accesses remain at risk (right portion of each panel).

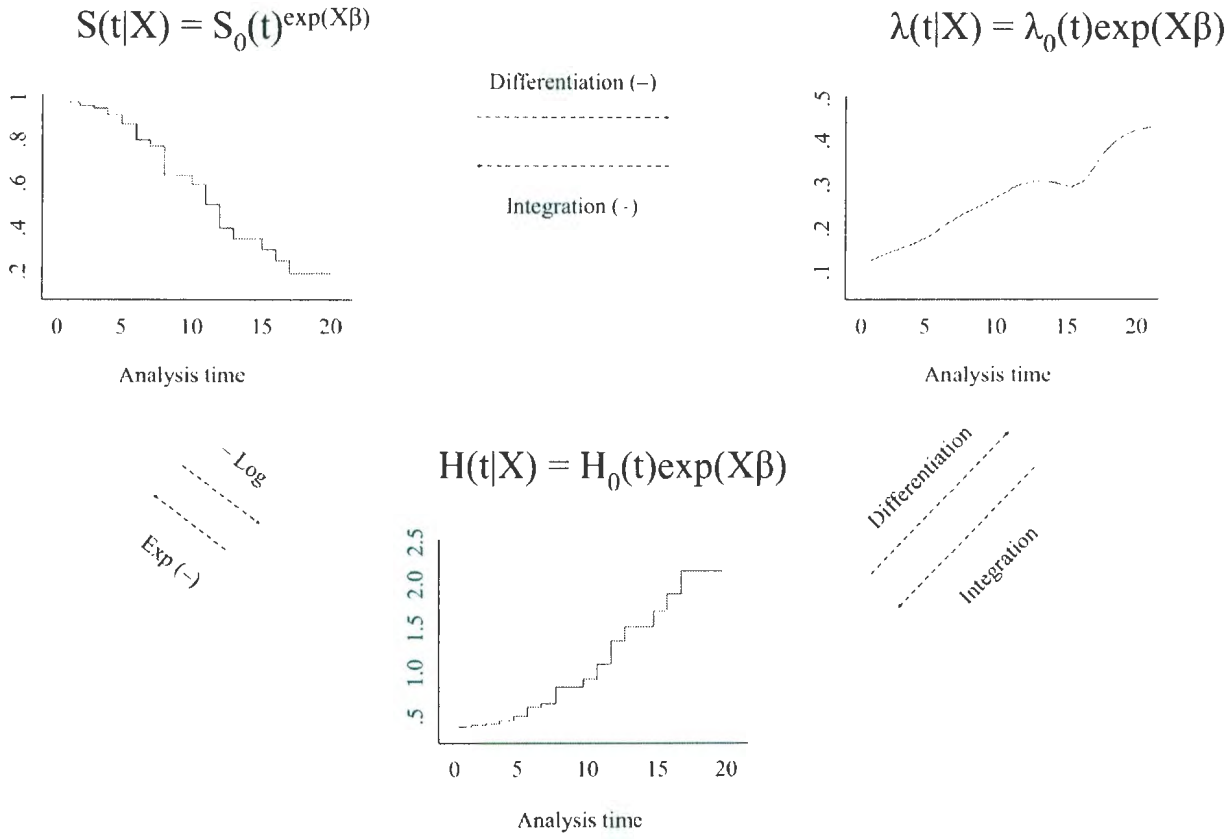


Figure 4: Characteristics and relationship among the hazard function λ(t), the survivor function S(t) and the cumulative hazard function H(t).

5. Making comparisons: Cox's and Poisson regressions

To compare hazards, survival functions, or times across groups, there are different approaches more or less free from specific distributional assumptions about the hazard function (Table 8). Furthermore, some parametric models have an accelerated failure time metric, i.e. the estimated coefficients (the co-variate effects) are interpretable as log-time ratios and some have both the proportional hazard and the log-time interpretation. The two interpretations are different. The proportional hazard metric focuses on the actual risk process (the hazard function) that causes failure and how the risk changes with the value of the covariates in the model. The accelerated failure time metric gives a more prominent role to time in the analysis (how the survival time changes with the value of the covariates in the model). The measures of effect estimated in the models reported in Tables 4 – 7 are hazard ratios (Cox's models) and incidence rate ratios (Poisson regression). This means that the estimated coefficients are interpretable as log-risk ratios (proportional risk metric). In other words, a hazard ratio of 1.316 (coefficient or log-risk ratio = 0.274) means that those referred to the nephrologist within 3 months of dialysis

Table 8: Forms of survival analyses

Form	**Parameter**	**Example**	**Metrics**	**Meaning**	**Exp(β)**
Non-P	None	KM	NA	Survival Prob.	NA
Semi-P	Effects	Cox's M	PH	Risk change	HR
Parametric	Effects / λ	Gamma	AFT	Time change	TR
		Log-N	AFT	Time change	TR
		Exponential	PH / AFT	T or R change	HR / TR
		Weibull	PH / AFT	T or R change	HR / TR

Legend: Exp(β), is the number e to the power of β, the estimated value of the coefficient; NA: not applicable; PH, proportional hazards; HR, hazard ratio; gamma, log-normal, exponential and Weibull are the names of some parametric regression models; AFT, accelerated failure time; TR, time ratio.

start experienced a risk 31% greater than those followed for longer (Table 4, Model 1). The model coefficients represent differences in logs by unit change of each predictor (for late referral indicator from 0 to 1). The meaning of the coefficients is interpreted taking the exponential. Each exponentiated coefficient represents the ratio of the risk between two levels or unit of exposure. If the model is formulated in the so called proportional risk metric, differences on the log scale are assumed to be constant and ratios on the exponential scale proportional (proportional hazard requirement). These effect measures can be estimated with semi- and fully parametric models.

The Cox's model is by far the most commonly used survival procedure. It is a *semi-parametric* model since it formulates the analysis of survival data where no parametric form of the hazard function (output) is specified and yet the effects of the covariates (inputs) are parameterized (i.e., modelled based on assumptions) to alter the baseline hazard function (the hazard for which all covariates are equal to zero). The Cox's model makes estimation possible assuming that the covariates multiplicatively shift the baseline hazard (Figure 5 and Figure 6).

Besides the ease of coefficient interpretation, freedom from distributional assumption is the greatest advantage of Cox's regression. The cost is a loss of efficiency (precision) since the parameters are estimated comparing subjects at the times when failures happen to occur whereas parametric models maximize the use of the information in the data.

Figure 5: Hazards proportionality (12): The individual hazard at time t (i.e., at any time during follow-up) given the exposure "X" is a function of the baseline hazard (λ_0) and the hazard ratio (HR) associated with each unit change of the input, i.e. $\lambda(t) = \lambda_0(t)$ times HR. This HR is estimated as "exp(β)". As the estimated coefficients are constant, and constant differences on the log scale correspond to constant ratio on the exponential scale, the model assumes proportional hazards (Hazards Proportionality – HP). If more inputs are in the model each HR is adjusted for the effect of all the other independent variables.

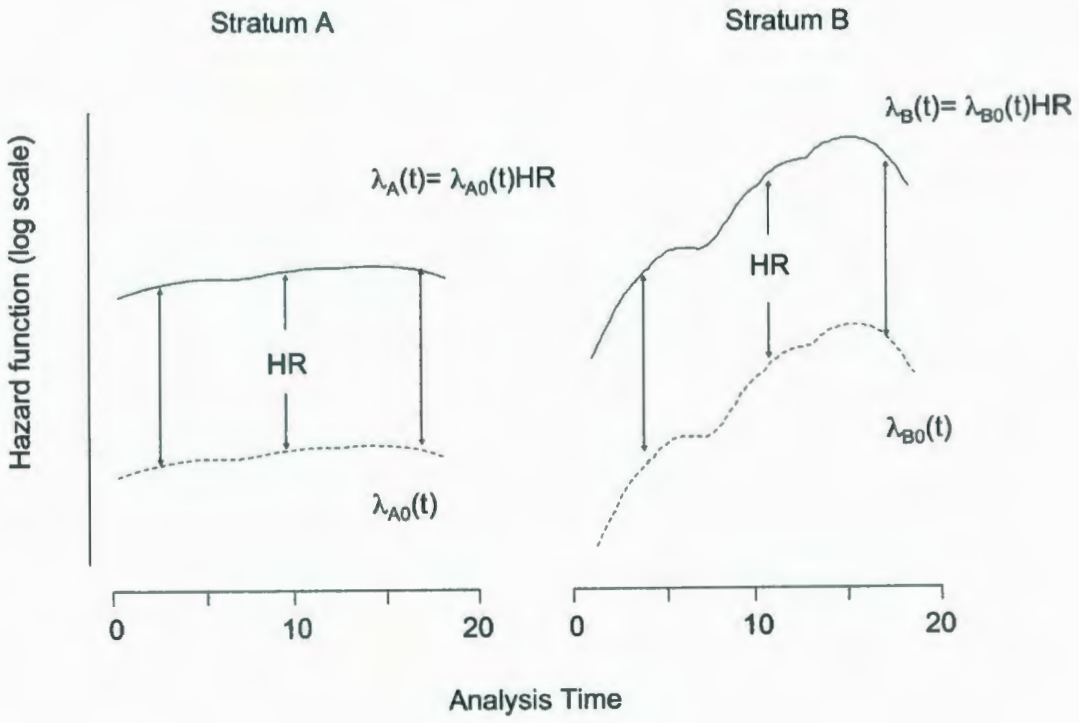


Figure 6: Stratified Cox' model (12): There can be groups with different baseline hazard $\lambda_0(t)$, such as race groups (e.g. race A and B). If the effect on survival of race A vs. B is not of interest (does not need to be estimated) but requires to be controlled for, a stratification variable can be used to specify the model (e.g., with possible values A and B). The model contains the same predictors (e.g., gender) but allows the basal risk to vary, i.e. $\lambda_{A0}(t) \neq \lambda_{B0}(t)$. The difference in the linear predictor (β) between two groups of subjects (e.g., men=1 and women=0) in terms of hazard (e.g., for cardiovascular event) is still constant (Hazards Proportionality – HP), and is the same in both strata. However, as $\lambda_{A0}(t) \neq \lambda_{B0}(t)$, also $\lambda_A(t) \neq \lambda_B(t)$.

Thus it is not surprising that the results of the Cox's models (Table 4 and 5) differed from those of Poisson models in terms of significance testing (Table 6 and 7). In fact, Poisson regression is a fully parametric model which can be applied also to survival data. What should cause alarm about the models in Table 6 and 7 are the coefficient estimates rather than the results of hypothesis testing. Indeed, the aspect to be verified is whether the observed hazard distribution (Figure 3, right panel) is consistent with the distributional assumptions implied by the Poisson model. In the access study the rate of access failure was not constant over the entire follow-up time. This violates the distributional assumption of the Poisson model and makes the models reported in Table 6 and 7 wrong. Thus, the inconsistency of the coefficients points out the inadequacy of a model that assumes constant baseline hazard, as the assumption of constant hazard is incorrect for these data. The main difference between the Cox's model and the other parametric models is that with Cox's regression how rates change with time can be ignored (no specific distribution of the baseline hazard is specified), while with all parametric models this must be taken into account. Some parametric models (such as the exponential and the Weibull models) are proportional hazards models. This implies that the log-relative hazard has the standard interpretation, i.e. the exponentiated coefficient is the hazard ratio for one unit change in the predictor. The direct comparability to Cox's regression is probably the most appealing feature of this class of models. In fact, when engaging a parametric estimation it is prudent to compare the estimated coefficients to those from a Cox's model fit, to verify that they are roughly similar. If they prove not to be similar, then there is evidence of mis-parameterization of the underlying baseline hazard. When the chosen parametric model is correct, or the effect of time is controlled for, coefficient estimates from parametric models are close to those of Cox's regression (Table 9). In the example, the follow-up time was divided into pieces within which the hazard appeared to be approximately constant (0 – 3, 3 – 6, 6 – 12, 12 – 20, 20 – 30 and > 30 months). However, sometimes rates vary so quickly with time that it would be necessary to split the follow-up into too many pieces, or it would be difficult to decide how to split the

Table 9: Models of the primary failure of the 1st AV access in the access study (N = 513)

Poisson model adjusted for the effect of time bands (0-3, 3-6, 6-12, 12-20, 20-30 & > 30 months)

Co-variate	IRR	P>|z|	95% Confidence Interval
LR	1.360	0.070	(0.975 to 1.896)
HF	1.146	0.525	(0.753 to 1.745)
HXL	2.007	0.056	(0.981 to 4.107)
VASC	1.377	0.025	(1.040 to 1.822)

Cox's model (no band adjustment)

Co-variate	HR	P>|z|	95% Confidence Interval
LR	1.328	0.095	(0.952 to 1.854)
HF	1.142	0.537	(0.749 to 1.739)
HXL	1.902	0.078	(0.929 to 3.890)
VASC	1.359	0.032	(1.027 to 1.798)

Legend: IRR = incidence rate ratio; P>|z| two sided P value of the Wald test on the coefficients; LR = late referral (referral to the nephrologist less than 3 months of the dialysis start); HF = history of heart failure; HXL interaction term between HF and LR; and VASC: presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events). Subjects receiving an AV graft were excluded from both models (N = 22).

data. With Cox's regression it is not necessary to split the data into time pieces because Cox's regression implicitly controls for time continuously. Both methods assume that the risk ratios are constant over the entire follow-up. In Poisson regression the effect is the ratio of two incidence rates, each constant within the time band whose effect is controlled for in the model. In the Cox's model the hazard ratios (HR) estimate the true risk ratios (RR) of instantaneous event rates. In fact, as with rates, HR is an instantaneous RR, the limiting value for the RR as time approaches zero. As time approaches 0, the risks also approach 0. However, the value of HR is different from zero and approaches that of the true RR. In survival analysis, the incidence rate ratio is the limiting value for the RR as time approaches 0. Poisson regression models rates, which are assumed to be low and constant with variance equal to the mean. Often complex survival data can be aggregated and Poisson regression and other generalized linear models can be used provided that the above assumptions are satisfied.

6. Parametric survival models

Linear regression, logistic regression and Poisson regression are examples of parametric models. With these models the outcome is assumed to follow a distribution from a certain family (normal, binomial, Poisson) with unknown parameters. For example, if one distribution has some parameters (e.g., mean systolic blood pressure 130 with SD 20 mmHg) and another distribution has different parameter values (e.g., 110 ± 10), the two distributions belong to the same family (normal) but they are different distributions. In parametric regression models data are typically used to estimate the values of the unknown parameters that fully specify the chosen distribution. The exponential and the Weibull models are two parametric proportional hazards models. As with the Cox's model they have the following structure $\lambda(t) = \lambda_0(t)e(X\beta)$, which means that the hazard at

any time t during the observation period depends on the baseline hazard $\lambda(t)$ times the hazard ratios $HR = e(X\beta)$. The term "parametric" means that in addition to the log-hazard ratios (betas) they produce direct estimates of the baseline hazard $\lambda_0(t)$. In the Cox's model this baseline hazard is left un-parameterized and through conditioning on failure times the hazard ratios are obtained anyway[1]. Parametric models estimate additional parameters defining the shape of the baseline hazard. These include an intercept and the so called "ancillary parameters". These additional parameters are the coefficients of the hazard function implied by the model. From these parameters it is possible to obtain the predicted baseline hazard function and the other related functions, i.e. the survivor and the cumulative hazard functions (Figures 4 and 7). Since the exponential and the Weibull models have both the proportional hazard and the accelerated failure time interpretation (Table 8), these two models will be used to fit the AV access study data and compare the two interpretations. Parametric models need not be PH models. Some have only the proportional hazard formulation (Gompertz) and others the accelerated failure time formulation (log-normal, log-logistic, and gamma). Some of these models are naturally non-proportional (log-normal) and some have several ancillary parameters possessing a highly flexible hazard function which allows for a large number of possible shapes (gamma). However, this is beyond the scope of the present work and will not be discussed further.

---

1 Although the baseline hazard and survival functions are not estimated with Cox's regression, these can be derived generalizing the Kaplan-Meier method and using the estimated effects. However, baseline hazard estimation is not necessary in Cox's regression because in the computations of the effects (HR) the baseline hazard cancels out.

Table 10 shows the primary survival of the first AV access (fistulae only) from the access study data fitting the Cox's, the exponential and the Weibull models in the (log) proportional hazard metric. Once the effect of time has been controlled for, the exponential model provides results similar to those of the Cox's models (Table 10) at the cost of estimating 6 additional parameters: the effects of each of the 5 time bands relative to the first band (month 0 to 3), and the intercept "a", which indicates that the baseline hazard is constant over time. Actually, splitting time in several pieces may even improve the results but would make the model more complex. The exponential model is useful to check "non-parametrically" the validity of any parametric form the researcher whishes to use. This piece-wise exponential model was used to identify the chosen time band categories along with the observed data from Figure 3. The exponential model can be used even if the overall hazard varies with time, provided that the process under study is understood and taken into account. For example, it is reasonable to believe that the hazard of fistula thrombosis is greater in the first days after surgery than later on. Based on the intercept and the effect of time (how the hazard changes with time) it is possible to predict the instantaneous baseline hazard within each time band, i.e. the instantaneous risk for failure in those not referred in a timely manner to the nephrologist and without cardiovascular disease. The effect of time is to alter the baseline hazard by an amount estimated by the coefficient associated with each time band, and the model is valid provided that the baseline hazard controlled for time remains constant within that time

Table 10: Models of the primary failure of the first AV fistula (N = 513)

Cox's model (see the HR in Table 4, model 2)

Co-variate	**log-HR**	**P>|z|**	**95% Confidence Interval**
LR	0.410	0.006	(0.116 to 0.705)
CVD	0.397	0.004	(0.127 to 0.667)

Exponential model adjusted for time bands (0-3, 3-6, 6-12, 12-20, 20-30 & > 30 months)

Co-variate	**log-HR**	**P>|z|**	**95% Confidence Interval**
LR	0.444	0.003	(0.150 to 0.738)
CVD	0.415	0.003	(0.145 to 0.685)
a	– 2.702	<.001	(– 2.954 to – 2.450)

Weibull model (no band adjustment)

Co-variate	**log-HR**	**P>|z|**	**95% Confidence Interval**
LR	0.429	0.004	(0.135 to 0.724)
CVD	0.406	0.003	(0.136 to 0.676)
a	– 2.276	<.001	(– 2.545 to – 2.007)
p	– 0.797	<.001	(– 0.916 to – 0.678)

Weibull model with band adjustment (effect of the time bands 3-30 & > 30 v 0-3 months)

Co-variate	**log-HR**	**P>|z|**	**95% Confidence Interval**
LR	0.414	0.006	(0.119 to 0.708)
CVD	0.400	0.004	(0.130 to 0.670)
a	– 2.235	<.001	(– 2.503 to – 1.968)
p	– 0.616	<.001	(– 0.783 to – 0.449)

Legend = log-HR: log-hazard ratio (beta coefficients); P>|z| two sided P value of the Wald test on the betas (log-HR); LR = nephrology referral < 3 months of dialysis; CVD = history of heart failure and / or presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events). The coefficients "a" and "p" are respectively the intercept and the ancillary (log) parameters.

band. From the baseline hazard (controlled for time) and the coefficients of the covariates in the model it is possible to estimate the hazard of those referred late (baseline hazard times the HR of late referral) or with cardiovascular disease (baseline hazard times the HR of cardiovascular disease) or both (baseline hazard times the HR of late referral times the HR of cardiovascular disease – multiplicative model).

Also the Weibull model provides results that seem close to those of the Cox's model, with the additional advantage of estimating fewer parameters than the exponential model to do the same job (Table 10). In fact the second new parameter "p" is the ancillary parameter of the Weibull model (Figure 7). This parameter determines the form of the baseline hazard, i.e. how the hazard changes over time. This is why the ancillary parameter is called the *shape parameter* as opposed to the other parameters (intercept and the co-variate coefficients) which are called *scale parameters* as they alter the level of the hazard proportionally, i.e. independent of time. In the Weibull model of the fistula data the hazard is monotonically decreasing as (log) $p < 0$ (it would be monotonically increasing if $p > 0$ and flat if $p = 0$, the model reducing to the exponential case). As with the Cox's model the baseline hazard is altered multiplicatively by the effect of the covariates in both the exponential and Weibull regression models (i.e. these baseline hazards increase by approximately $\exp(0.4) = 50\%$ in presence of any cardiovascular disease or if the patient is referred late). The ancillary parameter can also be modelled as a function of some (or several) co-variates, allowing each risk group to have its own

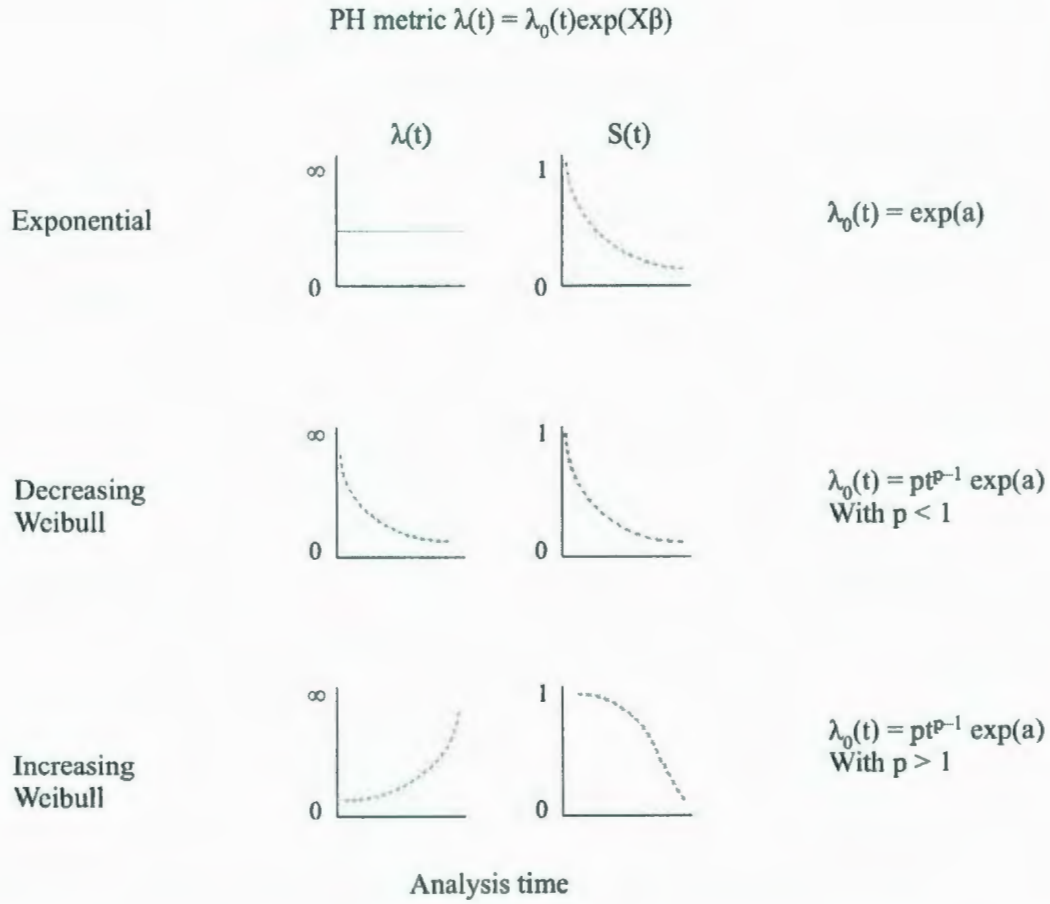


Figure 7: Relationship between three parametrically specified hazards and the corresponding survival probabilities. The first $\lambda(t)$ is a constant hazard rate over time (e.g., a constant speed); the following are the decreasing and increasing hazard rates based on a Weibull model (e.g., decreasing and increasing speed over time). These curves are illustrative examples and other shapes are possible. As opposed to the Cox's model where the baseline hazard $\lambda_0(t)$ was simply left un-parameterized and the coefficient estimates were obtained anyway through conditioning on failure times, in parametric models a functional form of $\lambda_0(t)$ is specified through 1 (e.g. the intercept "a" of the exponential and Weibull models) or additional parameters (e.g. the ancillary parameter "p" of the Weibull model).

hazard shape. For example, the hazard might decrease more slowly ($p < 1$ but closer to 1), or more quickly ($p < 1$ but closer to 0), in subjects with vascular disease than in those without vascular disease, or might even increase instead ($p > 1$ as opposed to < 1). Including presence of vascular disease in the co-variate list only assumes that this co-variate has an effect on the scale of the hazard (HR) but that the shape of the hazard (p) is the same for both its levels (as was the case in these data for all covariates). It is possible to assume that the shape of the hazard (p) changes with the level of a co-variate and not the scale (HR). It s also possible to assume that both the scale and the shape of the hazard are affected as the co-variate values change.

Table 11 shows the primary observed event rates of the first AV fistula and the hazards predicted by the exponential and Weibull models within each time band. It can be seen that despite its parsimony, the Weibull model does not seem to work well in predicting the risk at the beginning and at the very end of the observation times. The predictions of the Weibull model improve when the effect of two time pieces are controlled for (i.e. if the model allows additional changes of the hazard over time). However, it should be noted that the table reports averages which reflect the implications of the exponential model (as this implies a constant hazard within each time band) but not those of the Weibull model which has an ancillary parameter. In fact the hazard decreases (or increases) monotonically according to the Weibull model by an amount estimated by this shape parameter. In fact according to the second Weibull model in Table 10, the hazard

Table 11: Observed primary failure rates of the first AV fistula and hazards predicted by the exponential and the Weibull models (Table 10). The exponential model appears to work better than the Weibull models. The predictions of the Weibull improve when the effects of two time pieces are controlled for.

Time band	Observed	Exponential	Weibull (1)	Weibull (2)
0 to 3 months	0.0931	0.0957	0.0708	0.0807
3 to 6 months	0.0248	0.0250	0.0242	0.0209
6 to 12 months	0.0155	0.0156	0.0165	0.0152
12 to 20 months	0.0134	0.0134	0.0123	0.0119
20 to 30 months	0.0134	0.0104	0.0098	0.0099
> 30 months	0.0053	0.0053	0.0077	0.0053

Table 12: Parametric models of the primary failure of the first AV fistula in the access study (N = 513) formulated in the accelerated failure time metric.

Exponential model adjusted for the effect of the time bands (0-3, 3-6, 6-12, 12-20, 20-30 & > 30 months)

Co-variate	**log-TR**	**95% Conf. Int.**	**TR**	**95% Conf. Int.**
LR	– 0.444	– 0.738 to – 0.150	0.641	0.478 to 0.861
CVD	– 0.415	– 0.685 to – 0.145	0.660	0.504 to 0.864
a	2.702	2.954 to 2.450		

Weibull model with band adjustment (effect of the time bands 3-30 &> 30 v 0-3 months)

Co-variate	**log-TR**	**95% Conf. Int.**	**TR**	**95% Conf. Int.**
LR	– 0.766	– 1.321 to – 0.210	0.465	0.266 to 0.810
CVD	– 0.741	– 1.253 to – 0.228	0.476	0.285 to 0.795
a	4.138	3.435 to 4.841		
p	– 0.615	– 0.783 to – 0.448		

Legend: log-TR = log-time ratio (coefficients); TR = exponentiated coefficient (time ratio); LR = late referral (referral to the nephrologist < 3 months of dialysis); CVD = history of heart failure and / or presence of vascular disease (previous clinically documented coronary, cerebral or peripheral events). The coefficients "a" and "p" are respectively the intercept and the ancillary (log) parameters.

Table 13: Estimated median times to failure (in months) from the models in Table 12.

	Exponential	Weibull
LR– / CVD –	53.53 (10.33 to 164.66)	74.91 (31.80 to 172.82)
LR– / CVD +	33.58 (6.82 to 108.82)	34.90 (15.16 to 82.36)
LR+ / CVD –	30.46 (6.63 to 105.64)	32.66 (14.78 to 80.34)
LR+ / CVD +	17.89 (4.37 to 69.75)	14.33 (7.04 to 38.29)

Legend: LR = late referral absent (–) or present (+); CVD = cardiovascular disease absent (–) or present (+). Each cell reports the mean, the minimum and the maximum estimated category values.

decreases quickly during the first month, from 0.28 at day 1 to 0.11 after 1 week and to 0.06 at the end of the third week. This may explain why the coefficient estimates from the Weibull model are closer to those of the Cox's model than those of the exponential model (Table 10).

Table 12 shows the results of the same exponential and Weibull models controlled for time in the accelerated failure time formulation. When these models are expressed in this metric, the coefficients are log-time ratios. In other words, the exponentiated coefficients are interpretable as time ratios, i.e. the ratio of the predicted time to failure associated with a level of the regressor to the predicted time of the previous level. The underlying assumption for AFT models is that the effects of the covariates are multiplicative (proportional) with respect to survival times (time comparison) as opposed to PH models where the effects of the covariates is multiplicative with respect to hazards (hazard comparison). Mean and median survival times (and measures of dispersion) can be predicted using this formulation (Table 13).

The word "accelerated" is used to describe these models because they follow the parameterization[2] $\log(t) = -X\beta + \log(\tau)$. This last term $\log(\tau)$ has a distribution defined by the model. After taking the exponential $\tau = e(X\beta)*t$, where $e(X\beta)$ is the TR. Its reciprocal $e(-X\beta)$ is the acceleration parameter (AP) which estimates how the "speed of

[2] The AFT formulation is reported as $\log(t) = X\beta + \log(\tau)$ by some, where $X\beta$ is the $\log(AP) = -\log(TR)$. After taking the exponential $\tau = e\{\log(t) - \log(AP)\}$ and $\tau = t/AP = t*TR$ (Figure 8).

the process" changes. In fact, if AP = 1 (TR = 1) then $\tau = t$ and time passes at its "normal" rate (the covariates do not have any effect); if AP > 1 then time passes more quickly for the subject (time is accelerated) and the failure event is expected to occur sooner (TR < 1); and if AP < 1 then time passes more slowly (time is decelerated) and failure is expected to occur later (TR > 1). These models are called "accelerated" because the effect of a covariate measured in time units increases with time. For example, if the β coefficient of a covariate x is – 0.75 (TR = e(– 0.75) = 0.472 and AP = 1/TR = 2.117), one unit increase in x would speed up the process by a factor 2.117 resulting in a decrease in the expected value of log(t) by 0.75. In fact, the effect of the covariate would anticipate the occurrence of failure to $\tau = e(\log(1) - 0.75) = 0.472$ at t = 1; to $\tau = e(\log(5) - 0.75) = 2.361$ at t = 5; and to $\tau = e(\log(10) - 0.75) = 4.723$ at t = 10. Of note the predicted times to failure per unit change in the covariate x are "proportional" as AP is constant (1/0.472 = 5/2.361 = 10/4.723 = 2.117 = e[– Xβ]). However, at larger times greater absolute anticipations are expected when the coefficient $\beta < 0$ (TR < 1 and AP > 1) and longer absolute delays are expected when $\beta > 0$ (TR > 1 and AP < 1). This is why the marginal effect of x "accelerates" (or "decelerates") as time goes by. Figure 8 shows predicted survival times over t for different values of AP (page 57).

The coefficient β can also be viewed as an accelerating factor stretching or contracting time given a certain value of the survival probability. For example, dogs are said to grow older seven times faster than humans. The coefficient of the covariate "being dog as

opposed to human" in an AFT model would be $\beta_D = \log(1/7) = -1.94591$. The TR of dogs versus humans would be $e(\beta_D) = 0.1428571$ (1/7). In terms of probability of surviving "past some age", this means that a 10 year old dog would have the same survival probability as a 70 year old human. In fact, according to the AFT assumption at any value of the survival probability S(t) the ratio of times (TR) is constant, i.e. $S_D(10) = S_H(TR*10) = S_H(70)$ and $S_H(70) = S_D(AP*70) = S_H(10)$. In terms of time $t_D = e(\log(70) - 1.94591) = 10$. In other words, the life-span of dogs is contracted (that of humans is stretched out) by a quantity estimated by the model coefficient β formulated in the AFT metric. Figure 9 shows predicted survival probabilities over t for different values of TR (page 58).

An obvious question at the end of this brief review of standard analytical methods for risk estimation in longitudinal studies is: "Why should we bother with parametric models if the Cox's model remains the "safest" estimation method or "gold standard" if you prefer? Actually Cox's regression was used for the analysis of the Italian access data (1 – 4). However, often researchers are unaware of methods for the analysis of repeated events and fully parametric estimation procedures. In some circumstances these methods may be more appropriate. There are at least 4 aspects that should be considered when one is engaged in the analysis of risk data from longitudinal studies: 1) The parametric estimation scheme is more efficient (more precise) because it makes better use of the information in the data using probabilities that depict what occurs over the whole

observation time, and not only at the times when events happen to occur; 2) Parametric estimation methods are appropriate when one has an idea of what the baseline hazard looks like (such as in the case of AV access failure), and wants to impose that idea in order to obtain the most efficient estimate of the parameters and predict the baseline hazard subject to that constraint; 3) Parametric models are useful if one is interested not only in the hazard ratios but also in predicting the time to failure which requires some sort of parametric assumption; finally 4) Parametric models offer some advantages to study correlated data in terms of frailty. In fact the parametric form of the baseline hazard that parametric models posit allows easier description of the shared frailty (within individual or group dependencies). In addition, in the context of parametric models, this shared frailty can be contrasted with the latent individual un-shared frailty (over-dispersion or heterogeneity) which cannot be modelled with Cox's regression. These issues will be discussed in the next sections.

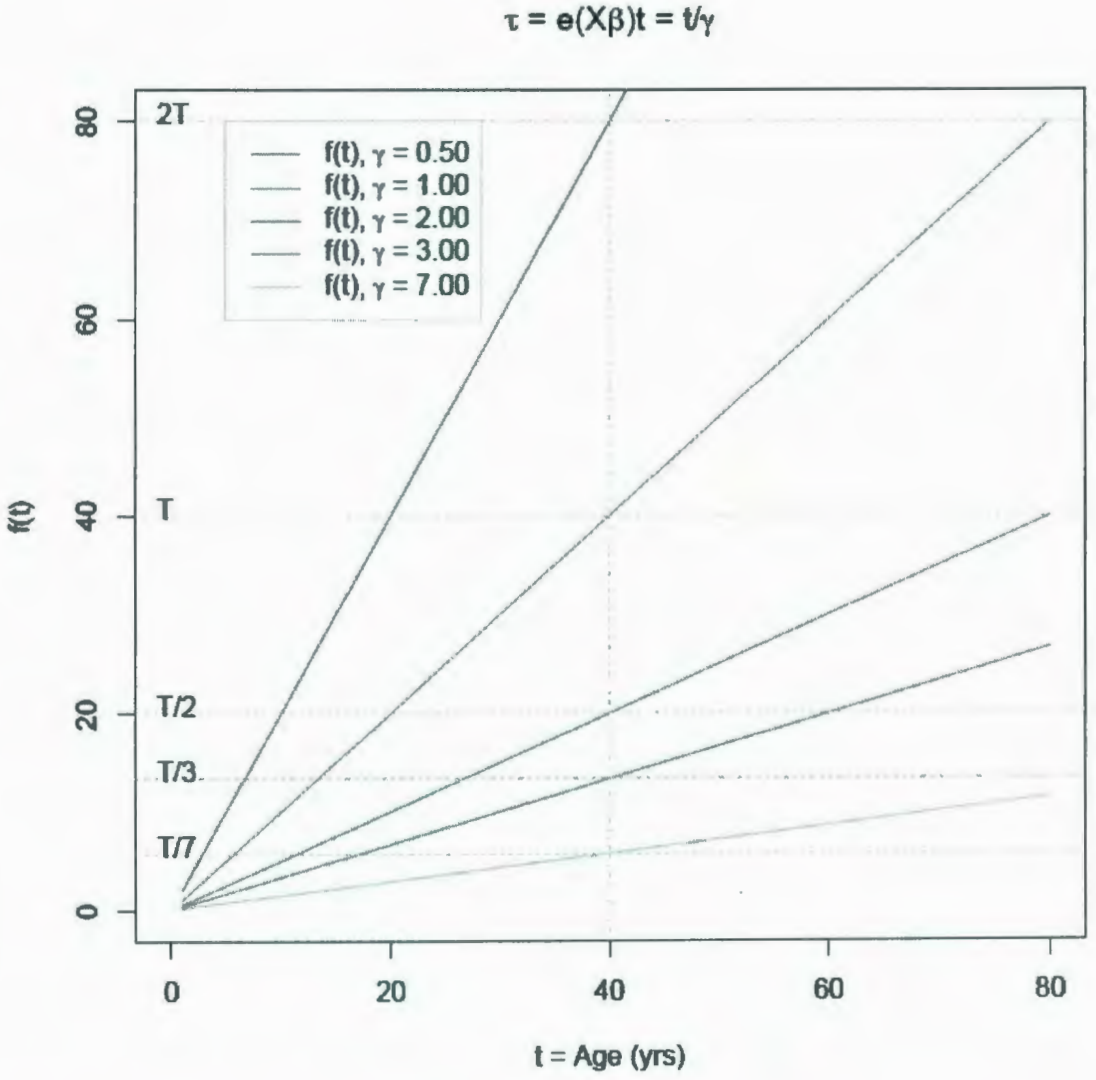


Figure 8: Predicted times to event (τ) as a function of the acceleration parameter AP (γ) and the follow-up time t. AP is the reciprocal of the time ratio (TR = e[Xβ]), i.e. γ = e(– Xβ). Time is stretched out to greater values when γ < 1 (e.g., if non smokers have a survival time twice as long as smokers or a TR = 2, then γ = e(– log[TR]) = 0.5). Time is contracted when γ > 1 (e.g., sicker patients die sooner than healthy individuals, and pets grow older faster than humans).

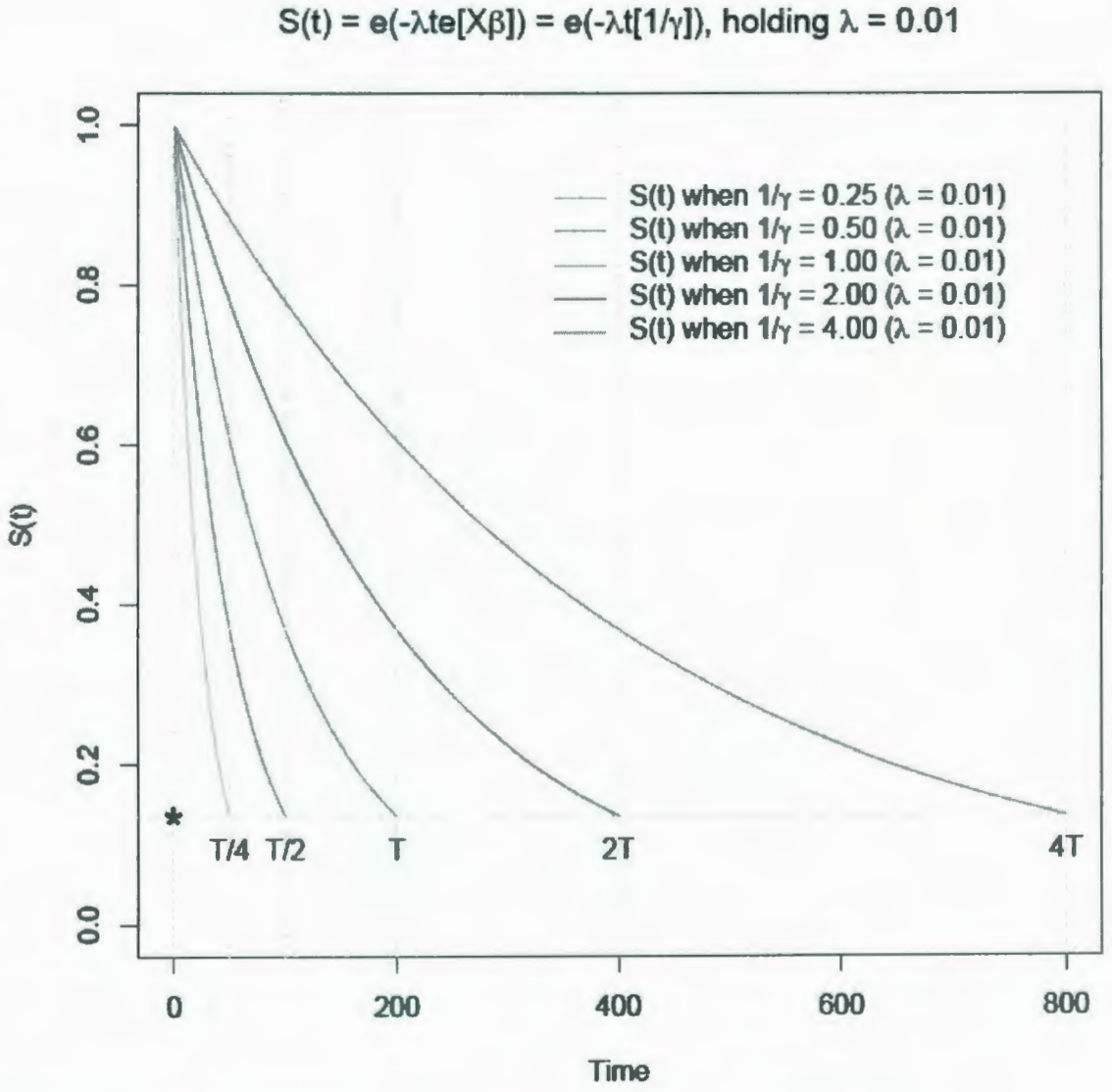


*Figure 9: Survival probability S(t) as a function of the reciprocal of the acceleration parameter AP (γ) or time ratio TR (e[Xβ] = 1/γ), the hazard λ (held at 0.01), and the follow-up time t. The effect of the covariate x is to extend (1/γ > 1) or contract (1/γ < 1) the expected time to failure. At the same expected survival probability (in the example * = 0.135) the expected time to failure is multiplicatively affected by 1/γ.*

3) Chapter 3: Methods for the analysis of correlated events

Chapter overview

Chapter 3 introduces definitions of and analytical approaches to correlated data. Statistical methods for correlated data can be distinguished into two major families: random effects models and pragmatic methods. Random effects modelling relies upon assumptions about the distribution of the extra-variability in the data once the fixed fixed effects have been estimated. Pragmatic methods correct the variance of the fixed effects without requiring extra-assumption to be satisfied. Pros and cons of each method is discussed at the end of the chapter.

1. Analysis choice and implications

Analysis of repeated AV access failure is possible considering more than one event per subject. As opposed to the analysis of the first event that occurs which ignores additional information in the data, analyses of event rates including multiple events per person have the potential to provide measures of disease burden in a population often more relevant and clinically interpretable. For example, in the AV access study the question is "Does late referral to the nephrologist impact the risk of failure of the *first and further* AV accesses?" as opposed to "Does late referral to the nephrologist impact the risk of failure of the *first* AV fistula?". Separate standard models of the fate of the first, second and subsequent accesses per person represent another option. However, the analysis power of the second, third and subsequent AV accesses would become smaller and smaller as fewer and fewer people have further AV accesses after the first. Multiple failure times analysis still allows testing whether the risk varies by access number while using all the information in the data.

Considering multiple AV accesses per patient or different possible events in each individual such as dialysis and death requires the recognition that some subjects may be especially likely to experience recurrent or multiple events, or become more prone to further events during a recurrent process or to competing events of different type. This tendency is called *frailty* and generates correlation in the data threatening the validity of traditional analytical tools. Alternative methods make use of all information in the data

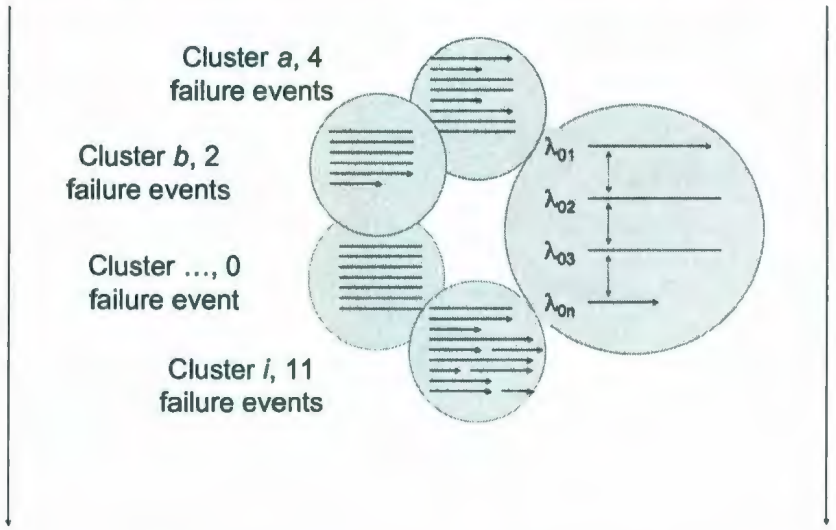
while accounting for the lack of independence of the event counts or failure times.

2. Sources of correlation

Correlation in the occurrence and timing of repeated or multiple events may ensue when individuals experiencing a single event belong to group or clusters, when a subject experiences some event more than once due to a recurrent event process, or when a subject experiences different events. This correlation may result from differences in the general tendency to fail across individuals and varying tendency to fail further once the recurrence process has started (Figure 10). Heterogeneity across subjects (also called *un-shared frailty*) may be due to unknown, unmeasured, or un-measurable effects (different lifestyles, genetic traits, environmental factors and experiences) which influence the likelihood to succumb to disease. As a result, some individuals are more (and others less) prone to disease, experiencing their first, second, third, etc., recurrence more (less) quickly than others. Event dependence within a subject emerges when the threshold for further events changes once previous events have occurred (e.g., the baseline risk of failure of the second and third access is progressively higher or lower than that of the first access). Further events become more or less likely according to whether the process induces a biological weakening or strengthening of the organism and whether the subject is more or less frail (*shared frailty*). In either case the risk for an event is a function of previous occurrences. Medical research and clinical experience suggest that both individual un-shared tendencies and varying shared susceptibility to fail during the

Heterogeneity: different general tendency to fail across individual (different hazards, $\lambda_{0a} \neq \lambda_{0b} \neq \ldots \neq \lambda_{0i}$)

Event dependence: within cluster *dependence* of the failure times (varying baseline hazard $\lambda_{01} \neq \lambda_{02} \neq \ldots \neq \lambda_{0n}$)



This heterogeneity can be incorporated into the model as **random effect** term
$\lambda(t|X) = \lambda_0(t) \exp(X\beta + \textbf{frailty})$

This shift in the baseline hazard can be controlled **stratifying** the model
$\lambda(t|X) = \lambda_{0n}(t) \exp(X\beta)$

Figure 10: Sources of correlation within risk data (survival approach). Unknown (or unmeasured) cluster / individual level factors can be responsible for heterogeneity across groups (with consequent different baseline group risk – λ_{0a}, λ_{0b}, λ_{0c}, etc.) and within group dependence of the failure events (varying baseline risk within cluster / subject during the recurrent or multiple failure process – λ_{01}, λ_{02}, λ_{03}, etc.). Heterogeneity across cluster / individuals can be modelled as a random effect term. Event dependencies can be controlled for by stratification.

recurrent process are likely to be the rule, rather than the exception, in the study of multiple events and that each may enhance the effect of the other (22 – 25).

It has been previously mentioned that independence of the failure times is a key requirement for survival analysis to be valid. Also Poisson regression assumes that the events are independent. Any correlation among events violates the assumption that the timing of events, or the event counts, is independent. This has two important consequences: the regression model is both biased and inefficient in a typical repeated events context. Variations of survival and count models, namely variance-corrected and random effects models have been proposed to account for, and possibly correct, the biasing effect due to the correlation among event times or counts.

3. Random effects modelling

To understand the philosophy of random effects models it is useful to use the analysis of variance approach to the general linear model. Figure 11 shows the change in left ventricular mass index as a function of study time and two blood pressure targets (standard or rigorous regimen). The bell curves represent the distribution of the response values measured at yearly intervals. Repeated measurements within the same subject make it possible to identify random individual deviations from the treatment group average. Put simply, around the line of the average response to treatment there are other lines each with its intercept and slope. Random effects are sources of variability in the

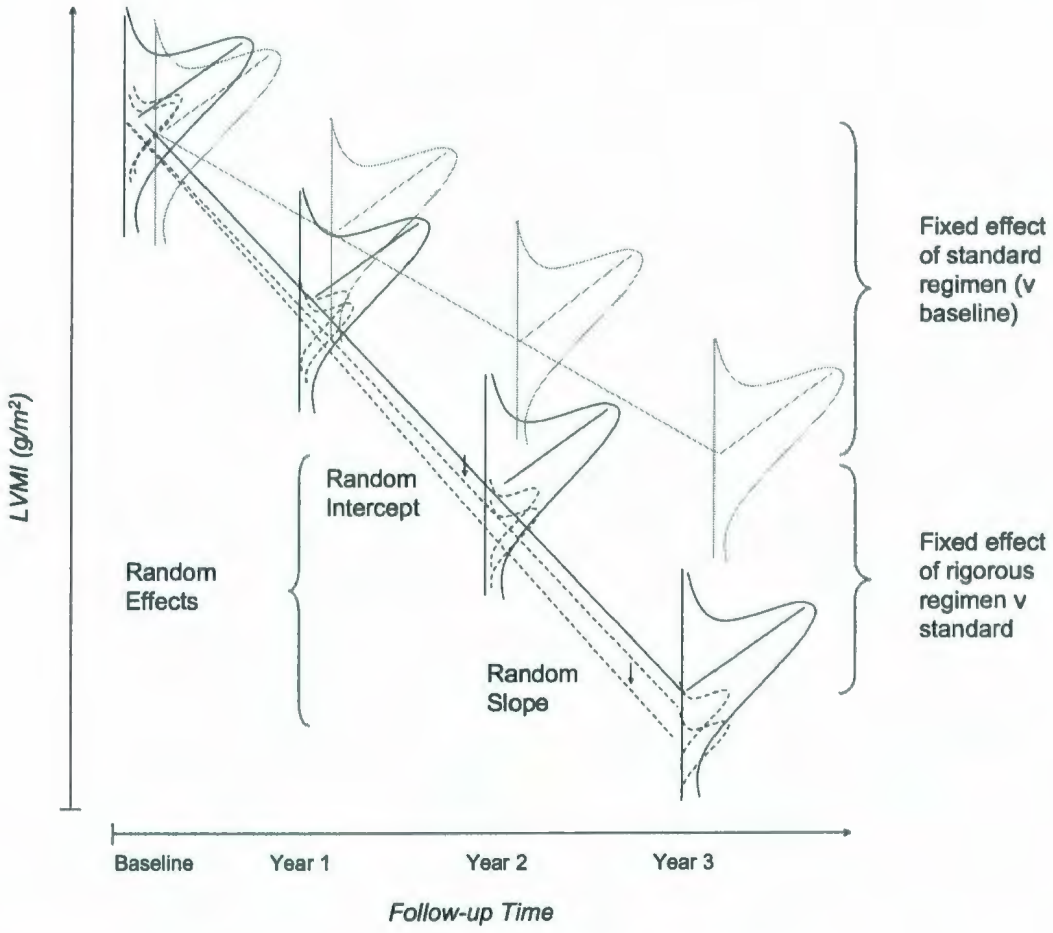


Figure 11: Graphical representation of fixed and random effects. Fixed effects are differences in the response – left ventricular mass index (LVMI) by exposure levels – blood pressure targets (continuous curves and lines). Random factors are responsible for the deviations from the average fixed effects (2 subjects are indicated for simplicity with dashed curves in only one treatment group). Random effects (dashed lines) can affect either or both the intercept and the slope of the line defining the input-output relationship. In other words, independent of the fixed effects, different subjects (in both treatment groups) may have different average values of the response either or both at any given value of the exposure (random intercept – vertical distance between lines) and by level of exposure (random slope – rate of average change).

intercept and slope of the group lines (fixed effects) that can be taken into account because they are due to subjects. In other words, these random or individual effects account for the variation in the response that the predictors of interest fail to explain but can be controlled for. Although not as easy to represent graphically, these random effects exist also in other generalized linear models (including those for counts) and survival models (13).

More generally, it is always useful to look at the study outcome variability as a mixture of different components. The regression coefficients of the model covariates estimate the "explained variability" of the response (systematic component). These are called fixed effects, because they are associated with fixed factors (or continuous inputs) whose levels of interest are actually measured or measurable. Fixed effects are unknown constant population parameters (e.g., the "true effect" of blood pressure target on left ventricular mass). The levels of interest of the fixed covariates are known or chosen by design (e.g., gender or exposure levels, or categories of a continuous covariate such as blood pressure targets). However, other covariates are often measured in some studies. They are called random classification variables because their levels can be thought of as being 'randomly sampled from a population of levels', such as individuals A, B, C in repeated / longitudinal designs, Drs A, B, C, or hospitals A, B, C and so on in clustered studies. All possible levels of these random factors are not present in a single study, but researchers still intend to make inferences about the entire "population of levels". In the above

example, study participants are random factors. To distinguish between random and fixed factors, it is useful to answer the following question: "Were the study to be repeated would the same groups / levels be used again?" If yes (e.g., gender, smoking, quality of pre-dialysis care, treatment A v B, age groups): this implies fixed effects. If not (e.g., centres, regions, subjects) it implies random effects. However, the same variable may be treated as fixed factor in some studies, and as a random variable in others, depending on the study question (e.g., health policy effects).

Of course all models have a random component that represents what remains to be explained once the model has been fitted (11, 12). However, a model containing a random effect splits its random part into two layers, the variation explained by the random factors and what remains unexplained by the combination of fixed and random factors (13). Random effects are unobserved random changes of the response by levels of the random factors or deviations from the relationship described by the fixed factors. For example, suppose that an outcome such as peripheral blood flow is measured twice in the same subject before an experiment is undertaken (Figure 12). Response values in the same subject tend to be closer to each other than values obtained from different individuals. Consequently, two error components exist rather than one. One is due to subject (between subject variation) and it is the random effect shared within individual but varying across them (heterogeneity). The other is due to the measurement occasion nested within subject (within subject variation). Within subject residuals are closer

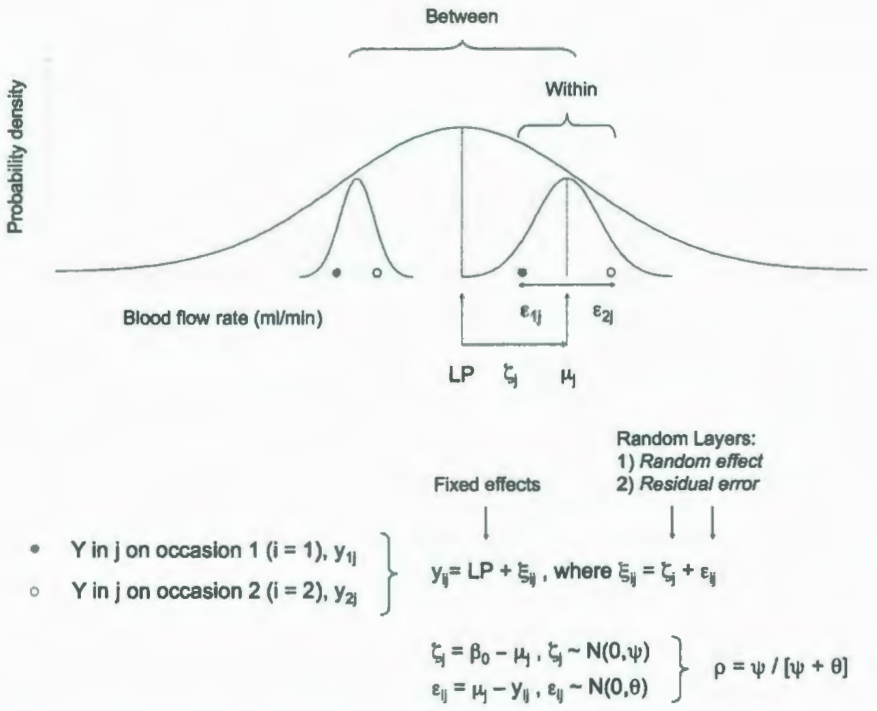


Figure 12: Between (B) vs. Within (W) Subject (j) Correlation and Variance Components: Blood flow rate values recorded on two occasions in the same subject are correlated as they (1) tend to lie on the same side of the overall mean (linear predictor, LP); and (2) be closer to each other than those taken on different individuals (within subject variability < between subject variability). The response value (y_{ij}) of subject (j) in occasion (i) equals LP plus an error term (ξ_{ij}). This includes two components: the variability due to subject "j" (random effect ζ_j equal to the difference between the subject mean (μ_j) and LP); and the variability due to measurement on occasion "i" (effect of occasion ε_{ij} nested in subject equal to the difference between μ_j and each response measured on j, y_{ij}). Usually it is assumed that both these components are normally distributed (~ N) with mean zero and some non-zero variance (ψ and θ). In linear random intercept models the intra-class correlation coefficient ρ estimates the outcome correlation as the proportion of the total variance explained by the variance (θ) of the random effect (ζ_j).

together than between subject residuals due to "shared characteristics". The variability due to measurement can be estimated when more than one measurement is performed in the same subject, although it exists independent of the number of measurements performed. A random effect model estimates both these variance components. When the variance of the random effect is significantly different from zero, the null hypothesis of absence of correlation in the data is rejected. The proportion of the total variance due to subject estimates the correlation in the data, and the accuracy of the measurement tool[3].

The heterogeneity among clusters / subjects and the dependencies within cluster / subject are incorporated into the estimated portion of the model by making assumptions about the distribution of the resulting random effect (13). As a result, under random effects models the outcome response (event times or counts) are assumed to be independent *conditional* on the patient's underlying frailty and inference can be made in the standard fashion. Random effects models estimate the variance of this latent effect. When this variance is significantly different from zero, the model supports the hypothesis of a significant heterogeneity / dependencies in the data.

4. Robust variance method

Robust variance or variance-corrected models represent another way to deal with the

3 When one observation per subject has been collected, un-shared random effects can be modelled using parametric models. In such models the frailty term is used to correct the over (under) dispersion implied by the chosen distribution. When more than one record per individual is available or subjects belong to clusters, it is also possible to predict the frailty value for each individual or cluster using shared frailty models.

problems produced by heterogeneity across individuals and failure-time or count dependencies. These methods do not incorporate any random effect into the estimates themselves but were developed to account for the lack of independence by using "robust" standard errors (also known as "sandwich estimator" from the matrix algebra). This method corrects the variance of the coefficient estimates (the random part of the model) for the correlation in the data by incorporating the dependencies in the process of computations. This is done by removing one cluster at a time, and providing an honest estimate for correlated data whenever the observations left out at any step are independent of the observations left in (22, 23). The standard errors of the coefficients are usually (but not always) larger, depending on the sign of the correlation in the data. Put simply, the statistical testing is more conservative (the confidence intervals larger) as compared to the corresponding generalized linear model applied to the same data as though each observation was independent (independent correlation structure). This empirical method is called robust because the variance estimation is consistent, even if the chosen correlation structure is incorrect (robustness to mis-specifications). This robust or corrected variance method can be used both for count and survival models, and both for semi-parametric and parametric models.

Although it is possible to run a standard Poisson model using robust standard errors, a popular approach to the analysis of correlated count data is represented by Generalized Estimating Equations (GEE). GEE are a class of analytical methods which represents the

extensions of standard generalized linear models for repeated / multiple measures (13). GEE have the same structure as standard regression models, i.e. a systematic component and a random component without specification of any additional random layer (random effect). The estimation method of GEE requires the specification of a working correlation for the observed responses to obtain unbiased estimates of the coefficients and their variances (Table 14). As a result, the standard errors of the GEE coefficients are corrected assuming that one or more correlation coefficients (parameters) describe the association of pairs of different responses from the same subject or cluster. Which structure best describes the real data correlation is not always obvious, although the research design may help decide. However, GEE analysis requires only a rough estimate of this structure to get started. The final parameter estimates (fixed effects, their standard errors, and the ρ correlation coefficients) are not usually dependent on the accuracy of the initial assumptions about the correlation matrix. In fact they are consistent (i.e., converge to the true value) as the number of clusters / subjects increases even if the initial structure is incorrectly specified, unless the fraction of missing data is large or they are not missing at random[4]. Although the correlation structure is not necessarily the same for all clusters / subjects, GEE assume one set of ρ parameters common to all clusters / subjects to avoid estimating too many parameters. Given the importance of the chosen correlation

4 Data are said to be missing at random (MAR) if, conditional on the observed data, the missingness is independent of the unobserved measurements (as opposed to data not missing at random, MNAR). Maximum likelihood estimation still retains its desirable properties (validity, consistency, efficiency) provided that data are MAR. For GEE to be valid this may not be enough and data need to be missing completely at random (MCAR), i.e. the missingness must be independent of both unobserved (missing value of the response) and observed data (previous response data and covariates values).

Table 14: Examples of correlation structures used in GEE: Each panel represents a correlation matrix between any two of 4 possible observations in the same subject taken at time 1, 2, 3, 4 (or in subject 1, 2, 3, 4 of the same cluster). Each matrix has a value of 1 along the main diagonal (as each measure correlates perfectly with itself) and some non-1 value off the diagonal. In the absence of correlation (independent errors) the correlation structure is independent (identity matrix); it is exchangeable if there is only one parameter ρ for any pair of measurements (e.g., in a cross-sectional study the order of measurement is arbitrary and it may be assumed that any two responses within a cluster have the same correlation); unstructured if there are n(n–1)/2 different parameters ρ (e.g., if it is assumed that there are as many ρ parameters as there are paired combinations of n measurements); autoregressive if there is only one ρ raised to the power of the absolute difference between the response times (e.g., in longitudinal designs it is reasonable to assumed that the correlation is greater for observations taken closer in time than farther apart as the order of measurement is not arbitrary).*

	T1	T2	T3	T4
T1	1	0	0	0
T2	0	1	0	0
T3	0	0	1	0
T4	0	0	0	1

Independent

	T1	T2	T3	T4
T1	1	ρ	ρ	ρ
T2	ρ	1	ρ	ρ
T3	ρ	ρ	1	ρ
T4	ρ	ρ	ρ	1

Exchangeable

	T1	T2	T3	T4
T1	1	ρ_{12}	ρ_{13}	ρ_{14}
T2	ρ_{21}	1	ρ_{23}	ρ_{24}
T3	ρ_{31}	ρ_{32}	1	ρ_{34}
T4	ρ_{41}	ρ_{42}	ρ_{43}	1

Independent

	T1	T2	T3	T4
T1	1	ρ	ρ^2	ρ^3
T2	ρ	1	ρ	ρ^2
T3	ρ^2	ρ	1	ρ
T4	ρ^3	ρ^2	ρ	1

Exchangeable

structure, the possibility of mis-specification in real life situations and, most of all, the possibility that data may not be missing completely at random, robust standard errors are usually recommended with GEE. Examples of random effects and GEE Poisson modelling of the AV access data will be presented in chapter 6.

The survival counter part of such "robust" approaches is represented by the family of robust variance methods. These survival techniques have been developed as extensions of the standard Cox's model. Variations within the variance corrected survival models are based on the definition of the risk sets (when the risk starts and ends for each subject) and whether event specific baseline risk is allowed by stratification. These models will be presented separately in their semi-parametric (chapter 5) and parametric formulation (chapter 6) with examples.

5. Method choice

The choice of the analytical tool to correct for the correlation in the data can be guided by different considerations.

As opposed to random effects models for counts, GEE are based on only one level of clustering, are not designed for inferences about the covariance structure (the working correlation structure is formulated with no distributional assumptions) and do not give predicted response values for each cluster. Using random effects models involves making extra assumptions, but gives more efficient estimates, and allows estimating

contributions to variability from different sources, including multilevel correlations. Finally, GEE are marginal models as they assume a model holding over all clusters (population average). Therefore, the coefficients represent the average change in the response over the entire population for a unit change in the predictor. Random models are conditional models in that they assume a model specific to each cluster / subject. Therefore, the coefficients represent the average change in the response for each cluster / individual, given a unit change in the predictor. Although population effects can be derived averaging cluster effects, conditional models are most useful when the objective is to make inferences about clusters / individuals rather than the population.

Similar considerations apply to survival analysis. In survival analysis stratification represents an additional advantage of variance corrected (robust) survival methods versus GEE for counts. Parametric survival models can formally distinguish between un-shared and shared frailty, and estimate the so called "frailty effect" (chapter 6).

4) Chapter 4: Failure process and event types

Chapter overview

Chapter 4 shows how to define the risk set and how to set up the data layout for survival or event-history data analysis. Survival analysis requires that time to event is defined for each individual assigning values to the study start date, the observation end date and the censor status variable at the last observation date. Different risk sets definitions exist for the analysis of recurrent and multiple events.

1. Defining the risk sets for survival (event-history) data

The risk sets for survival analysis are defined through different possible organizations of event-history data in the data layout. This task is accomplished by specifying the 3 components of the response variable (time start, time stop and censor status), and possible different basal risk categories by event number / type (in addition to some covariate if necessary) using stratification. Valid risk estimation depends on the extent to which data layout organization reflects the nature of the underlying failure process. To achieve this goal the following aspects need to be considered: a) classification of the type and order of the failure events: whether the events are of different (e.g. dialysis start or cardiovascular event) or the same type (e.g., repeated catheter infections), and whether they occur with (e.g., repeated infection or rejection episodes) or without natural order (e.g., cardiovascular events and infections in the same individual); b) definition of the time at risk: when the risk starts and ends; c) consideration of the mechanisms through which the predictor is involved in the process: whether / how the same predictor affects different outcomes; and d) definition of what is being modelled: the time to each event, the total course of a recurrent process or the time segments to each recurring event. Different risk sets have been formulated in the past to address different questions while fitting recurrent or multiple failure time data (Table 15 and Table 16).

2. Recurrent failure events of the same type

Events of the same type may or may not follow a natural order. Catheter infections or

Table 15: Risk sets for survival analysis: structure and implications

Risk set	Order / type	Example	Strata	Time zero	Modelling	Assumptions
Counting process	Ordered / Same	Catheter infections or dysfunction episodes; fistula thrombosis; repeated peritonitis or transplant rejection recurrences: they are ordered events in that they may be seen in a study that records the time to first, second, third event, and so on, and the subject is not at risk for further events until a prior one has occurred. Four layout options are available for ordered recurrences	No	Each observation time is event defined	Total time course of the recurrent process	Order not important; no tied times; same baseline hazard
Stratified Marginal risk sets	Ordered / Same		Yes	Time measured from subject enrolment	Total time course of the recurrent process	Events as independent processes; order ignored
Conditional risk sets from entry (elapsed time)	Ordered / Same		Yes	Enrolment (time measured continuously from entry)	Total time course of the recurrent event process	No risk for further event until a prior has occurred
Conditional risk sets from event (gap time)	Ordered / Same		Yes	Clock set to zero after each event	Time segments between events	No risk for further event until a prior has occurred
Un-stratified Marginal risk sets	Unordered / Same	The same lesion in paired organs such as the eye	No	Time measured from subject enrolment	Total time to each event	The same process for all events
Competing risk sets[5]	Unordered / Different	Uraemia and mortality in Chronic Kidney Disease	Yes	Time measured from subject enrolment	Total time to each event	Predictor involved in competing processes

5 In the absence of correlation and dependent censoring each observation continues until the first event that occurs giving the same results as the time to combined event analysis (competing risk model of Lunn-McNeil). In the presence of correlation the observations continue beyond the first event that occurs (marginal model of Wei-Lin-Weissfeld), each event can occur only once per subject, ad all subjects are at risk for all events.

Table 16: Risk sets for survival analysis: example of repeated events in 2 subjects (3 observed in subject 1 and 1 in subject 2 – models 1 to 6 for event of the same type); and triple failure times in other 2 subjects (model 7 and 8 for event of different type).

	Id = 1			Id = 2			Notes
Model	**Time**	**Event**	**Stratum**	**Time**	**Event**	**Stratum**	
1) Counting processes (VC)	(0,10] (10,15] (15,20] (20,30]	1 1 1 0	1 1 1 1	(0,20] (20,25]	1 0	1 1	# id records id = # events + 1; order not important; no tied times; same baseline hazard; t experience broken in event defined segments
2) Marginal risk sets (VC)	(0,10] (0,15] (0,20] (0,30]	1 1 1 0	1 2 3 4	(0,20] (0,25] (0,25] (0,25]	1 0 0 0	1 2 3 4	Strata present but order not important; start and stop times are the margins of the follow-up
3) Conditional risk sets from entry – elapsed time (VC)	(0,10] (10,15] (15,20] (20,30]	1 1 1 0	1 2 3 4	(0,20] (20,25]	1 0	1 2	Total time measured continuously from entry; stratification keeps track of the event # (sequential assumption)
4) Conditional risk sets from event – gap time (VC)	(0,10] (0,5] (0,5] (0,10]	1 1 1 0	1 2 3 4	(0,20] (0,5]	1 0	1 2	Clock set to zero after each failure event; stratification keeps track of the event # (sequential assumption)
5) Unconditional Frailty – elapsed time (RE)	(0,10] (10,15] (15,20] (20,25]	1 1 1 0	1 1 1 1	(0,20] (20,25]	1 0	1 1	No stratification; same risk set as the counting process (elapsed time); heterogeneity controlled through modelling random effects
6) Conditional Frailty – gap time (RE)	(0,10] (0,5] (0,5] (0,10]	1 1 1 0	1 2 3 4	(0,20] (0, 5]	1 0	1 2	Dependencies controlled through stratification; same risk set as the conditional from event (gap time)
7) Independent Competing risk sets (standard)	(0,10] (0,10] (0,10]	1 0 0	1 2 3	(0,12] (0,12] (0,12]	0 1 0	1 2 3	Marginal risk set stratified by event type; censoring at the t of the first event (combined end-point)
8) Correlated Competing risk sets (VC / RE)	(0,10] (0,30] (0,20]	1 0 1	1 2 3	(0,25] (0,12] (0,20]	1 1 1	1 2 3	Marginal risk set stratified by event type; follow-up continues beyond the first event

Legend: VC = variance corrected method; RE = random effects; unordered events of the same type (such as the same eye lesions in either or both eyes where the order does not matter) are modelled using a marginal risk sets such as # 8 but without stratification.

catheter dysfunction episodes, acute transplant rejection or peritonitis recurrences, AV fistula and graft failures are ordered events. In these cases, the underlying failure process can be captured in a longitudinal study that records the time to the first, second, third event, and so on. A common assumption is that the subject is not at risk for further events until a prior one has occurred. Several analytical options are available for this type of repeated events within the family of the variance corrected methods and frailty models (Table 15 and Table 16): the counting process (26), marginal risk sets (27, 28), and the conditional risk sets (29) were originally formulated to address correlated events using robust variance methods, but similar risk sets have been proposed using random effects modelling (23 – 25).

As it can be seen from Table 15 and 16, in the *counting process model* each subject becomes a "multi-event counting process" since the total follow-up time of the subject is broken into event defined segments, starting from entry into the study with as many records per individual as there are events plus one (26). Consequently, the subject returns "at risk" once an event has occurred until the study ends or observation becomes impossible. The ordering is taken into account to some extent by the sequentiality of the time pieces. However, the counting process is not stratified thus reflecting the assumption of similar basal risk for all events. Averaging potentially different baseline risks as events recur may induce bias. This constant basal risk assumption and the requirement for lack of tied times are often untenable. Despite these limitations, the

theory underlying the counting process and the way the risk sets are organized, make this approach appealing for a general approach to recurrent processes.

In the *marginal risk sets model* the risk set up is stratified by event number and thus treats each failure occurrence as a separate process. This accounts for varying basal risks as events recur. However, the marginal risk set actually ignores the event ordering as the time at risk begins at the initial observation time for all records – e.g., the possibility that a person is at risk for the fourth access failure before the first access has even failed is not excluded (27). This model may be useful to model the total time to each of the possible recurrent events, allowing basal risks to differ but with no strict order assumption. Such a model may be useful to model repeated hospitalizations from different causes, where the orderable event of the same time is represented by the hospital admission and the possible basal risk change after each episode is accounted for by stratification.

The assumption of the *conditional risk sets model* is that each patient is not at risk for a further event until a prior has occurred (29). Two variations with different time scales and risk sets have been implemented and both stratify the data by event number so that the baseline hazard is allowed to vary with each event. In the conditional risk sets model from entry (elapsed time) the data is set up as for the counting process (t measured from entry). This latter variation is useful when modelling the full time course of the recurrent event process. In the conditional risk sets model from previous event (gap time) the

clock is reset at each event (t from previous event with zero time at the beginning of each follow-up segment). This variation is useful to model the gap time between events. Both models are stratified by failure order to track the event number. The structure of the data set up reflects this sequence or ordering assumption (conditional risk). However, elapsed time estimation produces the hazard of an event since the study began, while the gap time formulation gives the hazard since the previous event. The choice of gap versus elapsed time approach depends on the research question at hand. Using gap time presumes there are substantive reasons to believe that the 'clock should restart' after each event in order to determine the effect of the covariates on subsequent events (e.g., when a previous infection has been cleared prior to the next catheter placement). In this case estimated effects mirror how the covariates affect the risk of failure for each access. In contrast, elapsed time models assess the effect of the covariates on the risk of failure from the start of the study through to the end (e.g., when there is no reason to reset the risk clock because the risk accumulates since entry). In such a case the estimated effects reflect how the covariates affect the risk of failure over the entire course of the recurrent event process.

Data set up for frailty models are the same as those of the conditional risk sets from entry (traditional un-stratified frailty) and previous event (stratified or conditional frailty). The heterogeneity between individuals / clusters is taken into account through modelling random effects. Stratification is an additional tool to control for the dependencies among

the failure times.

Examples of modelling ordered events of the same type using the AV access data will be discussed in chapter 5 and 6.

3. Unordered failure events

For unordered events of the same type (such as the same retinal lesions of either eye) the *un-stratified marginal risk sets* has been used (23). However, this recurrent failure process is not so common. For unordered events of different type, such as infections, rejection episodes and death in a follow-up study of transplant patients, the suggested risk set is the *marginal risk sets*. This risk set is stratified by event type to allow basal risks to vary (Table 15 and 16). Instead, the coefficients are restricted to be the same across strata, although stratum specific effects can be incorporated into the model. Other nephrology examples of failures of different types include diverse adverse reactions to therapy in an intervention trial, or uraemia and mortality in a follow-up study of chronic kidney disease patients (5 – 8). These events are unordered because they occur in random sequence. Depending on whether events are correlated or not, two different approaches are available.

In the absence of correlation and dependent censoring the competing risk model has been suggested for analysis (28). In this case the likelihood of being censored at time t does

not depend on the reason for censoring including failure from a competing risk. The competing risk model is stratified by event type (basal risk allowed to differ) and gives the same results as the combined end point analysis (time to the first event that occurs) with the data layout containing only one record per subject. In the competing risk set the number of observations per subject is a multiple of the number of considered events (if there are k possible events, each subject will appear k times in the data layout, once for each possible failure). All failure times within the same subject begin at the same observation start date and continue until the follow-up end date in the absence of any event. If one or more event occurred the first event date is the termination time for the corresponding event type record and the censoring date in the remaining records from the same individual. Further events after the first are consequently ignored as in the combined end-point with one record per subject approach. The advantage of the larger data set is that it allows for easy estimation of within-event-type coefficients (stratum specific effects). The variance correction is not necessary in this independent competing risks model as each subject may have at most one event (23).

When there are reasons to believe that the data are correlated, it is possible to analyse multiple events per subject using the marginal model of Wei-Lin-Weissfeld (27). As in the previous model all times are measured from the date of patients' enrolment (time zero) but each observation continues in each stratum beyond the first event that occurred. An important characteristic of these failure events is that each can occur only once per

subject and that all subjects are at risk for all events as in the un-correlated competing risk sets model. This model is appropriate when the predictors under investigation are plausibly involved in the pathways leading to more than one event type and, therefore, the censoring mechanism for one event may be informative for the other. For example, plasma levels of asymmetrical di-methyl-arginine (ADMA) have been shown to predict both progression of chronic nephropathies and death in patients with chronic kidney disease (5). In these situations, the terminating time for observing one event could be correlated with the other and, as a result, the assumption of independent censoring may be violated. In addition, considering only time to the first event that occurs reduces the study power. The variance corrected and frailty approaches (with stratification by event type) make better use of the information in the data and the analyses are thus more powerful.

Examples of modelling unordered events of the different type using the chronic kidney disease data will be discussed in chapter 5.

5) Chapter 5: Semi-parametric Cox's model

Chapter overview

Chapter 5 introduces extensions of the Cox's model for correlated events of the same type (recurrent events) or different type (multiple events). Both pragmatic (variance corrected) and random effects (frailty) method are available for Cox's regression. The dialysis access data are used as example of recurrent event analysis. The Vitamin D data are used as example of multiple event (competing risks) analysis.

1. Variance corrected and frailty models for recurrent events of the same type: the AV access data revisited

Repeated failure time models can be fitted using the Italian AV access study data. Table 17 reports the estimates of the effect of the interaction between heart failure and late referral on the risk of AV access failure from models of time to the first event (1 record per subject, N = 535), and time to multiple events per subjects for ordered events of the same type (variance corrected and frailty models).

From the table it can be seen that the confidence intervals for the estimated effect of the interaction are wider (greater P value) in standard survival analysis than those from models for repeated events. This is a consequence of the smaller sample size (smaller event number). However, the coefficient estimates are close, indicating that such effect may be the same for any AV access created in the same patient. The stratum by covariate interaction is a formal way to test such hypothesis (not supported by these data).

The counting process formulation provides a slightly higher effect estimate as compared to the time to first event analysis. However, such estimate may be biased since the absence of stratification may fail to control for potentially different baseline risks as events recur and these "baseline" risks are averaged (23).

Table 17: Multiple primary failure time models for ordered events of the same type (access failure): fits of the effect of the interaction between heart failure and late referral on the risk of arterio-venous (AV) access failure. Models include the main effects (heart failure and late referral) and are controlled for history of vascular diseases (cerebral, peripheral and coronary artery diseases), and AV graft vs. fistula. All models are independent of gender, BMI, proximal vs. distal location, centre effect, diabetes, hypertension, chronic lung and systemic diseases.

Model	N	F	Beta	NSE	RSE	P (β=0)	HR	CI_{95} for HR	Theta	P (θ=0)
First event	535	222	0.671	0.363	0.377	0.075	1.96	0.934, 4.1	NA	NA
C Process	700	310	0.793	0.298	0.329	0.016	2.21	1.16, 4.21	NA	NA
Marginal	2140	310	0.946	0.298	0.404	0.019	2.58	1.16, 5.68	NA	NA
Elapsed T	700	310	0.734	0.307	0.308	0.017	2.08	1.14, 3.81	NA	NA
Gap time	700	310	0.701	0.302	0.303	0.021	2.02	1.11, 3.65	NA	NA
U Frailty	700	310	1.180	0.465	NA	0.011	3.25	1.31, 8.09	1.04	<0.01
C Frailty	700	310	0.701	0.302	NA	0.020	2.02	1.12, 3.64	<0.01	0.670

Legend: N = number of records; F = number of failures; Beta (β): regression coefficient; NSE: naïve standard error of beta; RSE: robust standard error (adjusted for clustering on patient identity); P (β=0): two-tailed significance level for hypothesis testing (using RSE of beta) of Beta being = 0; HR and CI_{95} for HR: hazard ratio and 95% confidence intervals for hazard ratio estimate. Theta (θ) variance of the frailty (gamma distribution); P (θ=0): two-tailed significance level for hypothesis testing of θ being = 0. C Process: counting process; Elapsed T: elapsed time (from entry); U/C: unconditional and conditional frailty models.

Table 18: Multiple secondary (assisted) failure time models for ordered events of the same type (access failure): fits of the effect of heart failure on the risk of arterio-venous (AV) access failure. Models are controlled for late referral, history of vascular diseases (cerebral, peripheral and coronary artery diseases), and AV graft vs. fistula. All models are independent of gender, BMI, proximal vs. distal location, centre effect, diabetes, hypertension, chronic lung and systemic diseases.

Model	N	F	Beta	NSE	RSE	P (β=0)	HR	CI_{95} for HR	Theta	P (θ=0)
First event	535	176	0.429	0.192	0.198	0.030	1.53	1.04, 2.27	NA	NA
C Process	700	245	0.451	0.161	0.186	0.016	1.57	1.08, 2.26	NA	NA
Marginal	2140	245	0.506	0.163	0.223	0.023	1.65	1.07, 2.57	NA	NA
Elapsed T	700	245	0.389	0.164	0.162	0.016	1.48	1.07, 2.03	NA	NA
Gap time	700	245	0.389	0.162	0.160	0.015	1.47	1.07, 2.02	NA	NA
U Frailty	700	245	0.507	0.208	NA	0.015	1.66	1.10, 2.50	0.89	<0.01
C Frailty	700	245	0.389	0.162	NA	0.017	1.47	1.07, 2.03	<0.01	0.900

Legend: N = number of records; F = number of failures; Beta (β): regression coefficient; NSE: naïve standard error of beta; RSE: robust standard error (adjusted for clustering on patient identity); P (β=0): two-tailed significance level for hypothesis testing (using RSE of beta) of Beta being = 0; HR and CI_{95} for HR: hazard ratio and 95% confidence intervals for hazard ratio estimate. Theta (θ) variance of the frailty (gamma distribution); P (θ=0): two-tailed significance level for hypothesis testing of θ being = 0. C Process: counting process; Elapsed T: elapsed time (from entry); U/C: unconditional and conditional frailty models.

In agreement with simulation studies, the marginal formulation provides an even larger estimated effect probably due to the lack of any order implication of the risk set organization (23).

Point effect estimates from conditional models (both variance corrected and stratified frailty models) are closer to those of time to first event (standard) analysis than all other models. In addition, the underlying assumptions of these risk sets is consistent with the biological and clinical understanding of some recurrent event processes such as AV access failure. Conditional models assume event time sequentiality (ordering) and varying basal risks, achieved through event stratification and specific risk times definition (i.e., elapsed times from entry, or gap times from previous events). Most importantly, this may impact on the dependencies among the failure times. This problem can be studied looking at the model standard errors.

Considering the variance corrected methods, it can be seen from Table 17 that the robust standard errors of the estimators are higher in all non-conditional models. The counting process and the marginal model do not address the sequential nature of the event order and assume independence in their structure (the former is not stratified and the second has a marginal time structure, i.e. the time at risk starts at the initial follow-up date for all records). This fact induces higher errors. As mentioned, correction for robustness is based on these errors. Since in the conditional models the sequential nature of the order

of events mostly accounts for the lack of independence, the variance correction in the calculation of the robust standard errors has very little effect (little difference between the naïve standard errors and the corresponding robust standard errors of the estimated coefficients). This feature of conditional models makes them especially attractive for multiple access failure because they account for both potentially different baseline hazards and event order.

Random effects modelling offers further insight about the sources of correlation in the data. Frailty models indicate that the true data generating process is characterized by significant individual (cluster) heterogeneity. In fact the variance of the random effects (theta) is significantly different from zero according to the standard unconditional (un-stratified) frailty model. However, this heterogeneity disappears when the model is stratified. This suggests that the heterogeneity is induced by event dependence as the main source of correlation in the data. In other words, it may not be simply that some individuals are predisposed to access failure in general because of continuing factors such as tendency to thrombosis. It seems rather that once a failure has occurred in susceptible (frail) individuals, consequences such as having to use sub-optimal vessels increase the risk of subsequent failure events.

The same results are obtained analysing the effect of heart failure on the risk of secondary (final or assisted) AV access thrombosis (Table 18). Point estimates from

conditional models are the closest to those from the standard time to first event model, and little residual dependencies are addressed by robust errors. In addition, conditional solutions to both primary and secondary survival modelling offered better fits (standard model diagnostics and assumption verification analysis not shown).

2. Variance and frailty models for multiple failure events of different type: the vitamin D data revisited

Unordered events of different type such as chronic dialysis initiation and death in chronic kidney disease patients can be studied as a function of some exposure of interest using marginal models (5 – 8). If the exposure is supposed to be associated with an event only multiple failure time analysis is not necessary. In such a case, the association of the predictor with either event may be tested using standard time to single event analyses first. For example, if the predictor is associated with death only, such as Urotensin II in chronic kidney disease (7), the other event (dialysis status) may be treated as a time varying covariate (absent versus present) if it is supposed to change the risk for the final event (death). In other cases the exposure under study is thought to be associated with multiple events (such as ADMA and the risk of both requiring dialysis and death). When failure times are correlated the marginal risk sets model offers interesting analytical solutions (5, 6). Both time varying covariate and event specific effects can be incorporated into these models (8).

In a recent study Vitamin D deficiency predicted greater risk for dialysis and death among Vitamin D naïve chronic kidney disease patients (8). Vitamin D deficiency is defined as levels of 25-hydroxy-vitamin D below 15 ng/ml (i.e., inadequate levels of the nutritional or storage form of Vitamin D). Although Vitamin D is mainly seen as a compound pivotal for bone physiology it is also central to optimal functioning of other organ systems including the cardiovascular, endocrine and immune systems. The study hypothesis was that Vitamin D deficiency may be a condition associated with (a marker of) both progression to dialysis and death.

Table 19 reports the estimates of the effect of Vitamin D deficiency on the risk for dialysis or death in 168 patients with chronic kidney disease (8). The record number is greater than the number of subjects even in time to single event models because some covariates were allowed to vary over follow-up (time dependent models). However, the constant proportionality assumption held for all such models (as well as for multiple failure times models) particularly once updated covariate values were modelled (9 – 13).

From Table 19 it can be seen that the effect of Vitamin D may differ by event type, being greater on the risk for dialysis than the risk for death. The observational nature of the study does not allow causal inferences regarding either of such effects, provided that they exist. In other words, causal effects are possible, but it is just as possible that Vitamin D levels decline as people near dialysis or death. This is true also for multiple failure times

Table 19: Multiple failure time models for unordered events of different type (dialysis and death): fits of the effect of Vitamin D deficiency (initial follow-up levels < 15 ng/ml) on the risk of dialysis (standard model), death (standard model), dialysis or death (independent competing risks – combined end-point) and dialysis and death (correlated competing risks – double end-point).

Model	N	F	Beta	NSE	RSE	P (β=0)	HR	CI_{95} for HR	Theta	P (θ=0)
Dialysis	406	48	0.969	0.370	0.295	0.001	2.63	1.48, 4.70	NA	NA
Death	406	78	0.731	0.297	0.267	0.006	2.08	1.23, 3.51	NA	NA
LMN	812	105	0.889	0.256	0.203	<.001	2.43	1.63, 3.63	NA	NA
WLW	812	126	0.825	0.230	0.198	<.001	2.28	1.55, 3.37	NA	NA
Frailty	812	126	0.825	0.231	NA	<.001	2.28	1.45, 3.59	<0.01	0.92

Legend: Cox's models controlled for time-varying levels of kidney function, phosphate and use of angiotensin antagonists (both events); proteinuria (dialysis); and heart failure, age, smoking habit, C-reactive protein, serum albumin (death). All models are independent of gender, BMI, season, other comorbidities or therapies (including 1,25VD supplements), and labs (including 1,25VD). N: record #; F: event #; Beta (β): effect of 25VD deficiency; NSE: naïve standard error of beta; RSE: robust SE (adjusted for clustering on patient identity); P (β=0): two-tailed significance level for hypothesis testing (RSE of β) of β = 0; HR and CI_{95} for HR: hazard ratio and 95% confidence intervals for HR estimate. Theta (θ): variance of the frailty (gamma distribution); P (θ=0): two-tailed significance level for hypothesis testing of θ being = 0. LMN: independent competing risk sets (Lunn-McNeil); WLW: correlated competing risk sets (Wei-Lin-Weissfeld).

Table 20: Multiple failure time models for unordered events of different type (dialysis and death): stratum specific effects of Vitamin D levels at the date of nephrology referral on the risk of dialysis or death (independent competing risks – combined end-point) and dialysis and death (correlated competing risks – double end-point).

Model		N	F	Beta	NSE	RSE	P (β=0)	HR	CI_{95} for HR
LMN	1 = dialysis	406	48	0.968	0.370	0.295	0.001	2.63	1.47, 4.70
	2 = death	406	57	0.814	0.356	0.308	0.008	2.26	1.23, 4.14
WLW	1 = dialysis	406	48	0.968	0.370	0.295	0.001	2.63	1.47, 4.70
	2 = death	406	78	0.731	0.297	0.267	0.006	2.08	1.23, 3.51

Legend: Cox's models specification as described in Table 19 legend. The exposure is Vitamin D (25VD) deficiency (levels < 15 v >= 15 ng/ml) at the date of nephrology referral. N and F: stratum size and event #; Beta (β): effect of 25VD deficiency; NSE: naïve standard error of beta; RSE: robust SE (adjusted for clustering on patient identity); P (β=0): two-tailed significance level for hypothesis testing (RSE of β) of β = 0; HR and CI_{95} for HR: hazard ratio and 95% confidence intervals for HR estimate. LMN1/LMN2: stratum specific estimates from the LMN (Lunn-McNeil) independent competing risk sets model; WLW1/WLW2: stratum specific estimates from the WLW (Wei-Lin-Weissfeld) correlated competing risk sets model.

analyses.

Considering the first event (independent competing risks), 48 patients required dialysis and 57 deaths occurred in 168 subjects over follow-up (105 combined end-points). However, 21 subjects died subsequent to starting dialysis and if it is believed that dialysis and death are not independent (i.e., those experiencing greater risk for dialysis may also be at higher risk for death), 126 total events could be modelled using a variance corrected or frailty approach. In the presence of data correlation the reasons for censoring due to a competing risk may violate the non-informative censoring assumption.

Assuming the effect of Vitamin D on dialysis and death to be the same, the independent competing risks approach estimate is less precise, possibly because of power loss. In addition, the point estimate is greater than the correlated event model. Checking the results from single event models, and stratum specific effects (Table 20 – see below) it may well be that the composite event approach is inaccurate. Of note, the difference between naïve and robust standard errors is not as evident as it was in marginal models for ordered events shown in the previous chapter. This may be due to the fact that the risk sets include updated covariate values (defined as sequential elapsed times) and the measurement of only two possible events per subject. Finally, the frailty model shows that the evidence in support of heterogeneity is very weak if the model is stratified by event type (as recommended for competing risks studies).

The estimates in Table 20 are from models including stratum specific effects, the effect of Vitamin D on dialysis and death under the independent competing risks assumption and effect of Vitamin D on dialysis and death under the correlated competing risks assumption. The results of these models differ not only in terms of point estimates but also in terms of precision. However, the difference in precision may be simply the result of different model power, again being greater for the correlated competing risk model. Importantly, the point estimates from this last model only are practically the same as those from standard time to each event models. Thus, if it is believed that Vitamin D deficiency is associated with both outcomes (and thus there are reasons to believe that the independent censoring assumption is violated), the correlated competing risks solution may be a better choice.

What are the advantages of using a single model for two possibly correlated outcomes as opposed to two separate models for each if under the same specification they offer the same estimates? First, there is a clear power advantage. If the two single event models have some covariates in common (as was the case for phosphate), or stratifying variables (as was the case for levels of kidney function and use of converting enzyme / angiotensin receptor blockade), a greater number of additional predictors (following the rule of ten – one parameter estimate every ten events) can be included into the model, and more freedom is left for testing interactions and potential confounders. Of note, for the

purpose of comparability, the single event models, the independent multiple event model and the correlated event model were all specified in the same way (in terms of stratum specific and common effects and stratification). Results would differ if adjustment was made by including the covariate in the model instead of using stratification (8). Second, if a decision has to be made as to whether to believe or not in a similar effect on both outcomes based on the data, the correlated event model with stratum specific effects offers the advantage of double estimation, and even statistical testing of such possible effect difference.

The power implications of including additional events after the first can be assessed comparing the standard errors of the combined end-point model and the correlated (variance corrected) event model (e.g., Table 19). Prior to conducting any analysis the expected information gain was 20% (126/105). However, the variance change was slightly greater at 23.8% (ratio of the naïve variances). Thus, it is expected that each repeated event is worth at least as much as but probably even more than a new first event. This happened to be the case in this example. In fact, the expected naïve standard error (NSE) of the coefficient associated with Vitamin D deficiency in the multiple event model (event number 126) was close to (but slightly greater than) the estimated value from the model, and greater than the robust standard error (RSE). This expected NSE equals the observed NSE of the combined end-point model times the square root of the proportion of independent events: $0.256 * \sqrt{(105/126)} = 0.233$ (vs. the estimated values of

0.230 for the NSE and 0.198 for the RSE. Thus, the corresponding expected event number was $(0.256*\sqrt{105}) / 0.198$ to the power of 2, i.e. 175.5. This resulted in a realized information gain of (175.5 - 105)/(126 - 105) = 3.35. In other words, each further multiple event was worth about three times of a new first event (event number gain = 69).

Time independent models (i.e., models without time varying covariates) fitted to the same study data (8) resulted in similar information gain (ratios greater than one). Similar findings were obtained in different studies (23). These information analyses from observational data are important as they can be used to plan an event driven clinical trials whereby study size and duration are a function of the event number. For example, randomized clinical trials of an intervention impacting more than just one outcome can be planned to record multiple events in the same subjects. This is important because extending follow-up of fewer subjects can often be less costly than shorter follow-up of larger samples, and similarly efficient, provided that the information gain deriving from additional events in the same subjects is reasonable, the likelihood of experiencing further events is high, or the number of possible additional events is high. When there are reasons to believe that the intervention under study may affect a multiple failure events process, prior information data from observational studies may be used to estimate the information gain deriving from such multiple failure times measurements.

One obvious question at this point is whether different outcomes could be weighted

according to severity. For example, investigators planning a trial of Vitamin D supplement in chronic kidney disease may believe that patient death is twice as momentous as renal death, but still be willing to use both outcomes to compare intervention and control groups. This can be easily accomplished by giving each observation in the death stratum a weight of 2 and reassessing the information gain resulting from the use of further events in the same subjects.

6) Chapter 6: Parametric analysis of correlated recurrent events: the access data revisited

Chapter overview

Chapter 6 introduces fully parametric approaches to correlated recurrent events. Dialysis access data are used as example for either event count modelling or repeated failure time regression.

1. Models for event counts

The AV access study was planned to collect event history (survival) information on the first and further AV accesses created in an incident cohort of hemodialysis patients during a 6 year follow-up. These survival data can be aggregated by combination of the level of relevant covariates, leading to grouping of exposure times and sum of failure events (counts). Such data aggregation resulted in 78 records for primary failure count analysis and 80 records for secondary failure count analysis from the AV access study. Alternatively AV access level information can be used for count models assuming that the underlying failure process follows a Poisson distribution.

Table 21 and Table 22 report the effect of the interaction between late referral and heart failure on the risk of primary and secondary failure of any AV access created in the patients during follow-up. There were no stratum specific effects of any of the covariates in the models. All models are controlled for the effects of the main terms (late referral and heart failure), vascular disease, access type (prosthetic graft versus native fistula) and time bands (0-3, 3-6, >6 months). This last adjustment was necessary to reasonably meet the constant baseline incidence rate assumption implied by the adopted model.

Models of aggregated data

For each failure type (primary and secondary failure) five models are presented in Table 21: the first is the standard Poisson model; the second is its robust variance version; the

Table 21: Models for count data: aggregated data set

Models	β	SE	Exp(β)	CI_{95} of Exp(β)	P (β=0)	Extra parameters
Poisson (1)	0.704	0.298	2.02	1.12, 3.62	0.018	-
Poisson (2)	0.679	0.326	1.97	1.04, 3.74	0.037	-
RSE Poisson (1)	0.704	0.236	2.02	1.30, 3.21	0.003	-
RSE Poisson (2)	0.679	0.246	1.97	1.38, 3.20	0.006	-
SSE Poisson (1)	0.704	0.375	2.02	0.96, 4.22	0.061	Scale = 1.58
SSE Poisson (2)	0.679	0.324	1.97	1.04, 3.72	0.036	Scale = 0.98
BS Poisson (1)	0.704	0.294	2.02	1.13, 3.60	0.017	
BS Poisson (2)	0.679	0.288	1.97	1.12, 3.47	0.019	
NBR (robust) (1)	0.704	0.236	2.02	1.30, 3.21	0.003	θ = 0
NBR (robust) (2)	0.679	0.246	1.97	1.38, 3.20	0.006	θ = 0

Legend: Estimated effects of the interaction between late referral and heart failure from models of primary (1) and secondary (2) AV access failure. All models are controlled for the main terms (late referral and heart failure), access type (graft versus fistula), vascular disease and time band effects (0-3, 3-6, >6). N of records are 78 (1) and 80 (2). β: beta coefficient (log-rate difference); SE: standard error of β; Exp(β): Incidence Rate Ratio; CI_{95}: 95% Confidence Intervals of Exp(β); P value of the null hypothesis (β=0); θ: variance of the random effect (gamma distribution); RSE: robust SE; SSE: SE scaled using the square root of Pearson X2-based dispersion; BS: bootstrapping; NBR: negative binomial regression.

Table 22: Models for count data: AV access level data

Models	β	SE	Exp(β)	CI_{95} of Exp(β)	P (β=0)	Extra parameters
VC Count Models						
RSE Poisson (1)	0.704	0.314	2.02	1.09, 3.74	0.025	-
RSE Poisson (2)	0.679	0.335	1.97	1.02, 3.89	0.043	-
GEE (robust) (1)	0.689	0.317	1.99	1.06, 3.71	0.030	18 ρ parameters
GEE (robust) (2)	0.841	0.376	2.31	1.10, 4.84	0.025	18 ρ parameters
RE Count Models						
RI Poisson (1)	1.675	0.598	5.34	1.65, 17.2	0.005	θ = 2.7 (1.6, 3.8)
RI Poisson (2)	1.060	0.504	2.88	1.07, 7.75	0.035	θ = 1.1 (0.0, 2.2)
RC Poisson (1)	1.665	0.597	5.28	1.63, 17.0	0.005	θ_{11} = 2.8 (1.4, 4.2) θ_{12} = -0.1 (-.3, .15) θ_{22} = 0.1 (0.0, 0.2)
RC Poisson (2)	1.054	0.506	2.87	1.06, 7.75	0.037	θ_{11} = 1.1 (0.0, 2.3) θ_{12} = 0.0 (0.0, 0.0) θ_{22} = 0.0 (0.0, 0.0)

Legend: Estimated effects of the interaction between late referral and heart failure from models of primary (1) and secondary (2) AV access failure, controlling for the main terms (late referral and heart failure), access type (graft versus fistula), vascular disease and time band effects (0-3, 3-6, >6). N of records: 700; number of clusters (subjects): 535. β: beta coefficient (log-rate difference); SE: standard error of β; Exp(β): Incidence Rate Ratio ; CI_{95}: 95% Confidence Intervals of Exp(β); P value of the null hypothesis (β=0); θ: variance of the random effect (normal distribution); RSE: robust SE; GEE: generalized estimating equations; RE: random effects; RI: random intercept only; RC: random coefficients (intercept and slope, and their covariance). Variance subscripts for the random effects (95% CI): θ_{11} = RI (extra-variation due to subjects), θ_{22} = RC (extra-variation due to varying response to time bands), θ_{12} = covariance of RI and RC. For both (1) and (2) the variance of the RC (θ_{22}) was not significantly different from zero (NS likelihood ratio test of the RC and nested RI model).

third is its "Pearson chi square" scaled residual version; the fourth is its bootstrapping residual version; and the last is the negative binomial model with robust standard errors[6].

Results from all models indicate that the risk of both primary and secondary failure is twice as high in the presence of both late referral and heart failure (whereas each factor alone does not alter the incidence rate significantly – not reported but consistently found in all interaction models). Point estimates from robust, scaled and bootstrapping Poisson regression models are the same as those from the standard version of the model (as they must be) because these models correct the coefficient "variance" for the dependencies in the data. Of note, robust methods generate smaller standard errors in this example (with narrower confidence intervals) than standard Poisson regression (usually the opposite happens) and the residual scaling approach. Scaled variance models show that if a variance inflation exists, such inflation affects only primary failure data (scale > 1).

---

6 Unlike the normal distribution the Poisson distribution has no separate parameter for the variance but this is equal to the mean. More precisely, the variance is a function of the mean with dispersion parameter $\varphi = 1$ (whereas the variance of the normal distribution is an identity function of the dispersion parameter $\varphi = \sigma^2$). Different approaches exist to account for the extra-variability in the parameter μ of the Poisson distribution (i.e., if the variance is greater than the mean) due to omitted covariates or correlation in the data. One way is to allow μ to vary randomly to some extent (i.e., the variability unexplained by fixed effects) according to some (prior) distribution and assume that conditional on that random variation the response variable follows the Poisson distribution (random effects models). There are also pragmatic ways to address the problem of over-dispersion. One way is to assume that the variance is proportional to the variance function but estimating a dispersion or scale parameter φ rather than assuming the value of 1 appropriate for the distribution. This scale parameter can be estimated from the deviance or Pearson chi square $\Sigma(y - \mu)^2 / V_\mu$ where V_μ reduces to μ under the Poisson model. This scale parameter is then used to estimate the standard errors of the regression coefficients (quasi-likelihood estimation). Robust methods offer empirical estimates of the standard errors by removing the observations one at a time in the process of computation (one cluster at a time when clustering information is available – chapter 3). Finally in bootstrapping the sample is resampled with replacement to approximate what would happen if the whole population were sampled, and estimate the coefficients and their variance.

Negative binomial models incorporate also a random effect term in the equation, which may generate point estimates different form those of the standard Poisson model. This random effect is "record specific" or, more properly "combination specific" since data are aggregated (as opposed to "cluster specific"), and its variance is estimated from a theoretical distribution. There is no evidence from the data of this example that the variance of such random factor (if it exists) is different from zero. This indicates that the aggregated data do not show important heterogeneity. Results of the negative binomial regression were the same using standard rather than robust variance estimation (not shown). This means that once available information has been accounted for (including interaction terms between time bands and access number in these models) there remains little extra-variability in the data (over-dispersion)[7].

Count model of AV access level data

Count data can be analysed using the AV access information. One way is to treat the correlation among AV accesses from the same subjects using variance correction (although scaled residuals and bootstrapping are other possible approaches). Variance correction in such cases is made considering clusters of observations, i.e. treating

7 The Poisson distribution assumes that the variance be equal to the mean. This assumption is often violated in real life count data which presents variance greater than the mean (over-dispersion). Other common complications are zero inflation (excess zeros in the data) or zero truncation by designs (for example enrolment of subjects experiencing events only, or exclusion from the analysis of those without event). Negative binomial regression is useful to account for extra-variability unexplained by the measured covariate. Negative binomial versions for zero-inflated or zero-truncated models exist.

observations from different clusters as independent. Both variance corrected models replicate the results of aggregated data. In addition to variance correction (cluster robust variance), generalized estimating equations incorporate the assumed correlation structure of the data in the variance estimation. As opposed to correcting the estimators' variance only, this can have an impact on the point estimates. An unstructured working correlation matrix has been specified for these data as this is the most flexible – although the least parsimonious – approach (Table 14). GEE use the suggested working correlation to start the iterative estimation process, and then refine the estimates of the correlation matrix values using the observed data. From Table 21 it can be seen that results provided by these regression methods are consistent with those from previous models. The additional (nuisance) parameters estimated by the GEE are those of the final working correlation matrix from which the standard errors of the coefficients are estimated (values not shown).

Table 22 shows that the estimates generated by random effects Poisson models differ from those of pragmatic methods, especially for primary failure models (where aggregated data analysis showed the existence of some amount of extra-Poisson variation). It is possible that splitting some observations into time bands to account for the effect of time (to meet the constant incidence rate requirement) may have induced a type of over-dispersion that random effects fail to account for, or require extra-work to be

managed properly[8]. In fact, estimates were even larger if the chosen distribution for the random effects was the gamma distribution, rather than the chosen normal distribution (not shown). This limitation should not be seen as a limitation of random effects models in general, but rather as a problem related to the data set example chosen for the current discussion. The additional parameters estimated by random effects models are the variances of the frailties (or random effects). In the first two models only a random intercept has been assumed. Additional random effects (random slope in the last two models) did not improve significantly the model fit (non significant likelihood ratio test, not shown). However, these random coefficients models are reported to show the amount of extra-parameter estimation. It is important to notice that for any pair of random effects (1 intercept and 1 slope for example, but similarly for two slopes or two intercepts) the model estimates 3 parameters: the two variances plus the covariance of the two random effects. More generally, the number of estimated parameters in an "n" random effects model equals $n + n^*(n - 1) / 2$. For example, if the data contains two levels of clustering (e.g., one intercept for centre and one intercept for subject) and two random slopes (e.g., individual change in the response due to age and time), then the model will estimate $n = 4$ random effects plus 6 covariances (for a total of 10 parameters). An advantage of mixed effects modelling is the possibility to specify more than one level of correlation in the data and different types and number of random effects (intercepts and slopes). However,

8 For example, a within group correlation structure may need to be specified often for the computation of the residual matrix. This is done when the correlation in the data is not fully accounted for by the random effects (i.e., residual correlation remains after the random effects have been fitted). Some packages only are currently capable to run such models (S-Plus, R and SAS).

a thorough consideration of the benefits deriving from the inclusion of more than two or three random effects should be assessed carefully. In fact software procedures for non-linear mixed effects models currently run very slowly when estimation of random slope variances is required.

The current AV access data have only two layers of random errors (AV access at the lower level and subject at the higher level). Random effects due to subjects are meant to adjust for the error dependencies at the lower level. These random effects due to subjects were estimated through the random intercept only, though the amount of change in the incidence event rates as other covariates change may vary significantly by subjects (additional random slopes). There was no centre effect (upper level correlation), probably because the three participating centres were characterized by very similar policies and patient characteristics (centre random effects estimation would require a larger number of centres for reliable estimation). The variance "theta" reported in Table 22 refers to: 1) the extra-variability in the intercept (baseline incidence rate or incidence rate when all the model covariates are set to zero – or to the mean for continuous variables) unexplained by the covariate and due to subject characteristics (RI models); or 2) both to the intercept and slope or "extra-change" in the incidence as time passes by (RC models). This random slope estimates the extra-variability of the effect of time band due to different impacts of the effect of time in different subjects. Two additional parameters are estimated in RC models: the variance of the random slope and the

covariance between random intercept and slope (or correlation, if the covariance is divided by the product of the respective standard deviations). The sign and the size of this correlation between random effects is important. From this correlation it is possible to estimate the size and direction of the extra-variability in the response variable as the values of the slope covariate changes (heteroskedasticity). Prediction of the response by levels of the fixed effects of interest can be estimated taking into account both the residual error (standard errors) and the random variation components (intercept, slope and their covariance). The meaning of either one, or several random effects is the same: they account for the extra-variability in the data (called frailty for risk data) unexplained by the (fixed) covariates. If the mixed model is correctly specified, the estimated fixed effects are true (unbiased) conditional on the frailty effects. These effects (parameters) are interpretable as individual effects, as opposed to the population average effects estimated by marginal (robust) methods described above.

2. Models for repeated failure times

Different survival models can be fitted to the AV access data to estimate the effects of the interaction between late referral and presence of heart failure.

Robust methods

Table 23 shows the results from three robust variance models of primary and secondary AV access survival: the semi-parametric Cox's model in its gap time (conditional from

Table 23: Robust (cluster) survival models for continuous failure times data

Models	β	SE	Exp(β)	CI_{95} of Exp(β)	P (β=0)	Extra parameters
GT Cox's (1)	0.752	0.300	2.12	1.17, 3.82	0.012	Reference (1)
GT Cox's (2)	0.665	0.338	1.94	1.00, 3.77	0.050	Reference (2)
Exponential (1)	0.810	0.323	2.24	1.19, 4.24	0.012	$\beta_{0\text{-}3}$; $\beta_{3\text{-}6}$; $\beta_{>6}$; β_{zav}
Exponential (2)	0.734	0.335	2.08	1.07, 4.02	0.029	$\beta_{0\text{-}3}$; $\beta_{3\text{-}6}$; $\beta_{>6}$; β_{zav}
Weibull (1)	0.738	0.309	2.09	1.15, 3.79	0.015	p; $\beta_{0\text{-}3}$; $\beta_{3\text{-}6}$; $\beta_{>6}$; β_{zav}
Weibull (2)	0.722	0.317	2.05	1.10, 3.83	0.023	p; $\beta_{0\text{-}3}$; $\beta_{3\text{-}6}$; $\beta_{>6}$; β_{zav}

Legend: AV access data from 535 individuals receiving 700 accesses and experiencing 310 primary and 245 failures over 6 years. From each model the estimates of the interaction between late referral and heart failure are reported on the risk of primary (1) and secondary failure (2). Estimates are controlled for the effects of the main terms, access type, vascular disease and time band effects (0-3, 3-6, >6) for parametric models. β: beta coefficient (log-hazard difference); SE: standard error of β; Exp(β) Hazard Ratio ; CI_{95}: 95% Confidence Intervals of Exp(β); P value of the null hypothesis (β=0); GT: Conditional (Gap Time) Cox's models form previous event. As compared to the corresponding Cox's model the exponential model estimates 4 additional parameters: the effects of 3 time bands and access number ("zav" – used as stratifying variable in the Cox's models and thus not estimated); the Weibull model estimates 5 additional parameters: the effects of 3 time bands, access number and the shape parameter "p" (although only 2 time bands were necessary in this model – 0-6 and >6 – 3 bands were used for comparability).

entry) formulation[9], the exponential model and the Weibull model. The Cox's models (for recurrent primary and final events) are considered the reference model for these data following the intention to identify prognostic factors and their estimated effects in terms of risk change. The model fit is excellent for both primary and secondary survival and neither assumption violations were found, nor significant residual problems. It can be seen from Table 23 that robust estimates under parametric constraints provide similar results (the log-hazard parameterization is chosen for comparability). However, both parametric models estimate extra-parameters: the intercept (hazard when the covariates are all zero), the effects of time band (2 necessary for the exponential model, and 1 only for the Weibull, although a second was included (even if not significant) for reasons of comparability), and the hazard change in second and further access versus the first (used as a stratifying variable in the Cox's model leaving its effect un-parameterized). There was no evidence of any covariate effects on the shape parameter of the Weibull model[10]. Thus the Weibull model includes also one ancillary parameter "p" responsible for the model specific hazard shape (decreasing hazard, Figure 13). Had any covariate altered the hazard shape only, or both the shape and the scale of the hazard, its effect would have been estimated as shape parameter only or both as "p" specific in addition to the beta

---

9 Results differ from those reported in Table 17 and Table 18 because of the different definition of the stratifying variable and the exclusion of the interaction term from the models reported in Table 18. As opposed to the models in Table 17 and Table 18 where the number of strata was equal to the maximum number of accesses received by the patients (four), the stratifying variable "access number" in the models reported in Table 23 and Table 24 was simplified to a two level variable (first versus subsequent AV accesses) because the baseline hazards of failure of the second, third and fourth access were similar.

10 The Weibull model allows one (or more) covariates to affect or even change the sign of the ancillary parameter log-p leading to two (or more) specific hazard shapes within the same model.

coefficient (shape or / and scale change).

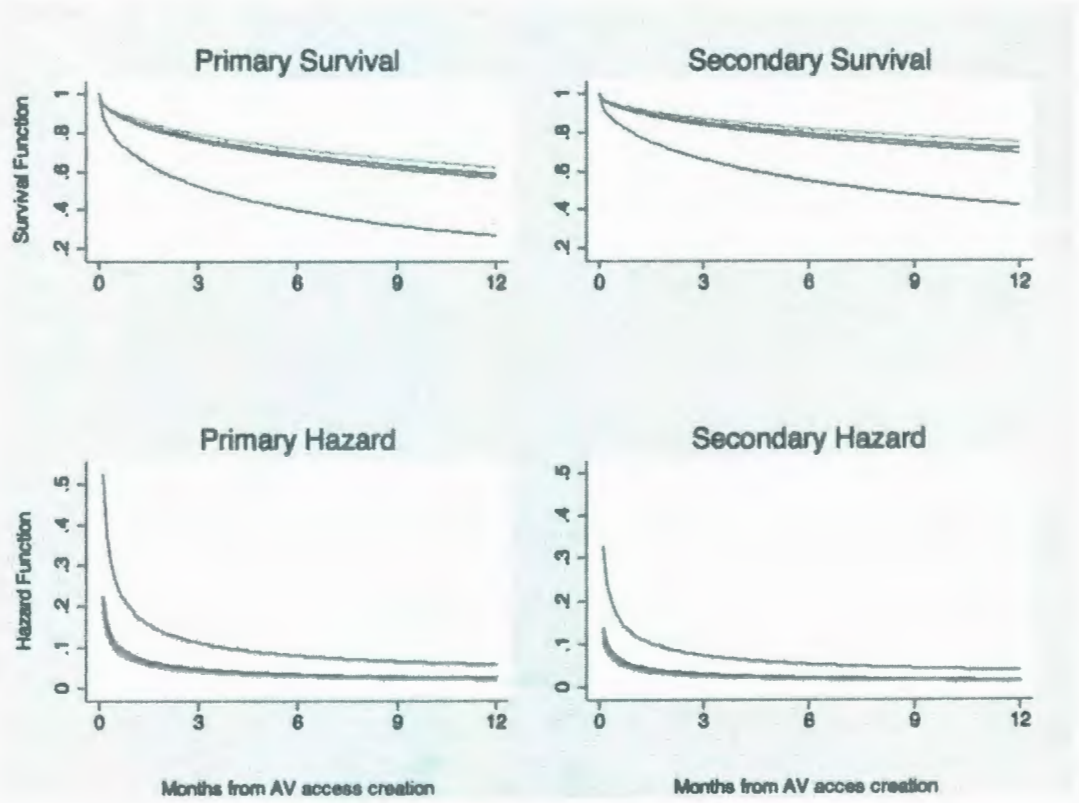


Figure 13: Primary and secondary survival and hazard function from the robust variance Weibull regression model of the recurrent AV access data reported in Table 24. Higher survival curves refer to the presence of both late referral and heart failure; the other curves to the presence of either or none of such risk factors.

Thus the chosen parametric models provide estimates of the effect of interest at the cost of extra-parameter estimation work. Precision has not improved substantially in parametric models of primary survival, but the evidence in support of an effect of the exposure (interaction between heart failure and late referral) on secondary survival is clearly stronger. There are advantages deriving from the choice of parametric models when their hazard scale estimates are consistent with those from the reference Cox's model and the chosen model shape is consistent with the observed data (Figure 14). First, there are advantages in terms of power because more parameters can be estimated in parametric regression models than in the corresponding Cox's models. In fact the estimation method of the parametric models makes use of all the information in the data and not only from the times when events occurred. Second when a phenomenon is sufficiently known to believe that a certain hazard shape can be modelled, estimates from parametric models are often more efficient (precise). Last, but not least, results can be expressed using the failure time interpretations and thus estimates and predictions of survival times can be obtained (Chapter 2).

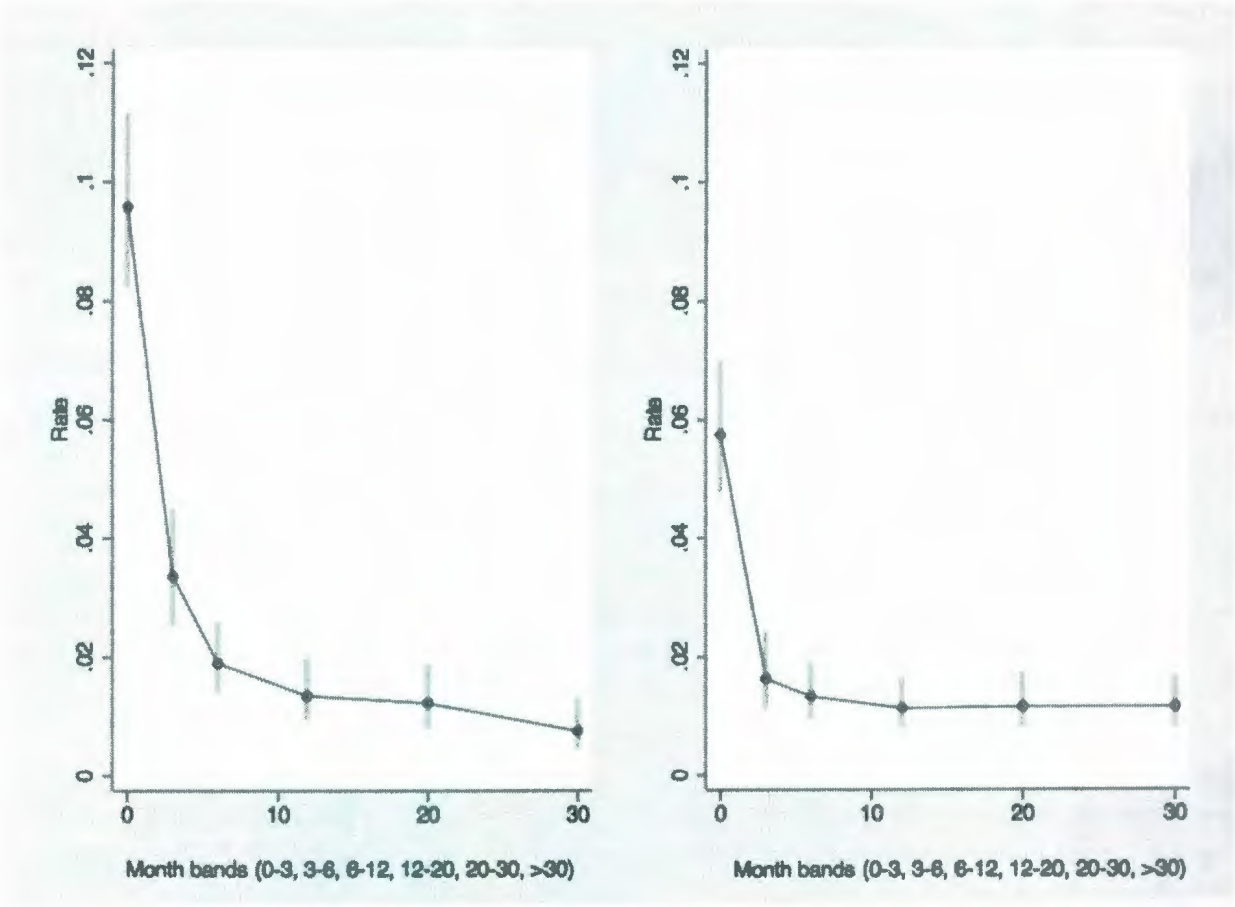


Figure 14: Observed primary and secondary recurrent event rate over time bands in months.

Frailty models

Survival models can be used to estimate the unshared frailty or heterogeneity among observations (corresponding to the random effect variance estimated by the negative binomial model) and the shared frailty due to within subjects dependencies. Frailty models can be best understood comparing and contrasting these unshared and shared frailties, which is possible only for parametric models. In fact unshared frailty models for reasons of identifiability do not exist with Cox's regression because studying extra-variability is impossible in the absence of any distributional assumptions.

Estimates from unshared frailty models in Table 24 confirm the existence of important heterogeneity in the data as already noted from Cox's analyses (Chapter 5). This heterogeneity results from differences in the risk of failure (frailty) *over observations* unexplained by the fixed effects. When the frailty is modelled as shared, i.e. *over clusters*, the coefficient estimates are closer to those from the reference Cox's models (assuming that these are the true ones). In either case, the frailty is an unobserved multiplicative effect on the hazard function of the individual observation or cluster of observations assumed to follow some positive distribution (the hazard can not be negative). The variance parameter θ of the random effects (over individual observations or clusters) is estimated from the data. As previously noted, stratification is an efficient method to control for the dependencies under the Cox's model (the variance of the

Table 24: Frailty survival models for continuous failure times data

Models	β	SE	Exp(β)	CI_{95} of Exp(β)	P (β=0)	Extra parameters
US Expon. (1)	1.438	0.582	4.21	1.34, 13.2	0.013	> 0; (4)
US Expon. (2)	1.418	0.643	4.13	1.16, 14.5	0.028	> 0; (4)
US Weibull (1)	1.402	0.574	4.06	1.31, 12.5	0.015	> 0; p(1); (4)
US Weibull (2)	0.722	0.326	2.05	1.08, 3.90	0.027	= 0; p(1); (4)
SH U Cox's (1)	1.246	0.561	3.47	1.15, 10.45	0.026	> 0
SH U Cox's (2)	0.963	0.554	2.63	0.88, 7.77	0.082	> 0
SH C Cox's (1)	0.653	0.298	1.92	1.07, 3.45	0.029	= 0
SH C Cox's (2)	0.606	0.328	1.83	0.96, 3.49	0.065	= 0
SH Expon. (1)	1.641	0.584	5.16	1.64, 16.2	0.005	> 0; (4)
SH Expon. (2)	1.326	0.563	3.76	1.24, 11.3	0.019	> 0; (4)
SH Weibull (1)	1.251	0.487	3.49	1.34, 9.08	0.010	> 0; p(1); (4)
SH Weibull (2)	0.995	0.474	2.70	1.06, 6.86	0.036	> 0; p(1); (4)

Legend: AV access data (700 accesses in 535 subjects; 310 primary and 245 secondary failures over 6 years). From each model the estimates of the interaction between late referral and heart failure are reported on the risk of primary (1) and secondary failure (2). Models controlled for the main terms, access type, vascular disease, access number and time band effects (0-3, 3-6, >6) for parametric models. β: beta coefficient (log-hazard difference); SE: standard error of β; Exp(β) Hazard Ratio ; CI_{95}: 95% Confidence Intervals of Exp(β); P value of the null hypothesis (β=0); θ: variance of the random effect; extra-parameters versus reference models in Table 23; US / SH: un-shared / shared frailty models; U / C: unconditional / conditional (stratified) frailty Cox's regression.

random effect becomes zero after stratification). In the corresponding shared parametric regression models the point estimates are still larger. This may be due to the fact that the data are extremely unbalanced (Figure 2). When data are unbalanced due to random "missingness", maximum likelihood estimation takes into account the correlation in the data in the process of computation and works using available data to provide the best guess of what the data would be had they been fully available. When there are missing data, the ordinary least square estimate of the mean (which ignores random effects) is affected by the values of missing data (e.g., if larger values are missing the sample mean will underestimate the true parameter value). In random effects models, maximum likelihood estimation provides parameter estimates closer to those of the full data set prior to random data being lost or eliminated in simulation studies.

An obvious question at this point is how to consider the current data set. Since data have been measured prospectively, data unbalance is the result of different subject frailty to succumb to disease and the limited observation time. Random effects models take into account the information in the data and provide estimates of what would have happened had the follow-up been longer / the data complete. This must be interpreted with caution and prior knowledge about the failure process has a role in making choices. For example, for the AV access story one may hypothesize that thorough longitudinal measurements remain a false version of truth, and time constraints or even competing risks may have induced underestimation of the true effects. However, in other circumstances this may be

not the case and pragmatic approaches (including parametric robust models) may represent safer choices.

Frailty parametric models offer other useful insights to clinical researchers. The ancillary parameter "p" of the primary robust Weibull survival model in Table 23 for example, is estimated at 0.45 (95% CI 0.4 to 0.49). The ancillary parameter implies that the "individual hazard" of primary failure is monotonically decreasing over time. In other words, as time passes by, the instantaneous risk of AV access failure each individual faces falls. If this is true, clinicians should tell their patients that if thrombosis has not taken place by a certain time after the AV access has been created then they need not worry too much as AV access failure is less likely as time goes by. On the other hand these estimates are obtained under the untenable assumption that all patients are identical other than the covariates in the model (late referral, heart failure, their interaction, access type, vascular disease and access number). When subjects differ in unobserved ways in their inherent risk to succumb (frailty) then the estimated ancillary parameter deserves more careful interpretation. The frailty version of the Weibull model of primary failure reported in Table 24 provides similar estimate of the ancillary parameter "p" (0.56, 95% CI 0.5 to 0.63) with significant variance "theta" of the frailty effects (2.2, from 1.4 to 3.4). In some situations the frailty term affects substantially the shape of the individual hazard, in terms of value and / or sign of the (log) ancillary parameter, or even number of ancillary parameters (i.e., parameterization). In fact, frailty parametric models allow

estimating two hazard curves: one describing the individual hazard pattern, with shape belonging to the chosen family distribution, and the other describing the population pattern, which is usually decreasing over time. Although both the individual and the population hazard declined over time in the AV access study, the two curves may differ in some studies. For example, Cleves et al re-analysed a dialysis catheter infection study using parametric models (30). They showed that according to the frailty Weibull model (which best fit the data), despite the fact that the population hazard for catheter infections initially increased after catheter placement and then fell as time passed by, the individual frailty monotonically increased ($p > 1$). In all frailty models individual and population hazards should be distinct because population and samples are heterogeneous. In fact in a heterogeneous population the population hazard can fall while the individual hazards all rise because over time more and more robust individuals survive while the more frail succumb. This phenomenon is known as *frailty effect* and results in a declining population hazard independent of the shape of the individual hazards. Parametric models allow exploration and distinction of these hazards as they may have different patterns when the variance of the frailty effect is significantly different from zero. These models may also be useful in the choice of the proper parameterization of the underlying failure risk process. Finally they tend to coincide as the variance of the frailty becomes zero because then the frailty model reduces to the standard regression model without frailty.

A final important point should be made about the implications of the frailty term on

model assumptions and effects interpretation[11]. If a (shared) frailty component is added to Cox's regression then the PH assumption is not satisfied for the population hazard. The estimated HR are thus individual level effects. In other words, the HR are "conditional on the same level of frailty". Similarly, for parametric models it is important to consider the possible existence of a (shared or un-shared) frailty effect. In presence of a frailty effect the PH assumption is satisfied at the individual level but not at the population (unconditional) level. Conversely, the AFT assumption is satisfied at both levels[12]. For this reason model checking and interpretation of a frailty model must take into account the possible existence of a frailty effect.

---

11 The most popular distribution chosen for the frailty is the gamma distribution, but others exist such as the inverse-normal distribution. The choice of the distribution affects the interpretation of the fixed effects. For example, in hazard-metric frailty models the exponentiated coefficients are interpretable as hazard ratios when t = 0. After time zero, as more frail subjects leave the population at risk, differences in the fixed effects level off. In gamma frailty models this vanishing of the fixed effects is eventually complete in favour of the frailty effect. Conversely, in inverse-Gaussian frailty models the fixed effects never disappears but tend to the square root of the effects at time zero (30).

12 The unconditional or population level hazard $\lambda u(t)$ with gamma frailty is $\lambda u(t) = \lambda(t) / \{1 - \theta \log[S(t)]\}$. Population level hazard ratio HRu for one unit change of an exposure is $\lambda u2(t) / \lambda u1(t) = HRu$. This HRu is estimated as $HRu = \{\lambda 2(t) / \lambda 1(t)\} \{1 - \theta \log[S1(t)]\} / \{1 - \theta \log[S2(t)]\}$, where $\lambda 2(t) / \lambda 1(t) =$ HR (PH) but where the last ratio is not constant as it depends on time. In the AFT metric the time ratio TR assumption holds instead because the unconditional survival probability $Su(t) = \{1 - \theta \log[S(t)]\}^{-1/\theta}$. This implies that if $S1(t) = S2(TR*t)$ then $Su1(t) = \{1 - \theta \log[S2(TR*t)]\}^{-1/\theta} = Su2(TR*t)$.

7) Chapter 7: Conclusions

Chapter overview

This concluding chapter presents an overview of the thesis work and summarizes the key concepts of risk estimation in longitudinal studies.

Risk estimation in longitudinal studies: challenges and opportunities

Recurrent events of the same type (e.g., infections episodes) and multiple events of different type (e.g., myocardial infarction and renal failure) are often observed in longitudinal studies. Risk estimation of these failure events poses analytical challenges due to the correlation of repeated events in the same subject. A common approach to evaluate risk in longitudinal studies is to model time to first event using standard regression methods, and disregard further events once the first has occurred. This approach may not only reduce the study power but also lead to biased estimates. Analytical methods for recurrent or multiple events are available taking into account the lack of independence of repeated observations in the same epidemiological unit.

In this thesis work the following steps are proposed for the design and analysis of longitudinal studies:

Consideration of the characteristics of the failure event process;

Identification of the possible reasons for the correlation in the data;

Choice of the statistical model that best represents the failure event process;

Organization of the data set to accomplish the analytical task.

Consideration of the characteristics of the failure event process. It is important to understand whether the failure process generates ordered recurrent events of the same type or multiple unordered events of different type. The Vitamin D study in CKD was

used as an example of two competing events of different type. In that study, the exposure of interest (Vitamin D deficiency) was hypothesized to be involved in the mechanism of both failure processes, one leading to death and the other to dialysis initiation. As compared to standard Cox's regression of separate events (used as reference models), the composite event model provided biased estimate of the relative risk (HR) for death in Vitamin D deficient individuals (HR 2.26 v to 2.08). This is probably due to the fact that composite end point models disregard further events after the first. This may be wrong when data are correlated. Composite end point analysis is valid (including analysis of competing risks) when events are not correlated, i.e. when the likelihood of being censored at time t does not depend on the reason for censoring (including competing events). When events are correlated the likelihood of being censored at time t does depend on the reason for censoring. The advantage of the competing risk model, as opposed to separate single event models, is that when stratum specific effect estimates are very close to each other, single effects can be estimated for multiple events, and fewer parameters are estimated in the model. Similar considerations apply to ordered recurrent events of the same type. The reanalysis of the AV access study data shows that the analysis of the first event only leads to imprecise estimates. For example the existence of an interaction between heart failure and late nephrology referral on the risk of AV access failure was not supported by time to first event analysis (HR 1.96, 95% CI from 0.93 to 4.1) but it was supported by recurrent event analysis (HR 2.02, 95% CI from 1.11 3.65). To model these recurrent events of the same type using a survival approach there are

several choices. However, most often the order condition and the difference in the baseline risks are important issues to be accounted for. For example, some models (marginal risk sets) are more appropriate for repeated hospitalizations, and others (conditional models) for repeated AV access failure or peritonitis episodes.

Identification of the possible reasons for the correlation in the data. Biological and physio-pathological mechanisms for both subject heterogeneity and event dependence need to be carefully assessed. Both individual heterogeneity and event dependence can be sources of correlation in the data. In the presence of event dependence without important heterogeneity the true variance of the frailty is close to zero. This seemed the case in both data sets examined in this thesis. However, when the variance of the frailty is very small or not significantly different from zero one can use pragmatic approaches such as stratified variance-corrected methods within either a semi-parametric or parametric context. The un-stratified frailty Cox's model may detect a random effect signal which is really the consequence of event dependence rather than heterogeneity. In presence of heterogeneity without event dependence stratification would not be necessary since the baseline risk does not change by event number. In this case variance corrected models may be inefficient and the unconditional frailty model would perform better. Yet, if repeated events data exhibit both heterogeneity and dependence, a stratified frailty model would be highly desirable.

Choice of the statistical model that best represents the failure event process. When there are reasons to believe that the risk change over time can be described well by some theoretical hazard distribution, then the choice of a parametric model may offer several advantages. Using the Cox's model as gold standard to check the amount of possible bias and consistency, one can check the information gain deriving from the inclusion of multiple and repeated events and the improvement in efficiency. Time can be split into pieces to control the change in rates as time goes by. This approach was shown using the AV access data. The Weibull distribution was found to describe the recurrent failure process well. Parametric survival models offer additional advantages in terms of effect interpretation (time metric) and individual v population hazard (frailty effect).

Organization of the data set to accomplish the analytical task. Risk sets organization is critical to event history analysis. However, even for count data analysis it is always necessary to define the time exposure correctly. Using the AV access data several data set ups were shown for different analytical purposes including models for count data. Usually investigators base their choices on assumptions about when the risk starts, changes and ends. Another example may be useful. If time to central venous catheter infection is measured and repeated episodes can occur in the same subject, it is important to know whether or not a plan is in place to remove the catheter, clear each infection episode and allow some additional time to pass before a new catheter is inserted and the patient returns at risk. This plan may make the assumption of resetting the risk clock

back to zero reasonable because the new catheter insertion date marks the risk onset for a new event and recurrence time is measured as time elapsed from this risk onset date rather than from the study start date. In other circumstances (e.g., no catheter removal and replacement, limited treatment time after each catheter infection episode) the risk clock is never reset, time is measured continuously from entry into the study while recurrences occur over time, and the risk accumulates from entry. It is important to make clear that the way analysis time is measured to set up the risk sets is based on clinical or biological considerations reflecting the dynamic and nature of the failure process rather than statistical assumptions.

In conclusion, different issues need to be considered in the design and analysis of longitudinal studies. These include a thorough evaluation of the biology and mechanisms of the failure event process, careful identification of the possible sources of correlation in the event data, and correct choice of the statistical model that best represent the phenomenon under study. This stepwise approach enhances the likelihood of successful generation and analysis of complex longitudinal data.

8) References

1. Ravani P, Marcelli D, Pecchini P, Malberti F: Early failure rates of arterovenous fistulas for hemodialysis: evaluation of six-year activity, J Vasc Access 2: 154-160, 2001

2. Ravani P, Marcelli D, Malberti F: Vascular Access Surgery Managed by Renal Physicians: the Choice of Native Arterovenous Fistulas for Hemodialysis. Am J Kidney Dis 40: 1264-1276, 2002

3. Ravani P, Brunori G, Mandolfo S, Cancarini G, Imbasciati E, Marcelli D, Malberti F: Cardiovascular comorbidity and late referral impact arteriovenous fistula survival: a prospective multicenter study. J Am Soc Nephrol 15:204-9, 2004

4. Ravani P, Barrett B, Mandolfo S, Brunori G, Cancarini G, Imbasciati E, Malberti F: factors associated with unsuccessful utilization and early failure of the arteriovenous fisula for hemodialysis. J Nephrol 2005 Mar-Apr;18(2):188-96

5. Ravani P, Malberti F, Tripepi G, Testa S, Mallamaci F, Zoccali C: Asymmetrical Di-Methyl-Arginine (ADMA) predicts progression to dialysis and death in patients with Chronic Kidney Disease (CKD): a competing risks modeling approach. J Am Soc Nephrol. 2005 Aug;16(8):2449-55

6. Scolari F, Ravani P, et al: The Challenge of Diagnosing Atheroembolic Renal Disease: Clinical Features and Prognostic Factors, Circulation:17;116(3):298-304

7. Ravani P, Tripepi G, Pecchini P, Mallamaci F, Malberti F, Zoccali C: Urotensin II is an Inverse Predictor of Death and Cardiovascular Events in Chronic Kidney Disease, Kidney Int. 2007 Oct 17

8. Ravani P, Malberti F, Tripepi G, Pecchini P, Mallamaci F, Cutrupi S, Pizzini P,

Zoccali C: Vitamin D Levels and Patient Outcomes in Chronic Kidney Disease: a Longitudinal Cohort Study. Kidney Int, 2009;75(1):88-95

9. Ravani P, Parfrey PS, Curtis B, Barrett BJ: Clinical Research of Kidney Diseases I: Researchable Questions and Valid Answers, Nephrol Dial Transplant. 2007 Sep;22(9):2459-68

10. Ravani P, Parfrey PS, Dicks E, Barrett BJ: Clinical Research of Kidney Diseases II: Problems of Study Design, Nephrol Dial Transplant. 2007 Oct;22(10):2785-94

11. Ravani P, Parfrey PS, Gadag V, Malberti F, Barrett BJ: Clinical Research of Kidney Diseases III: Principles of Regression and Modeling. Nephrol Dial Transplant, 2007;22:3422-30

12. Ravani P, Parfrey PS, Gadag V, Murphy S, Barrett BJ: Clinical Research of Kidney Diseases IV: Standard Regression Models. Nephrol Dial Transplant, 2008;23(2):475-82

13. Ravani P, Parfrey PS, Gadag V, Malberti F, Barrett BJ: Clinical Research of Kidney Diseases V: Extended Analytic Methods. Nephrol Dial Transplant. 2008 May;23(5):1484-92

14. Ravani P, Spergel LM, Asif A, Roy-Chaudhury P, Besarab A: Clinical epidemiology of arteriovenous fistula in 2007. J Nephrol, 2007; 20:141-9

15. Roy-Chaudhury P, Spergel LM, Besarab A, Asif A, Ravani P: Biology of arteriovenous fistula failure. J Nephrol, 2007;20:150-63

16. Lawrence M. Spergel, Pietro Ravani, Prabir Roy-Chaudhury, Arif Asi, Anatole Besarab: Autogenous Arteriovenous Fistula Options, J Nephrol. 2007;20(3):288-98

17. Arif Asif, Pietro Ravani, Prabir Roy-Chaudhury, Lawrence M. Spergel, Anatole Besarab: Vascular Mapping Techniques: Advantages and Disadvantages, J Nephrol. 2007;20(3):299-303

18. Anatole Besarab, Arif Asif, Prabir Roy-Chaudhury, Lawrence M. Spergel, Pietro Ravani: AVF in 2007: Surveillance and Monitoring, J of Nephrology. 2007;20:656-67

19. Asif A, Besarab A, Roy-Chaudhury P, Spergel LM, Ravani P: Interventional nephrology: from episodic to coordinated vascular access care. J Nephrol. 2007;20(4):399-405

20. Spergel LM, Ravani P, Roy-Chaudhury P, Asif A, Besarab A: Surgical salvage of the autogenous arteriovenous fistula (AVF). J Nephrol. 2007;20(4):388-98

21. Anatole Besarab, Pietro Ravani, Lawrence M. Spergel, Prabir Roy-Chaudhury, Arif Asif,: AVF in 2007: Research Needs, J of Nephrology. 2007;20:668-73

22. Lin DY, Wei LJ: The robust inference for the Cox proportional hazards model. Journal of the American Statistical Association 84:1074-1078, 1989

23. Therneau TM, Grambisch PM: Modeling Survival Data: Extending the Cox Model. Springer-Verlag: New York, 2000

24. Huang X, Wolfe RA: A frailty model for informative censoring. Biometrics 58:510-520, 2002

25. Box-Steffensmeier JM, De Boef S: Repeated events survival models: the conditional frailty model. Stat Med Dec [Epub ahead of print], 2005

26. Andersen PK, Gill RD: Cox's regression model for counting processes: A large

sample study. Annals of Statistics 10: 1100–1120, 1982

27. Wei LJ, Lin DY, Weissfeld L: Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association 84: 1065–1073, 1989

28. Lunn M, McNeil D: Applying Cox regression to competing risks. Biometrics 51: 524–532, 1995

29. Prentice RL, Williams BJ, Peterson AV: On the regression analysis of multivariate failure time data. Biometrika 68: 373–379, 1981

30. Cleves MA, Gould WW, Gutierrez RG: Generalizing the Parametric Regression model, in "An Introduction to Survival Analysis using STATA", pp 269-298, STATA press publication, College Station Texas, first revised edition, 2004

9) Further reading

Glantz SA, Slinker BK: A Primer of Applied Regression and Analysis of Variance. Mc Grow Hill, 2001 (second edition)

Kleinbaum DG, Kupper LL, Muller KE, Nizam A: Applied Regression Analysis and Multivariable Methods. Duxbury Press 1997

Dupont WD: A simple introduction to the analysis of complex data. Cambridge University Press 2002

Hosmer DW, Lemeshow S: Applied Logistic Regression. John Wiley and Sons, 2000

Cox DR: Regression Models and Life-Tables. Journal of the Royal Statistical Society, Series B, 1972; 34: 187–220

Hosmer DW, Lemeshow LS: Applied Survival Analysis, Regression Modelling of Time to Event Data, John Wiley & Sons, New York: 1999

Kleinbaum DG: Survival Analysis, a Self-Learning Text, Springer-Verlag, New York: 2005

West B, Welch K, Galecki AT: Linear Mixed Models: A Practical Guide using Statistical Software. Chapman & Hall/CRC, New York: 2007

Therneau TM, Grambisch PM: Extending the Cox Model. Springer-Verlag, New York: 2000