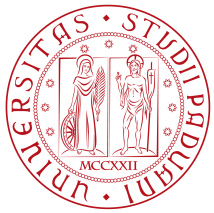
Università degli Studi di Padova

Corso di Laurea in Scienze Statistiche



JOINT MODELS FOR LONGITUDINAL AND TIME-TO-EVENT DATA. A STUDY OF THE INFLUENCE OF THE ASSOCIATION STRUCTURE ON EXPECTED SURVIVAL PROBABILITIES.

Relatore: Prof.ssa Fausta Ongaro
Dipartimento di Scienze Statistiche

Co-relatore: Prof. Dimitris Rizopoulos
Dipartimento di Biostatistica
Erasmus University Medical Center

Laureanda: Illary Sbizzera

Anno Accademico 2011/2012

Contents

Introduction

The two primary means by which alternative medical or surgical treatments are assessed is through the use of randomized controlled trials (RCTs) or observational studies (OS) (Hannan 2008). In RCTs, participants are randomly assigned to a treatment or control group (or to multiple treatment groups) so as to reduce bias by making the groups as equal as possible with respect to all patient characteristics that may have an impact on outcomes. Thus, in theory, the only difference between the groups is the treatment assignment and any differences that are identified. In contrast, OS do not randomize treatment but 'observe' differences in outcomes that occur after treatment decisions have been made, without regard to ensuring that patients in different treatment arms have similar characteristic related to outcomes. Treatments may be diagnostic, preventive or therapeutic and may include drugs, biologics, medical advices or methods of screenings. Treatments may also include procedures whose aim is to improve quality of life or to better understand how the intervention works in partecipants (Peace and Chen 2011).

Clinical trials and observational studies usually generate both longitudinal measurement data, with repeated measurements of one or more response variables at a number of time points for each participant, and event history data, in which times to recurrence or terminating events are recorded (Henderson et al. 2000). A widely used example is in AIDS research where a biomarker called CD4 lymphocyte count is measured intermittently and both its progression and relationship with time to seroconversion or death

is of interest (e.g. DeGruttola et alt., 1993; Tsiatis et al., 1995; Wulfsohn and Tsiatis, 1997). Another example comes from studies following men who have been treated with definitive local therapy (i.e. radical prostatectomy or radiation therapy) for localized prostate cancer. Despite the shift to earlier treatment due to a recent widespread use of early detection programs, a substantial percentage of these patients shows evidence of biochemical recurrence within 10 years after treatment (Taylor et al., 2005). Prostate-specific antigen (PSA) is the covariate commonly used to monitor patients after they have been treated. In fact, several studies have shown that a rise of post-treatment PSA is highly predictive of clinical recurrence (Ali et al. 2006).

Thus, at the end of studies similar to those described above, we often have the following situations (Wu 2010):

- "the main focus is on modeling the survival data, with modeling longitudinal data being secondary". For example, our attention is direct to the event outcome and we wish to account for the effect of the longitudinal outcome as a time-dependent covariate.

- "the main focus is on modeling longitudinal data, with modeling survival being secondary". We could be in this situation when we analyze longitudinal data where dropouts are informative so that the survival data could be only used to account for the informative dropouts.

- "the main focus is on modeling both the longitudinal data and the survival data, with goal of understanding the association between the two processes". This situation arises, for instance, when we wish to characterize the relationship between features of a time-dependent covariate's trajectory and the survival endpoint in order to assess the prognostic value of the longitudinal covariate and, in presence of treatment, to verify the possibility to use it as a 'surrogate marker' (Fitzmaurice et al. 2008).

In all three situations, traditional approaches used to analyze survival or longitudinal data are not applicable without incurring in biased results. Taking

in consideration the first situation, several problems are present. First, standard time-to-event models (e.g. proportional hazard models (Cox 1972)) require that time-dependent covariates are external: although the value of this covariate at time point u "influence the rate of failures over time, its future path up to time $t > u$ is not affected by the occurrence of a failure at time u" (Kalbfleisch and Prentice 2002, Section 6.3). However, most of the time-dependent variables observed in longitudinal studies are internal so that they do not satisfy this condition. Internal variables, in fact, are usually the output of a stochastic process generated by the subject under study. Consequently, they are only observed as long as the subject is alive and uncensored and therefore they implicitly carry information about the failure time. Second, to apply Cox's methods for the estimation of the model parameters, it is necessary to have complete knowledge of the covariate history for all individuals while on study (Tsiatis 1995). In most studies, however, the time-dependent covariate is collected only intermittently on each individual during medical examinations whose frequency can change from once a day to once a year according to the study protocol. Third, to optimally estimate the parameter models, we would need to know the covariate value without measurement error (Wulfsohn and Tsiatis 1997). Prentice (1982) established that the presence of the measurement error (i.e. a random error) in a measured covariate causes the estimated parameters to be biased towards the null. Also Tsiatis (1995) stressed that the maximum explanatory power of the time-dependent covariate on the hazard can only be achieved after adjustment of measurement error. Using the observed value, the possible relationship between the covariate and the time-to-event or between the covariate and the treatment effect could not be detected. In the second situation there is the need to deal with informative dropout. Dropout happens when, during the follow-up, a subject 'drops out' or is withdraw from the study before the study is completed. The dropout mechanism is said to be informative when the probability of dropout is related to specific values that should have been obtained, have the subject not left the study. If an

analysis is conducted without taking informative dropout into account, its results could be potentially biased. Obviously, the third situation presents all those problems together.

In order to avoid all these problems and obtain valid inferences, joint models for longitudinal and survival data have been proposed. Wulfsohn and Tsiatis (1997) develop a methodology whereby a joint maximization of the likelihood from both the covariate process and the survival data occurs. In particular, their method uses not only the observed covariate data but also survival information to estimate the true covariate value at any time. Therefore the authors expects more precise and accurate estimates of the strength of the relationship between the covariate risk and the risk of failure than those obtained from previous methods such as the 'Last value of Last Observation Carried Forward' or the 'Two-Step' approach.

In the last decade, joint models have been mainly applied in medical studies to correctly estimate the association between a time-dependent covariate, usually a biomarker, and baseline covariates with the event outcome, verifying candidate risk factors and testing treatments' effect in prolonging survival. A biomarker indicates a change in expression or state of a biological measurement (e.g. concentration of a protein or an antigen in serum or other tissues, aneurysm diameter, blood pressure, ...) at a given time point. A biomarker, if validated, could be used as a surrogate for subsequent survival. Buyse et al. (2000) distinguish two different types of surrogacy: trial-level and individual-level surrogacy. The trial-level surrogacy coincides with the surrogacy also described by Prentice (1989). He defines a surrogate endpoint to be "a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint". This means that a biomarker that is influenced by the tested treatment if and only if the treatment has a significant effect on survival is a good surrogate endpoint. On the other hand, at the individual level, a biomarker would be a useful surrogate endpoint if the trajectory of irregularly observed values available

at any time for a single subject provides helpful prognostic information on subsequent survival of that subject (Henderson et al. 2000).

Indeed, one aspect of joint models that has recently gained some increasing interest is to obtain subject-specific predictions for either the longitudinal or survival outcomes (Sweeting and Thompson 2011, Yu and Taylor 2008, Proust-lima and Taylor 2009). The ability of incorporating the whole biomarker's trajectory over time gives to joint models the possibility to produce dynamic prognostic tools that could more accurately guide clinical decision making. For example, the complete post-treatment PSA pattern of a patient can be used to predict the probability of biochemical recurrence within two years from the last visit time. If the PSA pattern is suggestive of an increase risk of clinical recurrence, the physician may decide to put the patient on hormone therapy to slow progression of the disease or to perform a prostatectomy.

The aim of this thesis is to compare the predicted individual survival probabilities estimated by models where different types of association between the longitudinal biomarker and the survival outcome are considered. In fact, although it is often assumed that the risk for an event at a particular time point t depends on the true level of the longitudinal marker at the same time point, it is not realistic to expect that this parameterization will always be the most appropriate in expressing the correct relationship between the two processes. This is in line with the more challenging nature of time-dependent covariates with respect to baseline covariates. Since the true functional form of a time-dependent covariate is often not self-evident and that the choice of how to model it can considerably influence the derived results, the researcher should not always rely on the standard formulation but rather prudently face the problem and investigate different parameterizations and possibly their combinations. In addition to what just said, Henderson (2002) also found that different models, even if they fit average characteristics equally well, may give quite different predictions for individual patients.

In Chapter 1 we will underline the importance of prognosis in everyday

medical practice and the necessity of having accurate survival predictions in order to properly program patients' therapy. Moreover we will show how some selected biomarkers could be used to increase the precision of prognosis.

In Chapter 2 we will describe the traditional approach used to analysis survival data where time-dependent covariates are involved and we will point out the problems arising from such approach. Thus, in Chapter 3, we will present the joint model approach which solves, or at least greatly reduce, those problems.

In Chapter 4 we will present how survival predictions can be computed within the joint modeling framework and we will indicate the different association structures which could be used as an alternative to the usual one.

In Chapter 5 we will show how the use of different association structures can substantially influence survival predictions. In our analysis we used a dataset based on a clinical trial involving subjects affected by primary biliary cirrhosis (PBC).

Chapter 1

Prognostic models in medicine

1.1 The meaning of prognosis

The word prognosis comes from the Greek $\pi\rho\acute{o}\gamma\nu\omega\sigma\iota\varsigma$. It is composed by the prefix $\pi\rho o$- ("before") + $\gamma\nu\tilde{\omega}\sigma\iota\varsigma$ (gnosis, "inquiry, investigation, knowing"). Thus literary it means "to know beforehand" or, as a noun, foreknowledge. Prognosis is the prediction of what is judged likely to happen in the future, especially in connection with a particular situation and relative to the information available at the time of prediction.

A field where prognostication is a daily practice is medicine. According to Abu-Hanna and Lucas (2001), medical prognosis is defined as "the prediction of the future course and outcome of disease processes, which may either concern their natural course or their outcome after treatment". Thus a prognosis is made every time a patient is diagnosed with a disease and its importance increases whether the disease can lead to a significant decrement in the patient's quality of life or to death. Indeed, physicians specialized in cardiology, neurology, intensive care medicine and oncology, from newly diagnosed to terminally ill patients, are especially required to be able to prognosticate and this does not only imply making survival predictions. Physicians are required to be able to indicate disease evolution, possible disabilities that the patient would be facing, response and side effects associated with different treatment

options and, eventually, costs of health care (Glare et al. 2008). Prognosis is interlaced with other tasks of clinical management of the patient such as diagnosis, therapy selection and planning. Information obtained at diagnosis and knowledge about feasible medical actions, with therapy in the first place, have a strong influence on prognosis. Therefore, prognosis could be considered a comprehensive decision tool that, according to all the information available at the time of diagnosis, allows clinicians to choose the most appropriate way to tackle the specific disease.

From now on, the aspect of prognosis that will be considered is the prediction of survival time, in the sense of time elapsed before disease recurrence or death.

Since prognosis has such a deep influence on patients' future, it is universally recognized that prognosis has to be as much accurate as possible. For cancer patients, for example, Mackillop and Quirt (1997) individuated three reasons why accurate predictions are required. These reasons can easily be generalized to other types of disease. First, physician's prognosis has an influence in the choice of treatment. The probability of cure associated with each potential treatment and the patient's expected survival have to be carefully assessed since many cancer treatments (e.g. biological therapy, radiation therapy, chemiotherapy, ...) have important side-effects which may initially decrease the quality of life and may only be considered acceptable if the patient is likely to live long enough to experience any subsequent treatment's benefit. Patients with an incurable cancer or at the latest stages should be advised on palliative care programs. Second, accurate prognostic judgment contributes to the efficient use of health care resources since treating patients with expensive therapies that will not give any benefit not only submits the patient to useless toxicity, but also waste important resources that could be spent elsewhere. Third, good predictions may help patients and their families to make appropriate plans for the remaining time in order to take advantage of every day.

Also national and local Health Authorities are particularly interested in

the fact that survival predictions are as much accurate as possible since patients may qualify for specific financial benefits if their life expectancy is below a specified threshold. For instance, Henderson et al. (2001) refer that "in U.K., a patient can claim additional financial support without the usual waiting time if a doctor certifies that the patient has 'progressive disease' and it is not expected to live longer than six months". In the U.S.A., the Medicare insurance programme[1] introduced hospice care for its beneficiaries. In order to be eligible for Medicare hospice benefits, the patient's doctor and the hospice medical director have to certify that the patient is terminally ill and probably have less than six months to live (Medicare). These two examples are enough to further underline the importance of giving accurate prognosis, both for patients' interest and health programs' efficiency.

1.2 Accuracy in survival predictions

In most of the cases, patients diagnosed with life-threatening diseases want a high level prognostic information. A study involving 126 patients with incurable metastatic cancer found that 95% of them wanted information about side effects, symptoms, and treatment options and the 98% indicated that the doctor should be realistic, provide an opportunity to ask questions, and acknowledge the patient as an individual when discussing prognosis (Hagerty et al. 2005). Although patients clearly want prognostic information, doctors feel this task stressful and difficult. Christakis and Iwashyna (1998) in a study involving 697 American internists investigating attitude and practice about prognostication found that doctors "believed that patients expect too much certainty and that both patients and (to a lesser extent) colleagues will judge them adversely for prognostic errors". Those doctors believed also that

[1]Medicare is a social insurance program administered by the United States government (CMS), providing health insurance coverage to people who are aged 65 and over; to those who are under 65 and are permanently physically disabled or who have a congenital physical disability; or to those who meet other special criteria such as terminally ill patients.

"they should accentuate the positive in making predictions and avoid being too specific" and that it is better not to volunteer prognostic assessments. One of the possible reasons that make clinicians so uncomfortable when they are asked for a prognosis is the intrinsic uncertainty that characterizes the course of the disease. The probability of error will never be completely removed when predicting future outcomes, especially when considering the complex dynamic of the human body and the multiple interactions between the human body and illness (Glarea et al. 2008). Nevertheless, the impossibility to formulate 100% accurate prognosis does not have to stop the research of tools that could improve clinicians' prognostications.

The two approaches used in order to formulate the prognosis are: clinical prediction of survival (CPS) and use of statistical tools such as prognostic models. CPS is a procedure in which the judge puts clinical data together using informal, subjective methods. Thus, CPS is essentially based upon physicians' experience. Even though, as it has been said before, prognosis, and hence CPS, does have a decisive impact on many aspects, the studies assessing accuracy in physicians' survival predictions are few and most of them are based on data from terminally ill patients. It is not easy to compare the results from these studies since there is not a worldwide definition of prediction accuracy. Nevertheless, in most of the cases, the conclusions are the same. Probably, the very first study about this subject was published by Parkes in 1972. The research was conducted in a hospice in Sydenham (close to London) where the author asked general practitioners or hospital medical staff to state the expectation of life in weeks for all those patients with a diagnosis of cancer who were admitted to the hospice during 1970-1. At the end of the study, there were 293 predictions of survival made on 168 cancer patients. In order to analyse the data, Parkes used his own definition of accuracy: prediction is in 'serious error' if it differs from the actual survival (AS) by a multiplicative factor of two, that is, if AS>2CPS or AS<0.5CPS. Using this criterion, overall the rate of error was 53% with a 83% of them in an optimistic direction. More recently, Christakis and Lamont (2000) con-

ducted a large, prospective cohort study of terminally ill patients to evaluate the extent and determinants of prognostic error. In the end, the study had information from 343 physicians caring for 468 patients and both CPS and AS were made in days. In order to verify prognostic error, the authors "divided the observed by the predicted survival and deemed prognoses accurate if this quotient was between 0.67 and 1.33". Hence, CPSs' values with quotients between 0.67 and 1.33 times the AS were accurate, values less than 0.67 were optimistic prognostic errors, and values greater than 1.33 were pessimistic. The authors found that 92 (19.7%) of 468 predictions were accurate, 295 (63.0%) were optimistic, and 81 (17.3%) were pessimistic. Whereas using Parker's definition of prognostic error, 159 (34.0%) of 468 predictions were accurate, 256 (54.7%) were optimistic, and 53 (11.3%) were pessimistic. Moreover, the authors reported that "physicians in the upper quartile of practice experience were the most accurate" and that "as the duration of the doctor-patient relationship increased and time since last contact decreased, prognostic accuracy decreased".

Given the importance of accurate prediction and the evidence that CPS is a not so precise tool in this sense, an interesting question arises as to whether the use of objective methods based on statistical models can replace or inform subjective clinical judgment. Survival analysis is a widespread tool used to determine covariate effects, to compare different groups (e.g. treatment groups), and to form prognostic indices. However, as Henderson et al. (2000) states, these models are not often used to make individual point or interval lifetime predictions. They claim that a possible answer "is that statistical methods have not convincingly been demonstrated to lead to accurate and thus helpful predictions". Indeed their study showed that the point predictions from four different survival models were generally quite similar. Moreover, for all models and all individuals the probability of being in 'serious error' using Parker's criterion was around 50%.

The studies presented so far were all based on point predictions. Therefore, it may seem appropriate to present some type of reliability measure

as well. One option is represented by predictive intervals (Henderson and Keideng, 2005). Predictive intervals can be obtained from survival curves, and they give to each patient a range of outcomes within which AS will lie with a specified probability. Nevertheless, the authors correctly recall that, although interval estimates accurately quantify the uncertainty in prognosis, intervals are often as wide as to be of little practical use.

In conclusion, multivariate regression models, such as survival models, have been proven to be good tools in individuating risk factors and in comparing groups of patients. In particular, they are often used in distinguishing patients with a high risk from patients with a moderate or low risk. Moreover, even though physician's prognostic estimates are not accurate enough to be reliable at the patient level, they still provide valuable prognostic information that can be used to improve statistical models. Muers et al. (1996), in a study of advanced non-small-lung cancer patients, found that the Cox's proportional hazard model incorporating, among prognostic factors, the physician's point prediction of survival, was the best model discriminating between poor and moderate/good prognosis groups. This improvement made the authors suggest that physicians not necessarily use the same information included in the model to predict survival and thus they might be using additional elements on which to base their prognostic judgement. Notwithstanding the great usefulness these models have at the population level (e.g. knowing the modifiable risk factors linked to cardiovascular pathologies such as hypertension, diabetes, obesity and smoke, governments can promote health programs with the aim of reducing the exposition of those risk factors and thus the incidence of such diseases), their prediction ability at the patient level is, in most of the cases, useless. The main point that Henderson and Keideng (2005) wanted to underline with their work was that "in all realistic scenarios we can imagine, the intrinsic statistical variations in life times are so large that predictions based on statistical models and indices are of little use for individual patients. This applies even when the prognostic model is known to be true and there is no statistical uncertainty in parameter estimation".

The fact that both physicians and prognostic models are not able to give accurate individual point prognosis, certainly will not stop patients from asking "How long do I have, doctor?". If clinicians do not want to base their answer only on their subjective judgment, the only alternative is to rely on epidemiological results. Nonetheless, as indicated by Scuchter (1996) and Hollnagel (1999), the physician must carefully distinguish prediction of the outcome for a population of patients from that for a singular patient. In particular, Hollnagel highlights two important points which any clinician should be aware of. "The first is that epidemiological research can only incorporate factors which can be measured quantitatively". For example, patients' important features such as life experience, lifestyle, physical capabilities, hopes and fears about the future and, even more, attitude matter so much (Gould, 1986). But, as these factors are not usually measured and not easily amenable to epidemiological analyses, they are largely ignored and thus results do not account for them. "Second, epidemiological knowledge about risk factors is group-based as opposed to individualised knowledge". This is a cornerstone issue that has to be tackled by every physician when he/she is in the consulting room with a patient. The physician could be able to assign the patient to the correct risk group and thus inform him/her about the median survival time, treatment efficacy and so on. Nevertheless, since the variability between patients belonging to the same risk group is often not null, the possibility that this patient would behave in a different way should not be ignored. Mackillop (2006), in his paper *The importance of prognosis in cancer medicine* states that, in the future, "the challenge will be to increase the 'particularizability' of medical knowledge in such a way that the individual characteristics of the patient and the tumor are appropriately factored into treatment decisions. This will require characterizing patients, not only in terms of the diagnostic group to which they belong, but also in terms of all those individual characteristics that may influence the outcome of treatment". This sentence is referred to prognosis in oncology but it is undoubtedly valid in other medical branches where accurate

individualised prognosis are fundamental in supporting both physicians' and patients' decision making.

Returning to the point prediction issue, what has emerged so far is optimally summarized by Schumacher et al. (2003). Examining the difficulties with the use of point predictions, the authors pointed out two key features. First, the duration of survival or time to the event of interest itself cannot be predicted adequately. Second, there seems to be no widely agreed statistical methodology to assess the accuracy of predictions derived from expert opinion or from a survival model. Consequentely, the authors suggest to abandon predictions on the time axis and to consider only predictions on the probability axis. Although Henderson et al. (2000) argues that the time axis is the most natural measure, statements like 'the chance of surviving 5 years is 70%' seem to be quite understandable and reveal the uncertainty about the exact moment of death. Second, the commonly used survival models allow to obtain individual survival probability curves which can be used as predictions. Such curves could provide confidence intervals to account for uncertainty as well. Moreover, physicians could find easier to make predictions on the probability axis and give more accurate judgments. Indeed, Weeks et al. (1998) found that physicians' estimates of 6-year survival of cancer patients was not so poor. For example, among patients with more than 90% probability of surviving 6 months, 71% (41 patients over 58) actually survived, while among those with a less than 10% probability, only 11% (17 patients over 158) survived.

Albeit probabilistic predictions have been found to be more accurate than temporal predictions, it has to be noted that they are not immune from uncertainty (e.g. confidence intervals of individual survival curves can be so wide to make probabilistic predictions pointless). Traditionally, statistical survival models used prognostic factors such as performance status, symptoms and simple laboratory results. Even though they are often highly statistically significant prognostic factors, they might explain a small fraction of the variation between individuals. Thus, in the recent years, the research's attention

has been directed to the discovery of new prognostic factors, strongly linked to disease progression and/or outcome, easily measurable with non-invasive techniques, and that may be of help to clinicians for more individualised prognosis. Easily measurable prognostic factors would also respond to the necessity of having constantly updated prognosis. Glare et al. (2008) specify that "prognosis is often misunderstood as a static phenomenon, reinforced by the research studies focusing on one point in time (e.g. survival after admission to hospital or referral to hospice, survival after the surgery, etc). The illness trajectory changes over time, so that, as the illness evolves, new issues must be considered and the prognosis should be revised".

1.3 Biomarkers

From Section 1.2 it can be inferred that there is an increasing interest in variables that could be used alone, or in more complex statistical models, in order to provide objective information to be used by clinicians in deciding which treatment would be better for the individual patient. Indeed, decision making would be a lot easier if specific measurements of the patients' clinical status could predict the risk of progression to death. This necessity made research to focus its attention on biomarkers. In 2001, a working group, belonging to the National Institute of Health, gave this standardized definition of biomarker (NIH Definitions Working Group, 2001): "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention". According to this definition, a biomarker may be measured on a biosample (e.g. blood, urine, sample of tissue), it may be measured directly from a person (e.g. blood pressure, electrocardiogram, spirometry), or it may be an imaging test (computed tomography, ecography, magnetic resonance spectroscopy) (Vasan 2006).

Biomarkers can be involved in several clinical activities whose aim is to assess patients' health or disease characteristics. According to their applica-

tion, biomarkers can be classified as follow (Vasan 2006):

- *antecedent biomarkers*: they identify the risk of developing an illness; usually they are genetic markers that can indicate the genetic predisposition of an individual to develop a specific disease. An example comes from the Huntington's disease, "a neurodegenerative genetic disorder that affects muscle coordination and leads to cognitive decline and dementia. The disease is caused by an autosomal dominant mutation in either of an individual's two copies of a gene called Huntington, which means any child of an affected parent has a 50% risk of inheriting the disease" (Wikipedia 2012). Since genetic test can be performed even before the symptoms' onset, a person can know in advance if he/she will develop the disease in the future;

- *screening biomarkers*: they are biomarkers that indicate the presence of a disease in asymptomatic people. For instance, the prostate-specific antigen is a biomarker for prostate cancer used in prostate cancer screening in an attempt to identify individuals at the early stages of the disease;

- *diagnostic markers*: they are used to confirm the presence of the disease among symptomatic people. An example is given by the celiac disease. It can be diagnosed in individuals having sign or symptoms of malabsorption or malnutrition. However, since other diseases manifest the same symptoms, it is important to confirm the disease with specific tests: small intestinal biopsy and antibody tests are suggested;

- *staging biomarkers*: they are used to characterized disease severity;

- *prognostic biomarkers*: they are useful in predicting disease progression, including recurrence and response to therapy. They are also used in monitoring efficacy of therapy. For example, PSA is used to monitor biological recurrence in prostate cancer patients after radiation therapy or radical prostatectomy while serum bilirubin level is used to monitor

disease progression in patients affected by primary biliary cirrhosis. An example of biomarkers used to predict treatment efficacy are genetic mutations that cause drug resistance in cancer patients. From the analysis of cancer tissue, genetic mutations, that cause specific drugs to be ineffective, can be found. Consequently, the physician will have to choose other drugs, if possible, in order to fight the cancer.

It is important to notice that the same biomarker can be used with different purposes. PSA gives a clear example: it can be a screening, diagnostic or prognostic biomarker.

Biomarkers may also serve as surrogate endpoints. A surrogate endpoint is a biomarker that is used instead of the clinical endpoint (e.g. death, biochemical recurrence, stroke or other specific events of interest) to evaluate safety and effectiveness of new treatments in clinical trials. In a seminal work, Prentice (1989) gave the guidelines for studying surrogate endpoints and a formal definition of conditions that the biomarker should satisfy to be used as a valid surrogate endpoint in a specific trial. More specifically, the three conditions where the following (Tsiatis 1995, Taylor 2002):

1. The biomarker should be associated to the clinical endpoint;

2. The treatment should have an effect on the biomarker;

3. The effects of the treatment should be mediated though its effect on the marker. That is, patients with the same value of the biomarker should have the same survival probability, independently of the treatment they are receiving.

The great attention shown in surrogate endpoints is due to several features which they usually possess. First, candidate surrogate points are cheaper and easier to measure than clinical endpoint. As Aronson et al. (2005) state "it is easier to measure a patient's blood pressure than to use echocardiography to measure left ventricular function, and it is much easier to do echocardiography than to measure morbidity and mortality from hypertension in the

long term". Second, they can be measured more quickly and earlier. In fact, blood pressure can be measured every day, whereas it usually takes several years to collect mortality data. Third, surrogate endpoints need smaller sample sizes and can avoid ethical problems associated with measuring clinical endpoints (Aronson et al. 2005).

However, surrogate endpoints come with disadvantages as well. In practice, it is rare to find biomarkers satisfying all Prentice's three conditions. The main reason is that it is difficult for a single biomarker to completely account for all treatment effects and often it is not reasonable to think that the clinical outcome of interest is influenced by that biomarker only. That is why, recently, some studies have focused their attention on multiple biomarkers that could capture various components of complex disease evolutions thus giving a more comprehensive assessment of treatment effect (NHI 2001). For example, a recent study tested the use of multiple biomarkers to improve the prediction of death from cardiovascular disease (Zethelius et al. 2008). The authors concluded that "the incorporation of a combination of biomarkers that reflect myocardial cell damage, ventricular function, renal function, and inflammation to a model with established risk factors improved the risk stratification for death from cardiovascular causes". Another study reviewed by many researches focused on the simultaneous examination of multiple markers in order to increase the sensitivity of the screening test for early detection of ovarian cancer (Yurkovetsky et al. 2006). The authors concluded that "a multimarker approach for the generation of a prototype assay for early detection of ovarian cancer has a great potential to lead to the development of a screening test for this disease".

Regardless of the purpose of its use, a new biomarker will be of clinical value only if it is strongly associated with the outcome of interest, it can easily be obtained (i.e. with no harm for the patient), it has a clear interpretation for the clinicians, it explains, despite the presence of already established risk factors, a reasonable proportion of the outcome variability when tested in several studies, and least, but not last, the information based on its knowledge

have a direct impact in patient's management.

1.4 The use of biomarkers as predictors

In this section, prostate-specific antigen (PSA) will be used to show its potentiality for individualised predictions of disease progression following therapy (i.e. radical prostatectomy or radiation therapy) for localized prostate cancer. Then, the statistical techniques used for the analysis of similar biomarkers will be mentioned. A more exhaustive description will be given in Chapter 3.

Patients treated for prostate cancer are usually followed for several years in order to monitor their response to the treatment. Every patient is asked to do medical check-ups every few months and at each visit a blood test is performed in order to measure PSA levels. In fact, usually, after prostate removal surgery, PSA level in the blood decreases and it eventually becomes almost undetectable. After radiation therapy, PSA levels usually drop to a stable and low level (WebMD). The correlation between changes in PSA with time and prostate biochemical recurrence (i.e. local recurrence or distant metastases) has been long recognised (Oesterling 1991). However, a widely accepted definition of biochemical recurrence based on a series of PSA measurements has been a problematic and controversial topic. For instance, there is considerable variability in the amount by which PSA is reduced by radiation therapy, and the time interval before it starts to rise, if it does rise, and the rate at which it rises are very diverse (Zagars and Pollack 1993). Carter and Pearson in 1993 suggested to use PSA velocity, and, in particular, PSA doubling time, to identify men with prostate cancer that is destined to progress. Pollack et al. (1994) conducted a study on patients showing a rising PSA profile after radiation therapy and they confirmed PSA doubling time (PSADT) as a strong prognostic factor for patient with biological recurrence. They also found that "the timing of the progression from a rising PSA to clinical disease relapse [...] is estimated to be 40 months on average".

This means that when a rapid increase of PSA levels is detected, there is still time for the physician to choose a proper treatment.

Since detecting early signs of a recurrence is of major importance for patient's care and may guide the physician's decision to start further therapies, such as salvage androgen deprivation therapy, it is important to be able to use all the available information from PSA monitoring after treatment. In fact, it is believed that methods based on the pathway of prognostic biomarkers such as PSA could enhance their prognostic ability, hopefully giving more accurate survival probabilities at the individual level. Moreover, those models should allow to update their survival predictions after each new measurement of the biomarker reflecting, in this way, the dynamic nature of disease.

The methods that were able to answer all the needs described above, and solve all the problems arising from dealing with internal time-dependent covariates characterized by a great variability, are known as joint models.

Chapter 2

Handling time-dependent covariates

2.1 Introduction

Longitudinal studies such as clinical trials or observational studies often record two different kinds of data from each individual: a longitudinal sequence of repeated measurements at pre-specified measurement times and one or more time-to event outcomes such as death, development of a disease, clinical recurrence or dropout from the study. Two typical examples of this setting are HIV and cancer studies. In HIV studies, measures of immunological and virological status, such as CD4 T-cell count and viral RNA copy number (also known as viral load) are gathered longitudinally on each individual along with time to progression to AIDS or death. Likewise, in cancer studies, the event of interest is death or cancer recurrence and patients provide longitudinal measurements of antibody levels or of other biomarkers of carcinogenesis, such as PSA levels for prostate cancer, during follow-up visits. Obviously, also several time-independent covariates, which are often referred to as baseline covariates, are recorded, usually at the beginning of the study.

With these types of data available, a variety of questions may be of in-

terest, depending on the application. According to the research's interest, three types of setting can be defined:

1. The interest could be on the event outcome and we wish to account for the effect of the longitudinal outcome as a time-dependent covariate. In Chapter 1 it has been shown how PSA is a widely used biomarker to predict biochemical recurrence in patients treated for prostate cancer. Similarly, serum bilirubin is used to monitor primary biliary cirrhosis progression.

2. The interest may be on the longitudinal outcome but this is complicated by potentially informative dropout, which may be viewed as a time-to-event outcome. For example, in a study where the goal is to compare different treatments, participants may dropout from the study since they feel much better or, on the contrary, they do not see any improvement from the therapy. Thus, dropout is nonrandom, but linked to treatment effect;

3. The interest may be on the relationship between the longitudinal and the time-to-event outcome so that both types of outcome are of primarily interest. This setting often arise when the association between the biomarker trajectory and the time-to-event outcome needs to be assessed. For example, the goal may be to verify if hemoglobin level can predict renal graft failure.

The first setting is the one that will be considered in this thesis from now on. More specifically, our interest will be in predicting survival probabilities using models which exploit all the information available from longitudinal measurements of a specific biomarker. In Section 2.2, the traditional approach for the analysis of survival data will be presented along with the problems arising from the necessity to incorporate the information from a time-dependent covariate,i.e. the longitudinal response, measured intermittently and possibly with error. In Section 2.3, the three methods developed to solve these problems are presented.

2.2 The traditional approach for analyse time-to-event data

The traditionally used survival model for the analysis of time-to-event data is the relative risk (Cox) model (Cox 1972). As other survival regression models, it is used to find any possible relationship between survival times and important covariates. The main reason why standard regression models (e.g. linear regression model, logistic regression model, etc.) will often be inappropriate is that survival times may be censored. Censoring occurs when incomplete information is available about the survival time of some individuals. There are three types of cesoring: left censoring, interval censoring, and right censoring. When the event of interest already occurred before the beginning of the study, the event time is said to be left censored. When the event occurred in a time interval but we do not know the exact time point, the event time is interval censored. Finally, when the event did not occur while the subject was under study, the event time is right censored. This last type of censoring happens, for example, because the medical study ends before each patient experiences the event, or because the patient drops out of the study prematurely. In both cases, we only know that the patient will possibly experience the event in the future but we do not know exactly when. Since right censoring is the most common type of censoring in clinical trials and observational studies, we assume survival data to be subjected to right censoring.

Let T_i be the time to an event of interest, called survival time, for the i-th subject ($i = 1, 2, ..., n$). T_i is taken as the minimum of the true event time T_i^* and the censoring time C_i. Therefore, we observe $T_i = \min(T_i^*, C_i)$. Furthermore, we define the event indicator as $\delta_i = I(T_i^* \leq C_i)$, where $I(\cdot)$ is the indicator function that takes the value 1 if the event is observed (i.e. $T_i^* \leq C_i$), and 0 otherwise. Moreover, let z_i be the vector of baseline (time = 0) covariates (e.g. gender, height, ethnicity, treatment indicator, etc.), measured before or at the beginning of the study. For the longitudinal outcome,

we consider only the case of a single covariate which is measured repeatedly over time. Let $y_i(t)$ denote the value of the longitudinal outcome at the time point t for the i-th subject. It is important to emphasize that we do not actually observe $y_i(t)$ at all time points, but only intermittently, at specific time points t_{ij} at which the measurements were taken. Moreover, although visit times usually follow a time schedule, some participants could not follow it perfectly so that the times at which the longitudinal covariate is collected and the final number of covariate values can be different for each individual. Therefore, the observed longitudinal data consist of the measurements $y_i = \{y_i(t_{ij}), j = 1, 2, ..., n_i\}$, where n_i is the number of measurements available for subject i. Overall, the observed data available for each individual is $(T_i, \delta_i, z_i, y_i)$. We assume that censoring is non-informative in that, given z_i and y_i, the event and censoring time for the i-th patient are independent.

In survival regression models, the common approach is to model the hazard function rather than the means functions as in classical regression models. The hazard function, also called risk function, is defined as (Wu 2010)

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T^* < t + \Delta t \mid T^* \geq t)}{\Delta t}, \qquad t > 0,$$

which is the risk or hazard of death (or event) at time t, i.e., the probability that an individual experiences the event at time t given that he/she has survived to time t. Another essential function in survival analysis is the survival function, which is defined as

$$S(t) = P(T^* \geq t) = 1 - F(t), \qquad t > 0,$$

where $F(t) = P(T^* < t)$ is the usual cumulative distribution function (cdf). Survival at time t, i.e. S(t), is the probability that an individual survives at least to time t. This two functions are strongly linked as it can be seen from the following formula:

$$h(t) = -\frac{d}{dt}\log(S(t)), \qquad t > 0,$$

The probability density function can be defined as

$$f(t) = h(t)S(t), \qquad t > 0,$$

Therefore, the probability distribution of the observed data (T_i, δ_i) is given by

$$f(T_i, \delta_i) \propto f(T_i)^{\delta_i} S(T_i)^{1-\delta_i} = h_i(T_i)^{\delta_i} S_i(T_i), \qquad T_i > 0,$$

In order to quantify the effect of $y_i(t)$ and z_i on the risk for an event, we may use a relative risk model of the form (Thernau and Grambsch 2000)

$$\begin{aligned} h_i(t \mid \mathcal{Y}_i(t), z_i) &= \lim_{dt \to 0} \Pr(t \leq T_i^* < t + dt \mid T_i^* \geq t, \mathcal{Y}_i(t), z_i)/dt \\ &= h_0(t) \exp\left\{\gamma^T z_i + \alpha y_i(t)\right\}, \end{aligned} \tag{2.1}$$

where $\mathcal{Y}_i(t) = \{y_i(u), 0 \leq u < t\}$ denotes the history of the longitudinal process up to time point t, $h_0(t)$ denotes the baseline hazard function (i.e. the hazard function of a hypothetical individual with all covariates equal to zero), γ is a vector of regression coefficients related to the vector z_i, and α is the regression coefficient which quantifies the effect of the longitudinal outcome on the risk for an event. In the classical relative risk Cox model (Cox 1972), the baseline hazard function $h_0(t)$ is left unspecified. This is due to the fact that a hazard function may be too complicated to be modeled parametrically so that we can avoid a parametric assumption about its distribution and model it flexibly though nonparametric estimators. For this reason, the

relative risk Cox model is a semiparametric model: while a parametric form is assumed for the covariate effect, the baseline hazard function is treated nonparametrically.

In order to estimate parameters γ and α, statistical inference can be based on the likelihood method. In particular, Cox (1975) suggested to maximize not the entire likelihood function, but the part of it that contains all and only the regression parameters. This part is called partial likelihood and is defined as

$$L(\gamma, \alpha) = \prod_{i=1}^{n} \left[\frac{\exp(\gamma^T z_i + \alpha y_i(T_i))}{\sum_{j=1}^{n} \exp(\gamma^T z_i + \alpha y_i(T_i)) I(T_j > T_i} \right]^{\delta_i} \tag{2.2}$$

This is treated as a usual likelihood, and inference is carried out as usual. Nevertheless, it should be remember that it is not a likelihood in the ordinary sense of the word. It can be used for parameter estimation but any probability interpretation should be avoided (Kalbfleisch and Prentice 2002). Looking at (2.2) it can be noted that, thanks to the event indicator δ_i, the numerator depends only on the information from the individual who experienced the event, while the denominator uses all the information from individuals who had not experienced the event yet. This formula is only valid when there are no ties between the event times, that is, there are no events which happened at the same time. Indeed, it is usually assumed that the events actually happen in a continuous time so that two event times cannot happen at the same time. In real datasets, however, due to grouping and rounding, the recorded event times may coincide. Thus, several suggestions for handling ties in the partial likelihood can be found in the literature. The two most used belong to Breslow (1974) and Efron (1977) (see Klein and Moeschberger 2003 for more details).

It should be noted that, in order to be able to implement (2.2), several problems need to be solved. Fist, we need to know the value of $y(t)$ for each individual still alive at each observed event time. In practice, as already said, $y(t)$ is only available intermittently for each subjects. In fact, if

we think about post-treatment PSA value, it could be measured at approximately 6-months intervals or even once a year. This leads to missing data in time-dependent covariates in the survival model. Second, longitudinal covariates may be measured with error, such as PSA level, blood pressure, CD4 count, or viral load. The consequence is that the observed covariate values, i.e. y_i, are not their true values but mis-measured one. If covariate measurement errors are ignored in regression models, parameter estimates may be biased, hypothesis testing may lose power, and important features in the data may be masked. In our case, if covariate measurement errors are not taken into account, true covariate effects may not be correctly estimated or even detected. In fact, assessing the effect of measurement error in a longitudinal covariate on the parameter estimates of the Cox model, Prentice (1982) showed that the presence of measurement error causes the estimated parameters to be biased towards the null. Therefore, covariates with a strong prognostic effect could be considered of low importance and be discarded.

Let $m_i(t)$ denote the true and unobserved value of the longitudinal outcome at time t. Then, if we need to account for measurement error, the correct hazard function and partial likelihood on which statistical inference should be based are

$$\begin{aligned} h_i(t \mid \mathcal{M}_i(t), z_i) &= \lim_{dt \to 0} \Pr(t \leq T_i^* < t + dt \mid T_i^* \geq t, \mathcal{M}_i(t), z_i)/dt \\ &= h_0(t) \exp\left(\gamma^T z_i + \alpha m_i(t)\right), \end{aligned} \tag{2.3}$$

where $\mathcal{M}_i(t) = \{m_i(u), 0 \leq u < t\}$ denotes the true unobserved longitudinal process up to time point t; and

$$L(\gamma, \alpha) = \prod_{i=1}^{n} \left[\frac{\exp(\gamma^T z_i + \alpha m_i(T_i))}{\sum_{j=1}^{n} \exp(\gamma^T z_j + \alpha m_j(T_i)) I(T_j > T_i)} \right]^{\delta_i}$$

In Section 2.3, the most widespread methods proposed to address missing

data in the longitudinal outcome and to take into account measurement error are presented.

2.3 Methods for missing values and measurement error

2.3.1 Last Value or Last Observation Carried Forward (LVCF or LOCF)

This method is well described by Molenberghs and Kenward (2007) and it consists in replacing the missing value with the last available value. In our case, the missing value of the longitudinal outcome would be substituted by the value measured during the nearest preceding visit time. However, the simplicity of this approach cannot overcome its disadvantages. In fact, to ensure the validity of this method, very strong and often unrealistic assumptions have to be made. Fist, it has to be assumed that the subject's value of the covariate between two measurements times is constant. Clearly, this assumption is not biologically plausible since it is difficult to think that highly variable measures, such as CD4 counts for HIV patients, stay unchanged between two visit times (Bycott and Taylor 1998). Second, it treats observed values and imputed ones on the same level. Third, it does not consider the measurement error issue at all.

2.3.2 Two-stage approach

The subsequent approach proposed to reduce bias in parameter estimates was the two-stage approach. In the first stage, a model for the longitudinal process is fitted ignoring the survival outcome. Usually, a linear mixed effect model (LME) is used. Then, in the second stage, a survival model, often the relative risk model, is fitted using the subject-specific predictions of the time-dependent covariate based on the longitudinal model. This approach

showed to reduce bias compared to the simpler LVCF without completely eliminating it. Wu (2010, Chapter 8) reports two sources of bias characterizing this approach. First, the longitudinal outcome does not consider the possibility that the covariate trajectories of subjects who experienced the event or dropped out from the study may be different from those who were still alive at the end of the study. In this case the bias is due to the informative dropout, and may depend "on the strength of the association between the longitudinal process and the survival process". Second, since the inference in the second stage completely ignores the estimation uncertainty in the first stage, the standard errors may be under-estimated.

2.3.3 Joint models

The persistent bias characterizing parameter estimates of the two previous methods conducted researchers to focus their study on an approach based on the *joint likelihood* of all observed longitudinal and survival data. Since all parameters in the longitudinal and survival models are *simultaneously* estimated, the joint likelihood method avoids much of the bias in the previous two methods and provides the most efficient estimation if the assumed models are correct. Wulfsohn and Tsiatis (1997) were the first authors who proposed this approach assuming a relative risk model for the survival times with an unspecified baseline risk function. After them, many other researchers tried to extend their work by testing different model formulations (Proust-Lima and Taylor 2009; Yu, Taylor, and Sandler 2008; Henderson, Diggle, and Dobson 2000), assumptions validity (Rizopoulos and Verbeke 2008; Hsieh, Tseng and Wang 2006; Brown, Ibrahim, and DeGruttola 2005; Tsiatis and Davidian 2004), and suggesting improved estimation techniques (Rizopoulos, Verbeke, and Lesaffre 2009; Tsiatis and Davidian 2001).

In Chapter 3, the joint modeling approach will be formally presented and discussed in detail.

Chapter 3

Joint modeling of longitudinal and survival data

3.1 Introduction

Joint models were developed in order to properly answer the three principal interests already presented in 2.1. In summary, when emphinternal[1] time-dependent covariates are involved in survival analises, we need to account for their special characteristics in order to obtain valid inferences. Similarly, when the interest is on the longitudinal outcome, ignoring a possible non-random dropout process, may lead to biased estimates. In both cases, the class of joint models for longitudinal and time-to-event data constitutes a promising modeling paradigm to resolve these issues.

Tsiatis and Davidian (2004) state that a joint model is comprised by two linked sub-models, one for the 'true' longitudinal process $m_i(t)$ and one for the event time T_i^*, along with additional specifications and assumptions, which allow ultimately a full representation of the joint distribution of the

[1] An *internal* covariate is "typically the output of a stochastic process which is generated by the individual under study and in many instances is observed as long as the individual survives and is uncensored" (Kalbfleisch and Prentice 2002). Thus, biomarkers such as blood pressure, PSA level, CD4 count are examples of internal variables. The difference between internal and external covariates will be further discussed in Section 4.2.

observed data $\{T_i, \delta_i, y_i, t_i, z_i\}$, where $t_i = (t_{i1}, \ldots, t_{in_i})^T$.

At the end of Chapter 2, it was mentioned that a variety of models has been considered in the joint modeling literature. Therefore a choice needs to be made. For the remainder of this work, we will adopt those models firstly suggested by Wulfsohn and Tsiatis (1997) for joint models: a Gaussian linear mixed effect model for the longitudinal response and a relative risk model for the event times. The survival sub-model has already been described in Section 2.2 and is defined by the hazard function (2.3), but a modification has to be introduced. So far the baseline hazard function $h_0(t)$ has been left unspecified and thus estimated nonparametrically. However, for reasons which will be explained in Section 3.3.2, within the joint modeling framework, a parametric function should be employed. We could opt for a standard survival distribution such as the Weibull or Gamma distributions, or for more flexible solutions such as step functions or splines. In the following, ω will be the vector containing the parameters describing the parametric hazard function $h_0(t)$.

In Section 3.2, the problem of how to properly model the longitudinal process is addressed, while in Section 3.3, likelihood formulation and estimation will be described.

3.2 The measurement error model

In order to obtain non biased parameter estimates for γ and α in (2.3), it is crucial to estimate $m_i(t)$ and successfully reconstruct the complete longitudinal history, using the available measurements. It is considered appropriate to use the model known as *classical measurement error model* (Wu 2010, Chapter 5). Let $y_i(t_{ij})$ be the observed covariate value for the i-th individual at time t_{ij}, $i = 1, ..., n$, $j = 1, ..., n_i$, and let $m_i(t_{ij})$ be the corresponding unobserved true covariate value. The classical measurement error model assumes that

$$y_i(t_{ij}) = m_i(t_{ij}) + e_i(t_{ij}),\ E(e_i(t_{ij})|m_i(t_{ij})) = 0,\ i = 1, ..., n;\ j = 1, ..., n_i$$

where $e_i(t_{ij})$ represents the measurement error (and the possible biological variation) at time t_{ij}.

To model the true time-dependent covariate $m_i(t)$, it is common to assume that the covariate values change smoothly over time and thus a linear mixed effect model (LME) is the standard choice (Tsiatis and Davidian 2004). A LME model allows to take into account between-individual variation and within-individual correlation. In fact, Diggle et al. (2007) state that "when the goal is to understand the joint evolution of measurement and time-to-event at the level of an individual subject, we would favor random effects models", to whom LME belongs. More specifically, the chosen LME has the form

$$y_i(t) = m_i(t) + e_i(t) = u_i(t)^T\beta + v_i(t)^T b_i + e_i(t), \qquad i = 1, \ldots, n,$$

where $u_i(t)$ and $v_i(t)$ are known row vectors of the design matrices for the fixed and random effects, respectively, β is the vector containing unknown fixed parameters, b_i is the vector containing the random effects, and $e_i(t)$ is the measurement error term. We assume that b_i and $e_i(t)$ are independent, and $e_i(t) \sim N(0, \sigma^2)$ so that the within-individual covariate measurements are assumed to be conditionally independent given the random effects. This assumption is reasonable if the measurement times are sufficiently far apart that within-subject autocorrelation among observed values is essentially ignorable, or if the measurement error is large in comparison with biological fluctuations (Tsiatis and Davidian 2004). Therefore, at any given time t_{ij}, the unobserved true covariate value can be taken as

$$m_i(t_{ij}) = u_i(t_{ij})^T\beta + v_i(t_{ij})^T b_i, \qquad i = 1, \ldots, n; \qquad j = 1, \ldots, n_i.$$

One of the most discussed distributional assumptions is that of the random effects b_i with the multivariate normal distribution with mean zero and covariance matrix D, i.e. $b_i \sim N(0, D)$ being the usual choice. Verbeke et al. (2010, Chapter 2) underline two reasons causing the concern that relying on standard distributions, within the joint modeling framework, may influence the derived inferences. "First, the random effects have a more prominent role in joint models, because on the one hand they capture the correlations between repeated measurements in the longitudinal outcome and on the other, they associate the longitudinal outcome with the event process. Second, joint models belong to the general class of shared parameter models, and correspond to a non-random dropout mechanism. [...] As it is known from the missing data literature, handling dropout can be highly sensitive to modeling assumptions". Studies comparing estimates from models with normal random effects and with a more flexible distribution suggest a sort of robustness of the joint likelihood approach with normal random effects to departures from this assumption (Song et al. 2002, Davidian and Tsiatis 2004). Such robustness is seen both in parameter estimates and standard errors. This empirical conclusion has been recently corroborated from a theoretical point of view by Rizopoulos et al. (2008). They showed that, as the number of repeated measurements per subject n_i increases, misspecification of the random effects' distribution has a minimal influence in parameter estimators and standard errors. In order to obtain good estimates of $m_i(t)$, it is important to adequately specify $u_i(t)$ and $v_i(t)$ so that interesting features of the each subject's longitudinal trajectory can be captured. More specifically, when the interest is in modeling the longitudinal component only and not to make inference from it, and the covariate shows highly non-linear longitudinal trajectories, a linear mixed model where $u_i(t)$ and $v_i(t)$ are expressed in terms of high-order polynomials or splines[2] is strongly recommended. Particularly,

[2]In mathematics a spline is a sufficiently smooth piecewise-polynomial function. In the context of longitudinal data analysis, they are used to fit curves with a flexible but smooth form that is determined by the data. A detailed description about the use of splines in linear mixed effect models is given by Fitzmaurice et al. (2008, Chapter 11).

the use of splines allows researchers not to make rigid assumptions about the path of the biomarker over time and thanks to the subject specific random effects can well describe individuals' profiles across time. Indeed, Brown et. al (2005) affirm that, since "the model is able to change rapidly to reflect changes in the biomarkers over time, it may be preferable to parametric models for many types of data". The most used type of spline is the cubic B-spline. In comparison with other B-splines with a higher degree, the cubic B-splines ensures a great flexibility with a restricted number of parameters.

3.3 Likelihood formulation and estimation

The main estimation methods that have been proposed for joint models are based on a likelihood approach (Hsieh et al. 2006, Henderson et al. 2000, Wulfsohn and Tsiatis 1997) and a Bayesian approach developed via Markov chain Monte Carlo techniques (Hanson et al. 2011, Chi and Ibrahim 2006, Brown and Ibrahim 2003). In this work, we use the classical maximum likelihood method to obtain parameter estimates for the joint model.

3.3.1 Likelihood formulation

In the following part, according to the context, $f(\cdot)$ will be used as generic notation for distributions, densities, or likelihood contributions.

First of all, we need to define the joint distribution of the time-to-event and longitudinal outcomes $\{T_i, \delta_i, y_i\}$. Before doing that, however, there are two assumptions which need to be made:

1. It is assumed that "the vector of time-independent random effects bi underlies both the longitudinal and survival processes. This means that these random effects account for both the association between the longitudinal and event outcomes, and the correlation between the repeated measurements in the longitudinal process", i.e. $corr\{e_i(t_s), e_i(t_k)\} = 0$, for $t_s \neq t_k$ (Rizopoulos 2010). This assumption is also known as the

conditional independence assumption and, omitting covariates in the notation, can be formalised as

$$f(T_i, \delta_i, y_i|b_i; \theta) = f(T_i, \delta_i|b_i; \theta) f(y_i|b_i; \theta) \tag{3.1}$$

$$f(y_i|b_i; \theta) = \prod_{j=1}^{n_i} f\left\{y_i(t_{ij})|b_i; \theta\right\}, \tag{3.2}$$

where $\theta^T = (\theta_t^T, \theta_y^T, \theta_b^T)$ denotes the parameter vector, with θ_t denoting the parameters for the survival outcome, θ_y the parameters for the longitudinal outcomes and θ_b the parameters of the random effect covariance matrix D, y_i is the row vector containing the n_i longitudinal responses of the i-th subject. As it can be seen from (3.1) and (3.2), the conditional independence assumption allows the definition of separate models for the longitudinal and the survival outcomes by conditioning on the shared random effects b_i.

2. "Both the censoring mechanism and the visiting process (i.e., the stochastic mechanism that generates the time points at which the longitudinal measurements are collected) are non-informative, and thus they can be ignored" (Verbeke et al. 2010). If we define the observed longitudinal history as all the information available for the longitudinal process prior to time point t, i.e., $\mathcal{Y}_i(t) = \{y_i(u), 0 \leq u < t\}$, this assumption implies that the decision on whether a subject drops out from the study or appears at the study center for the scheduled visit depends only on $\mathcal{Y}_i(t)$ (and possibly on baseline covariates) and there is no further dependence on future measurements or on other subjects characteristics associated with the survival outcome but not taken into account.

We can now formulate the joint likelihood contribution for the i-th subject as

$$f(T_i, \delta_i, y_i; \theta) = \int f(T_i, \delta_i|b_i; \theta_t, \beta) \left[\prod_{j=1}^{n_i} f\left\{y_i(t_{ij})|b_i; \theta_y\right\}\right] f(b_i; \theta_b) db_i, \tag{3.3}$$

where the likelihood of the survival part has the following form

$$f(T_i, \delta_i | b_i; \theta_t, \beta) = \{h_i(T_i | \mathcal{M}_i(T_i); \theta_t, \beta)\}^{\delta_i} S_i(T_i | \mathcal{M}_i(T_i); \theta_t, \beta), \tag{3.4}$$

with $h_i(\cdot)$ given by (2.3), and

$$\begin{aligned} S_i(T_i | \mathcal{M}_i(T_i), z_i; \theta_t, \beta) &= \Pr(T_i^* > t | \mathcal{M}_i(t), z_i; \theta_t, \beta) \\ &= \exp\left\{ - \int_0^t h_i(s | \mathcal{M}_i(s); \theta_t, \beta) ds \right\}, \end{aligned} \tag{3.5}$$

$f\{y_i(t_{ij}) | b_i; \theta_y\}$ is the univariate normal density for the longitudinal responses, and $f(b_i; \theta_b)$ is the multivariate normal density for the random effects.

An important feature about the relative risk model considered here is underlined by (3.5): the risk for an event at time t is assumed to depend on the longitudinal history $\mathcal{M}_i(t)$ only through the current value of the time-dependent covariate $m_i(t)$, whereas the survival function depends on the whole history. This is another reason why, in order to obtain accurate predicted survival probabilities, it is important to properly model the longitudinal outcome.

3.3.2 Maximum likelihood estimation

Maximization of the likelihood function (3.3) with respect to θ is a computationally challenging task and it is the major drawback that prevented a higher diffusion of joint models. In particular, both the integrals with respect to the random effects in (3.3), and the integral in the definition of the survival function (3.5) do not usually have an analytical solution.

Tsiatis and Davidian (1997) were the first who proposed to use the expectation-maximization (EM) algorithm to maximize the joint likelihood of the observed data and thus obtain parameter estimates. The EM algorithm is an iterative method used to find maximum likelihood estimates of parameters when in the likelihood unobserved variables are present. The EM algorithm alternates between an E-step, which computes the expectation of

the log-likelihood using the current parameter estimates, and an M-step, where new parameter estimates are computed by maximizing the expected log-likelihood found with the E-step. These new parameter estimates are then used to determine the distribution of the latent variables in the next E-step. Fitzmaurice et al. (2008, Chapter 15) give a simple presentation of the EM algorithm applied in the joint modeling framework where the random effects are treated as unobserved variables (all parameters are combined into in the vector θ:

- Step 1: obtain initial parameter estimates through separate analyses of the longitudinal and survival data. Thus, the required random effects are included in the longitudinal model but not in the survival model. The parameter association between the longitudinal covariate and the event time, i.e. α, is set to zero.

- Step 2: write down the combined conditional log-likelihood $l(\theta; y_i, T_i, \delta_i, b_i)$ of the observed data given random effects b_i.

- Step 3: obtain the conditional expectation of each function of b_i, i.e. $g(b_i)$, appearing in $l(\theta; y_i, T_i, \delta_i, b_i)$, given (y_i, T_i, δ_i) and using the current estimates of θ.

- Step 4: replace each function of b_i appearing in $l(\theta; y_i, T_i, \delta_i, b_i)$ by its conditional expectation. Maximize the obtained log-likelihood to update the estimate of θ.

- Step 5: Iterate step 3 and 4 until convergence.

Steps 3 and 4 are the E-step and the M-step, respectively. In the E-step, the conditional expectations of several functions of the non observed random effects, i.e. $g(\cdot)$, often involve intractable integrals of the form

$$E\left\{g(b_i)|y_i, T_i, \delta_i; \hat{\theta}\right\} = \int g(b_i) f(b_i|y_i, T_i, \delta_i; \hat{\theta}) db_i$$

where $\hat{\theta}$ is the value from the M-step of the previous iteration. In order to make the computation of feasible, Wulfsohn and Tsiatis (1997) noticed that $f(T_i, \delta_i | y_i, b_i) = f(T_i, \delta_i | b_i)$ and used two integrals instead of one:

$$\begin{aligned} E\left\{g(b_i)|y_i, T_i, \delta_i; \hat{\theta}\right\} &= \int g(b_i) f(b_i|y_i, T_i, \delta_i; \hat{\theta}) db_i \\ &= \frac{\int g(b_i) f(y_i, \delta_i|b_i; \hat{\theta}) f(b_i|y_i; \hat{\theta}) db_i}{\int f(y_i, \delta_i|b_i; \hat{\theta}) f(b_i|y_i; \hat{\theta}) db_i} \end{aligned} \tag{3.6}$$

At this point we can take advantage of the last form (3.6) and use numerical integration to solve both integrals. Standard numerical integration techniques such as Gaussian quadrature and Monte Carlo methods are usually applied (Wulfsohn and Tsiatis 1997, Henderson et al. 2000, Wu 2010). The use of Laplace approximations has recently been discussed by Rizopoulos et al. (2009) in situations where a large number of random effects is involved (e.g., when splines are used in the random effects design matrix).

The M-step is more straightforward since closed-form maximum likelihood estimates for parameters D and σ^2 are available. On the contrary, maximum likelihood estimates for parameters β, γ, and α do not have a closed-form solution and therefore are computed using the Newton-Raphson method (see Rizopoulos et al. (2009) for more details).

We wish to go back for a moment to the choice made at the end of Section 3.1 about the risk function $h_0(t)$. Although this function is typically left unspecified, in the joint modeling framework Hsieh et al. (2006) warned researchers from using the Fisher Information to obtain parameters' standard errors since their values would be underestimated. According to the authors, this problem arises from the fact that "the nonparametric maximum likelihood for $h_0(t)$ cannot be solved explicitly under the random effect structure". Furthermore, this causes the maximum profile likelihood estimates of the other parameters to be implicit as well since they depend on $h_0(t)$. Consequently, the authors suggested to use the bootstrap technique to compute the standard errors but its validity has not been proven yet. We choose then to estimate parametrically the risk function $h_0(t)$ so that the usual para-

metric asymptotic arguments for the maximum likelihood estimates can be applied.

Chapter 4

Individual survival probability estimation

4.1 Introduction

In Chapter 1 we have shown how the need of accurate prognoses has increased in the last decades in everyday medical practice. In particular, the possibility of obtaining survival probabilities based on characteristics of the specific patient would greatly help physicians in their decision making and patient counseling, thus improving clinical output.

In the joint modeling framework, the interest on subject-specific predictions for either the longitudinal or survival outcomes has increased in recent years (Sweeting and Thompson 2011; Proust-Lima and Taylor 2009; Yu, Taylor and Sandler 2008; Taylor, Yu, and Sandler 2005). In these studies, the prognostic information contained in longitudinal measurements of covariates, such as biomarkers, has been used to make predictions about patients' survival probabilities. However, this aspect needs to be further analysed and tested in order to be safely used in clinical practice.

Using the joint modeling formulation presented in Chapter 3, Rizopoulos (2011) proposed a Monte Carlo approach to estimate survival probabilities and their standard errors based on the output of a fitted joint model. In par-

ticular, the author considered dynamic subject-specific predictions, and illustrated how survival probabilities are updated as additional measurements of the longitudinal outcome are available. In Section 4.2 we will formally present this approach and then, in Chapter 5, we will use it to predict patients' survival.

4.2 Expected survival probabilities

The interest lies in predicting survival probabilities for a new subject i which has provided a set of longitudinal measurements $\mathcal{Y}_i(t) = \{y_i(s); 0 \leq s \leq t\}$ (dependence on baseline covariates is assumed but suppressed for ease of exposition).

Before going any further, it is important to carefully take into account the already mentioned feature typical of covariates such as, for example, PSA level, blood pressure and serum bilirubin level. In our framework, $y_i(t)$ represents an internal time-dependent covariate. More specifically, Kalbfleisch and Prentice (2002) identify two kinds of time-dependent covariates: external and internal. The condition which is needed to be satisfy by a time-dependent covariate $x(t)$ in order to be external is

$$\Pr(x(t) \mid x(u), T \geq u) = \Pr(x(t) \mid x(u), T = u), \;\; 0 < u \leq t. \tag{4.1}$$

The direct consequence is that, whereas the covariate $x(\cdot)$ may influence the event mechanism over time, its future path up to any time $t > u$ is not affected by the occurrence of an event at time u. A type of external covariate is the baseline covariate for which $x(t) = x$. Its value is measured at the beginning of the study and it will not change over time. Other examples of external time-dependent covariates are treatment, air pollution, atmospheric temperature. Those time-dependent covariates which do not satisfy condition (4.1) are called internal. This type of covariates have the property that, in order to be measured, the individual has to be alive and uncensored.

Their path, thus, carries information on the time of dropout. Since the covariate $y_i(t)$ considered here is internal, a subject who provides longitudinal measurements up to time t gives also the information that he/she was alive and uncensored at that time.

Hence, it is more appropriate to focus on the conditional probability of surviving time $u > t$ given survival up to time t, that is,

$$\pi_i(u \mid t) = \Pr(T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t), _n ; \theta), \tag{4.2}$$

where $_n = \{T_i, \delta_i, y_i; i = 1, \ldots, n\}$ denotes the sample on which the joint model was fitted and on which we wish to base our predictions. Using the assumption that the vector of random effects b_i undelies both the longitudinal and survival process (assumption formalized by (3.1)), Rizopoulos (2011) observed that (4.2) can be written as

$$\begin{aligned}
&\Pr(T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t); \theta) \\
&= \int \Pr(T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t), b_i; \theta) f(b_i \mid T_i^* > t, \mathcal{Y}_i(t); \theta)\ db_i \\
&= \int \Pr(T_i^* \geq u \mid T_i^* > t, b_i; \theta) f(b_i \mid T_i^* > t, \mathcal{Y}_i(t); \theta)\ db_i \\
&= \int \frac{S_i\{u \mid \mathcal{M}_i(u, b_i, \theta); \theta\}}{S_i\{t \mid \mathcal{M}_i(t, b_i, \theta); \theta\}} f(b_i \mid T_i^* > t, \mathcal{Y}_i(t); \theta)\ db_i,
\end{aligned} \tag{4.3}$$

where $S_i(\cdot)$ is given by (3.5). It is important to notice that the longitudinal history $\mathcal{M}_i(\cdot)$, approximated by the measurement error model presented in Section 3.2, is a function of both the random effects and the parameters.

At this point, a first order estimate of $\pi_i(u \mid t)$ can be obtained using the empirical Bayes estimate of b_i. However, the derivation of its standard error and or confidence interval is rather difficult since we need to take into account the variability of both the maximum likelihood and empirical Bayes estimates. To overcome this problem and produce a valid standard error for the estimate of $\pi_i(u \mid t)$, Rizopoulos (2011) proposes to follow an asymptotic

Bayesian formulation of the joint model and therefore derive the posterior expectation of (4.2), which can be written as

$$\begin{aligned}\pi_i(u \mid t) &= \Pr(T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t), {}_n) \\ &= \int \Pr(T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t); \theta) p(\theta \mid {}_n)\, d\theta. \end{aligned} \tag{4.4}$$

The first part of the integrand is given by (4.3). The second part is the posterior distribution of the parameters given the observed data. The author, using arguments of the standard asymptotic Bayesian theory, assumes that the sample size n is large enough such that $\{\theta \mid {}_n\}$ can be well approximated by $N(\widehat{\theta}, \widehat{\text{Var}})$, with $\widehat{\text{Var}} = \widehat{var}(\widehat{\theta})$. Then, in order to obtain a Monte Carlo estimate of $\pi_i(u \mid t)$, he proposes the following simulation scheme:

- Step 1: Draw $\theta^{(l)} \sim N(\widehat{\theta}, \widehat{\text{Var}})$.
- Step 2: Draw $b_i^{(l)} \sim \left\{b_i \mid T_i^* > t, \mathcal{Y}_i(t), \theta^{(l)}\right\}$.
- Step 3: Compute

$$\pi_i^{(l)}(u \mid t) = \frac{S_i\left\{u \mid \mathcal{M}_i(u, b_i^{(l)}, \theta^{(l)}); \theta^{(l)}\right\}}{S_i\left\{t \mid \mathcal{M}_i(t, b_i^{(l)}, \theta^{(l)}); \theta^{(l)}\right\}}. \tag{4.5}$$

- Step 4: Repeat Steps 1-3 for each subject i, $l = 1, \ldots, L$ times, where L denotes the number of Monte Carlo samples.

If Steps 1 and 3 are straightforward to perform, the posterior distribution of the random effects given the observed data in Step 2 is not of standard form. Thus, the author proposes to use the Metropolis-Hastings algorithm[1]

[1]The Metropolis-Hastings algorithm is a Markov chain Monte Carlo method used to simulate complex, nonstandard multivariate distributions. The M-H algorithm is based on proposing values sampled from an instrumental distribution, which are then accepted with a certain probability that reflects how likely it is that they are from the target distribution (Chib and Greenberg 1995). In our case, the instrumental distribution coincides with the multivariate t distribution while the target distribution is the multivariate distribution of b_i.

with independent proposals from a multivariate t distribution centered at the empirical Bayes estimated $\widehat{b_i}$, with scale matrix

$$\widehat{var}(\widehat{b_i}) = \left\{ -\frac{\partial^2}{\partial b^T \partial b} \log f(T_i^* > t, \mathcal{Y}_i(t), b; \widehat{\theta}) \Big|_{b=\widehat{b_i}} \right\}^{-1},$$

and four degrees of freedom. The author chooses a multivariate t proposal for two reasons. First, with other two collegues, he has recently shown that, as n_i increases, the leading term of the log posterior distribution of the random effects is the logarithm of the density of the LMM model $\log f\left\{\mathcal{Y}_i(t) \mid b_i; \theta^{(l)}\right\} = \sum_j \log f\left\{y_i(t_{ij}) \mid b_i; \theta^{(l)}\right\}$, which is quadratic in b_i and will resemble the shape of a multivariate normal distribution (Rizopoulos, Verbeke and Molenberghs 2008). Second, if n_i is small, the heavier tails of the t distribution will ensure sufficient coverage.

Once the realizations $\pi_i^{(l)}(u \mid t)$, $l = 1, \ldots, L$ are available, we can derive estimates of $\pi_i(u \mid t)$, such as

$$\widehat{\pi}_i(u \mid t) = median\left\{\pi_i^{(l)}(u \mid t),\ l = 1, \ldots, L\right\} \tag{4.6}$$

or

$$\widehat{\pi}_i(u \mid t) = \frac{1}{L} \sum_{i=1}^{L} \pi_i^{(l)}(u \mid t). \tag{4.7}$$

The standard error of $\hat{\pi}_i(u \mid t)$ can be computed using the sample variance over the Monte Carlo samples, while its confidence interval can be obtained using the Monte Carlo sample percentiles.

Both (4.2) and (4.5) formally show how the expected survival probability is forced to be equal to 1 while $u \leq t$ and then, when $u > t$, is allowed to decrease. Moreover, the estimates of $\pi_i(u \mid t)$ can easily be updated when new covariate measures are available. In fact, in Step 2, the multivariate t distribution of the random effects b_i is based on the whole individual covariate history $\mathcal{Y}_i(t)$.

4.3 Considering alternative association structures

So far we have used the standard association structure for the joint model formulation, that is, the risk for an event at a specific time point t depends on the true level of the longitudinal covariate at the same time point. Its hazard function is defined as

$$h_i(t) = h_0(t) \exp\left\{\gamma^T z_i + \alpha_1 m_i(t)\right\}$$

where the parameter α_1 incorporates the strength of the association between the current level of the covariate and the risk. However, as Fisher and Lin (1999) noticed, "the choice of a time dependent covariate involves the choice of a functional form for the time-dependence of the covariate. This choice is usually not self-evident but may be suggested by biological understanding or biological hypothesis". The functional form issue needs to be properly tackled since choosing the wrong covariate functional form may substantially influence the derived results. The aim of this work is exactly that of verify whether different association structures (which can be practically considered as functional forms) actually influence individual survival probabilities and to individuate possible patterns characterizing these differences.

In the following we present seven alternative parameterizations to the standard parameterization which can be use to model the association between the covariate and the risk.

Slope parameterization

As we have seen in Chapter 1, the CD4 cell counts is a good longitudinal covariate which helps to monitor HIV progression. A decreasing CD4 count over time indicates a worsening of patient health. On the contrary, in patients with PBC, an increasing serum bilirubin level indicates a worsening of patient condition since it means that the liver is failing. In both cases, the slope could be a good prognostic factor since it indicates the rate at which

CD4 count and serum bilirubin level is respectively decreasing or increasing. In general, the slope of the longitudinal trajectory can be a good indicator of the velocity with which the disease is progressing. The relative risk survival sub-model takes the form

$$h_i(t) = h_0(t) \exp\left\{\gamma^T z_i + \alpha_2 m_i^{'}(t)\right\}$$

where

$$m_i^{'}(t) = \frac{d}{dt} m_i(t) = \frac{d}{dt}\left\{u_i^T(t)\beta + v_i^T(t) b_i\right\}.$$

Parameter α_2 measures how strongly associated is the value of the slope of the true longitudinal trajectory at time t with the risk for an event at the same time point.

Current value + Slope parameterization

This parameterization incorporates into the model both the information about the value of the true covariate at time t and the value of the slope at the same time point. With this parameterization we can distinguish between patients with similar $m_i(t)$ values but with different slopes. The relative risk survival sub-model takes the form

$$h_i(t) = h_0(t) \exp\left\{\gamma^T z_i + \alpha_3 m_i(t) + \alpha_{3.s} m_i^{'}(t)\right\}.$$

Parameter α_3 measures the strength of the association between the true value of the longitudinal covariate at time point t with the risk for an event at the same time point, given that $m_i^{'}(t)$ stays constant. Similarly, $\alpha_{3.s}$ measures how strongly associated is the value of the slope of the true covariate trajectory at time t with the risk for an event at the same time point, provided that $m_i(t)$ remains constant.

Cumulative effect parameterization

The two previous parameterizations assume that the risk for an event at a specific time point t depends on the true covariate level and/or slope value at the same time point. However, this assumption may not be always realistic.

Consider, for example, the effect of smoking on the risk of heart attack. The risk at time t of a patient who used to smoke 30 cigarettes per day and only in the last period has started to smoke only 3 cigarettes per day may be much higher than the risk of a patient who has always smoked 3 cigarettes per day, being equal all the other covariates.

One method to account for the cumulative effect of the longitudinal covariate to time point t is to use the integral of the true longitudinal trajectory up to that time point. The relative risk survival sub-model takes the form

$$h_i(t) = h_0(t) \exp\left\{\gamma^T z_i + \alpha_4 \int_0^t m_i(s)ds\right\}.$$

The parameter α_4 measures the strength of the association between the risk for an event at time point t and the area under the true longitudinal trajectory up to the same time point, which is obtained through the integral value.

Weighted cumulative effect parameterization

The cumulative effect parameterization places the same weight for all past values of the true longitudinal covariate. In situations where more recent values have a stronger influence on the risk than previous values, this type of parameterization may not be completely appropriate. In order to overcome this problem, we could multiply $m_i(t)$ with a weight function that places different weights at different time points and then compute the integrand of the adjusted covariate. In this case, the relative risk survival sub-model takes the form

$$h_i(t) = h_0(t) \exp\left\{\gamma^T z_i + \alpha_5 \int_0^t g(t-s)m_i(s)ds\right\},$$

where $g(\cdot)$ denotes the weight function. In this work, we have chosen to use as weight function the standard normal density $g(x) = \exp(-x^2/2)$. Since the variance is equal to 1, in practice, the three most recent years of the true covariate history $\mathcal{M}_i(t)$ are associated with the risk for death. The parameter α_5 measures the strength of the association between the risk for an event at time point t and the area under the longitudinal trajectory of $g(t-s)m_i(s)$,

$0 \leq s \leq t$.

Lagged value parameterization

There are situations where the risk for an event at time point t is mainly influenced by the covariate value at a previous time point. For example, Cavender et al. (1992) analysed data from a study on patients with coronary artery disease. Each patient was interviewed every 6 months and, among other variables, his/her smoking status was recorder. Surprisingly, the authors found that the estimated effect of the current smoking was positive although not statistically significant. From the analysis of the individual patient smoking histories, it turned out that most of those who had died were smokers, but many of them had stopped smoking at the last follow-up before their death. The consequence was that many of the patients who died have just quit smoking, whereas some of the patients who were still alive were smoking. This explains why the model gave that unexpected result.

One way to address this problem is to use time-lagged covariates. In this case, the relative risk survival sub-model takes the form

$$h_i(t) = h_0(t) \exp\left\{\gamma^T z_i + \alpha_6 m_i \left[\max(t - c, 0)\right]\right\}.$$

With this parameterization we assume that the risk at time point t depends on the true value of the longitudinal covariate at time $t - c$, where c specifies the time lag of interest. Parameter α_6 measures how strongly associated is the value of the true longitudinal covariate at time $t - c$ with the risk for an event at time point t.

Lagged slope parameterization

Similarly to what happen for the current slope parameterization, we can consider the effect of the slope value at time $t - c$ on the risk at time t. The relative risk survival sub-model takes the form

$$h_i(t) = h_0(t) \exp\left\{\gamma^T z_i + \alpha_7 m_i^{'} \left[\max(t - c, 0)\right]\right\}.$$

Parameter α_7 measures how strongly associated is the value of the slope of

the true longitudinal trajectory at time $t - c$ with the risk for an event at time point t.

Lagged value + lagged slope parameterization

The last parameterization which will be considered here involves both the value of the true covariate and the slope of the true longitudinal trajectory at time point $t - c$. The relative risk survival sub-model has the form

$$h_i(t) = h_0(t) \exp\left\{\gamma^T z_i + \alpha_8 m_i \left[\max(t - c, 0)\right] + \alpha_{8.s} m_i^{'} \left[\max(t - c, 0)\right]\right\}.$$

Parameter α_8 measures the strength of the association between the true value of the longitudinal covariate at time point $t - c$ with the risk for an event at time point t, given that $m_i^{'}(t-c)$ stays constant. Similarly, $\alpha_{8.s}$ measures how strongly associated is the value of the slope of the true covariate trajectory at time $t - c$ with the risk for an event at time point t, provided that $m_i(t - c)$ remains constant.

In Chapter 5 we will use a dataset based on a study involving patients diagnosed with primary biliary cirrhosis. We will compare the individual expected survival probabilities given by the model using the standard parameterization with those compute by the models using the other seven parameterizations. However, the eight models will not contain any of the baseline covariates z_i. In fact, each model has its own estimates of the parameters contained in the vector γ and this affects models' prediction. Since we want to study the influence of the association structure on the expected survival probabilities, the presence of the baseline covariates z_i in the survival sub-model would make it difficult to distinguish between their effect and that of the association structure.

Furthermore, the reader could notice that also the values of the parameters defining the parametric baseline hazard function $h_0(t)$ and the true covariate trajectory are not usually the same. However, the influence of this differences is small in most of the cases, although it should not be ignored a priori, as we will see in Chapter 5.

Chapter 5

Comparing expected survival probabilities: the PBC dataset

5.1 Primary biliary cirrhosis: the disease

The Mayo Clinic website reports that "primary biliary cirrhosis is a disease in which the bile ducts in your liver are slowly destroyed. Bile, a fluid produced in your liver, plays a role in digesting food and helps rid your body of worn-out red blood cells, cholesterol and toxins. When bile ducts are damaged, as in primary biliary cirrhosis, harmful substances can build up in your liver and sometimes lead to irreversible scarring of liver tissue (cirrhosis)"(Mayo Clinic). At the moment, it is not clear what triggers this disease and it is believed to have an autoimmune etiology.

The typical patient is a middle-aged women who reports fatigue and itching or who has no symptoms but has been found to have unexplained hepatomegaly or abnormal serum liver tests. The natural history of untreated primary biliary cirrhosis (PBC) is one of gradual progression through four phases: preclinical, asymptomatic, symptomatic (including systemic and portal hypertensive), and liver insufficiency (Mayo 2008). Despite extensive studies, medical therapy has not been shown to have a significant impact in slowing the progression of PBC so that the only truly effective treatment

is liver transplantation (Markus et al. 1989). Consequently, there has been considerable interest in developing models which could accurately predict patients' survival probabilities in order to better individuate those who are going to strongly need a new liver within a couple of years. In fact, it has been found that, from diagnosis to liver failure, can pass more then 20 years in asymptomatic patients (Pares and Rodes 2003).

Although the four phases are usually respected by most of the patients (i.e. it is rare that a patient skips one phase), the velocity of disease progression can significantly change from patient to patient: some patients will have a relatively slow, benign clinical course whereas others will progress more rapidly to portal hypertension and liver insufficiency. Thus, the ability to identify surrogate markers of disease progression is extremely important. Of the serum tests widely available, serum bilirubin is the single most useful predictor of clinical outcome, although it does not become elevated until later in the disease process (Mayo 2008).

Serum bilirubin level is the prognostic factor used in our analysis.

5.2 Description of the PBC dataset

This dataset contains information of 312 patients who were enrolled in two clinical trials evaluating the use of D-penicillamine for treating PBC. Patients were followed from January 1974 through May 1984. The first visit was made at the study entry, the second after six months, the third after one year and then at approximately 1 year intervals. However, as often happens in longitudinal studies, the visit schedule was not precisely followed. No treatment benefit in prolonging survival was found so that the follow up was extended to April 30, 1988 in order to study the natural history of the PBC disease, i.e. how the disease progresses in time and which factors may influence such progression. By 1988, 140 (44.9%) patients died, 29 (9.3%) were transplanted and the other 143 (45.8%) survived. As happened in previous studies, transplanted patients were censored at the time of transplantation.

At the end of the study, there were 1945 total visits. The median number of serum bilirubin values per patient was 5, with a minimum of 1 and a maximum of 16 measurements. The median number of years in the study was 6.296, with a minimum of 0.1123 (i.e. 41 days) and a maximum of 14.31 years. Some baseline covariates, such as gender and age, were recorded at the beginning of the study. Laboratory test results, such as serum biliribin, albumin and alkaline phosphatase, were measured at each visit.

5.3 Description of the models

The analysis were done using the statistical software R. In particular, Rizopoulos (2010a) presented the R package JM that can be used to fit joint models with the formulation proposed in this work. Before this package, only a separate analysis of longitudinal and event time data was possible. The procedure followed here to fit the eight joint models is composed by three steps:

- Step 1: fit a proper mixed effect model for the longitudinal covariate (R package used: lme).
- Step 2: fit a relative risk model for the survival part without using any covariate (R package used: survival).
- Step 3: fit the joint model using the parameterization that defines the chosen association structure (R package used: JM).
- Step 4: repeat Step 3 for each of the remaining parameterizations.

In Step 1, we concentrate our attention on modeling the longitudinal trajectory. Since serum bilirubin exhibits a right skewed shape distribution, its logarithm is preferred (Fleming and Harrington 1991). For the remainder of this work, our longitudinal covariate will be $log(serBilir)$, where $serBilir$ is the label given to the original serum bilirubin covariate in the PBC dataset.

A measurement error model with a quadratic trend was chosen to model the longitudinal trajectory of $log(serBilir)$. Therefore the model takes the form

$$\begin{aligned} \log(serBilir)_i(t) &= m_i(t) + e_i(t) \\ &= (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t + (\beta_2 + b_{i2})t^2 + e_i(t), \quad (5.1) \end{aligned}$$

where $e_i(t) \sim N(0, \sigma^2)$ and

$$\begin{pmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 \end{pmatrix} \right)$$

It should be kept in mind that, in this step, the parameter estimates are only temporary. Their final values will be calculated in Step 3, where also the information from the survival part is taken into account. However, the non significance of the parameter β_2 found in this temporary model persists in all the final longitudinal models (see Table 5.2). Nevertheless, we chose to keep the fixed quadratic term to preserve interpretability.

Step 2 is only a technical step used to create objects used by the JM package in Step 3.

In Step 3, all the parameters are estimated simultaneously. Therefore, the estimates of the survival part are also based on the information contained in the longitudinal part, and vice versa. We used a piecewise constant function to model baseline risk function $h_0(t)$. More specifically,

$$h_0(t) = \sum_{q=1}^{Q} \omega_q I(r_{q-1} < t \leq r_q),$$

where $0 = r_0 < r_1 < \ldots < r_Q$ denotes the split of the time scale, with r_Q being larger that the largest observed time, and ω_q denotes the value of the

hazard in the interval $(r_{q-1}, r_q]$. In our case, $Q = 7$ and the six time points that divide the time interval $(0, r_Q)$ are the quantiles of the observed event times (Rizopoulos 2010b).

About the three models containing lagged covariate values and/or lagged slopes, we used a lag equal to 2 years. In fact, in a separate analysis testing for different lag times where also baseline covariates were involved, the joint model with a lag of 2 years was the best model, according to both the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC)[1].

For ease of exposition, we have labeled with a short name the eight models corresponding to the eight different parameterizations. See Table 5.1. Furthermore, when not otherwise specified, we will use the noun '*covariate*' to indicate the longitudinal variable $\log(serBilir)$ without the measurement error, which coincides with $m_i(t)$ in (5.1).

Table 5.2 contains the estimates of the parameters belonging to the longitudinal part of the joint model for each of the eight models. Similarly, Table 5.3 contains the estimates of the parameters of the survival part.

Observing the estimates of the fixed effect parameters in Table 5.2 we see that they are quite similar. As already announced, the quadratic term is practically null. Moreover, from the values of the estimates of the covariances σ_{01} and σ_{12}, we can see that those patients with a high intercept tend to have an increasing trend, and that those with an increasing trend have a concave trajectory.

[1]Both AIC and BIC are measures of the relative goodness of fit of a statistical model. They are used when we need to compare models that are not nested so that the classical likelihood ratio test (LRT) cannot be applied. The AIC and BIC are defined as

$$AIC = 2p - 2l(\hat{\theta})$$

$$BIC = p\log(N) - 2l(\hat{\theta}),$$

where $-2l(\hat{\theta})$ is the maximized log-likelihood under the fitted model, $\hat{\theta}$ is the maximum likelihood estimate of θ, p is the number of parameters in the model, and N is the total number of observations used to fit the model (Wu 2010). When comparing different models, the model with the smaller AIC or BIC is to be preferred.

Model name	Parameterization
J1	current value
J2	current slope
J3	current value + current slope
J4	cumulative effect
J5	weighted cumulative effect
J6	lagged value
J7	lagged slope
J8	lagged value + lagged slope

Table 5.1: Model name and corresponding parameterization

	Model J1	Model J2	Model J3	Model J4	Model J5	Model J6	Model J7	Model J8
β_0	0.515	0.451	0.500	0.512	0.514	0.516	0.485	0.509
(s.e.)	(0.057)	(0.043)	(0.048)	(0.058)	0.058	(0.058)	(0.044)	(0.053)
β_1	0.167	0.172	0.180	0.164	0.162	0.164	0.197	0.186
(s.e.)	(0.023)	(0.019)	(0.022)	(0.023)	(0.023)	(0.022)	(0.020)	(0.023)
β_2	0.001	0.003	0.001	0.001	0.001	0.001	0	0
(s.e.)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
σ_0^2	1.001	0.976	0.996	1.001	1.000	1.002	0.960	0.995
σ_1^2	0.301	0.295	0.319	0.302	0.304	0.297	0.313	0.315
σ_2^2	0.025	0.020	0.025	0.025	0.025	0.024	0.022	0.025
σ_{01}	0.198	0.421	0.275	0.168	0.173	0.185	0.5067	0.286
σ_{02}	0.002	0.019	-0.049	0.017	0.022	-0.004	-0.309	-0.091
σ_{12}	-0.867	-0.739	-0.853	-0.883	-0.873	-0.880	-0.842	-0.869
σ^2	0.303	0.311	0.303	0.305	0.304	0.306	0.311	0.304

Table 5.2: Parameter estimates of the longitudinal part of the eight joint models

Considering the estimates of the survival part in Table 5.3, we notice that all the parameters are highly significant (p-value < 0.0001). This means that serum bilirubin, independently of the parameterization used, is strongly related with the risk for death. For example, according to model J1, each unit increase of the current value of the *covariate* is associated with a $exp(1.314)$ = 3.72-fold increase (95% CI: 3.37; 4.11) in a patient's risk. According to model J8, being $m_i'(t-c)$ fixed, each unit increase of the lagged value of the *covariate* is associated with a exp(1.106) = 3.02-fold increase (95% CI: 2.72; 3.36) in the patient's current risk. Similarly, being $m_i(t-c)$ fixed, each unit increase of the lagged slope is associated with a exp(3.740) = 42.10-fold

	Model J1	Model J2	Model J3	Model J4	Model J5	Model J6	Model J7	Model J8
α_j	1.314	7.241	1.237	0.184	2.435	1.216	6.943	1.106
(s.e.)	(0.099)	(0.777)	(0.115)	(0.020)	(0.183)	(0.092)	(0.770)	(0.106)
$\alpha_{j.s}$	-	-	3.006	-	-	-	-	3.740
(s.e.)	-	-	(0.624)	-	-	-	-	(0.582)
$\log(\omega_1)$	-4.573	-5.620	-5.641	-2.907	-3.698	-4.026	-5.883	-5.313
(s.e.)	(0.258)	(0.433)	(0.409)	(0.155)	(0.193)	(0.217)	(0.479)	(0.380)
$\log(\omega_2)$	-4.429	-4.344	-5.211	-2.928	-3.954	-3.607	-4.448	-4.554
(s.e.)	(0.274)	(0.318)	(0.361)	(0.180)	(0.236)	(0.214)	(0.334)	(0.304)
$\log(\omega_3)$	-4.710	-4.360	-5.352	-3.456	-4.281	-3.960	-4.214	-4.752
(s.e.)	(0.318)	(0.329)	(0.383)	(0.257)	(0.290)	(0.273)	(0.319)	(0.331)
$\log(\omega_4)$	-4.672	-4.304	-5.257	-3.608	-4.266	-3.976	-3.991	-4.651
(s.e.)	(0.370)	(0.376)	(0.424)	(0.331)	(0.347)	(0.334)	(0.354)	(0.376)
$\log(\omega_5)$	-4.396	-4.198	-4.973	-3.533	-4.012	-3.748	-3.686	-4.313
(s.e.)	(0.343)	(0.356)	(0.394)	(0.315)	(0.319)	(0.306)	(0.315)	(0.342)
$\log(\omega_6)$	-4.021	-3.919	-4.729	-3.486	-3.633	-3.420	-3.310	-3.878
(s.e.)	(0.353)	(0.388)	(0.431)	(0.379)	(0.329)	(0.321)	(0.333)	(0.347)
$\log(\omega_7)$	-4.549	-3.865	-4.752	-5.162	-4.302	-4.235	-3.559	-4.318
(s.e.)	(0.477)	(0.478)	(0.507)	(0.605)	(0.463)	(0.459)	(0.450)	(0.472)

Table 5.3: Parameter estimates of the survival part of the eight joint models

increase (95% CI: 23.52; 75.34) in the patient's current risk. Interestingly, comparing the estimates of α_1 and α_6 with those of α_3 and α_8, we see that the latter are slightly smaller that the former. This suggests that indeed the trajectory slope adds some information in the model, however, the *covariate* value still keeps its importance. The small value of the parameter α_4 should be kept in mind since it will have an important role in understanding the peculiar behavior characterizing the differences between model J1 and model J2 survival predictions.

5.4 Expected survival probabilities comparison

For each of the eight models, we computed the following expected survival probabilities (ESPs), which we remind are *conditional* probabilities: $S(t + \Delta t \mid t)$, where $t = 2, 4, 6, 8$ years and $\Delta t = 1, 2, 4$ years. Comparing the ESPs of the eight considered models we found that, indeed, there are differences which can be attributed to the specific association structure assumed in the survival sub-model. The ESPs are obtained using the estimator defined by (4.7). Conclusions would not have changed if we had used the estimator

defined by (4.6).

We compared the ESPs of the standard joint model, i.e. model J1, with those of the other seven models. The number of comparisons that can be made is rather high: there are 12 types of $S(t+\Delta t \mid t)$, and 7 ESP comparisons for each of them. Thus, in total, we analysed 84 comparisons. We noticed that the reasoning behind the 7 ESP comparisons for one type of $S(t+\Delta t \mid t)$ is also applicable in the other 11 cases. Thus, we have chosen to concentrate our attention on $S(4 \mid 2)$. Our choice is base on the fact that, at year 2, the majority of the patients are still alive and uncensored. Moreover, a Δt equal to 2 gives enough time to the ESPs to decrease and thus leave the usually high values typical of time points immediately after t. We remind, indeed, that, independently of the risk level of the patient and of the model, ESPs at time t are always equal to 1 and then start to decrease with a rate which depends on the risk level attribute to the patient by the model. This first part of the analysis will be discussed in detail in Section 5.4.1. We will show how different association structures substantially change survival predictions using $S(4 \mid 2)$ as example.

In Section 5.4.2, we will consider another aspect charaterizing the differences between ESPs. Concentrating on only those patients who were still alive and uncensored at year 6, we will show how the difference between the ESP of model J1 with respect to that of the other seven models can change as Δt increases.

For ease of exposition, the expected survival probability estimated by model Ji, $i = 1, \ldots, 8$, will be indicated with ESPi. For example, the ESP computed by model J1 will be indicated with ESP1.

Finally, before starting with the description of our findings, we wish to point out one aspect that characterizes the longitudinal part of the models we are considering here. At the end of Chapter 4, we mentioned that the values of the parameters defining the longitudinal trajectory differ from model to model. Observing the trajectories predicted by model J1 with those predicted by the other models, we noticed that the 'biggest' differences belong to model

J2. The first row of Figure 1 contains the trajectories for patients 180, 133, and 93 based on the serum bilirubin measurements collected in the first 2 years of study. Likewise, those plotted in the second row are based on the measurements collected in the first 6 years. Each plot contains the trajectory predicted by model J1 (solid line) and by model J2 (dashed line). Usually, the differences between the two curves become more evident in later years. Moreover, the addiction of new bilirubin values does not always guarantee a reduction of these differences (for instance, compare the plots of patient 133). Also the curves predicted by model J7 are sometimes different from those of model J1, but are closer to them than those predicted by model J2 (results not shown). About the other models, their curves are, if not coincident, very close to those of model J1 so that the differences are imperceptible. Having said that, in order to correctly understand the mechanism underlying the possible differences between ESP1 and ESP2, it is important to have both the curve predicted by model J1 and the one predicted by model J2. This would also help the comparison of ESP2s among different patients.

5.4.1 Comparing ESPs for $S(4 \mid 2)$

The probability of surviving year 4, given survival up to year 2, is computed only for those patients who were still in the study at year 2. These patients were 278 (89%) of the 312 study participants. Between year 2 and year 4, 53 patients dropped out from the study: 42 actually died whereas 11 were censored.

ESP1 vs ESP2

The left plot of Figure 5.2 shows the scatter plot of the ESP1 (x-axis) and ESP2 (y-axis). We see that the disagreement is more evident for those patients with a low ESP value. For these patients, model J1 tends to be more pessimist than model J2. The departure from the 1:1 line is due to the 43 patients with difference between ESP1 and ESP2 lower than -0.10. These

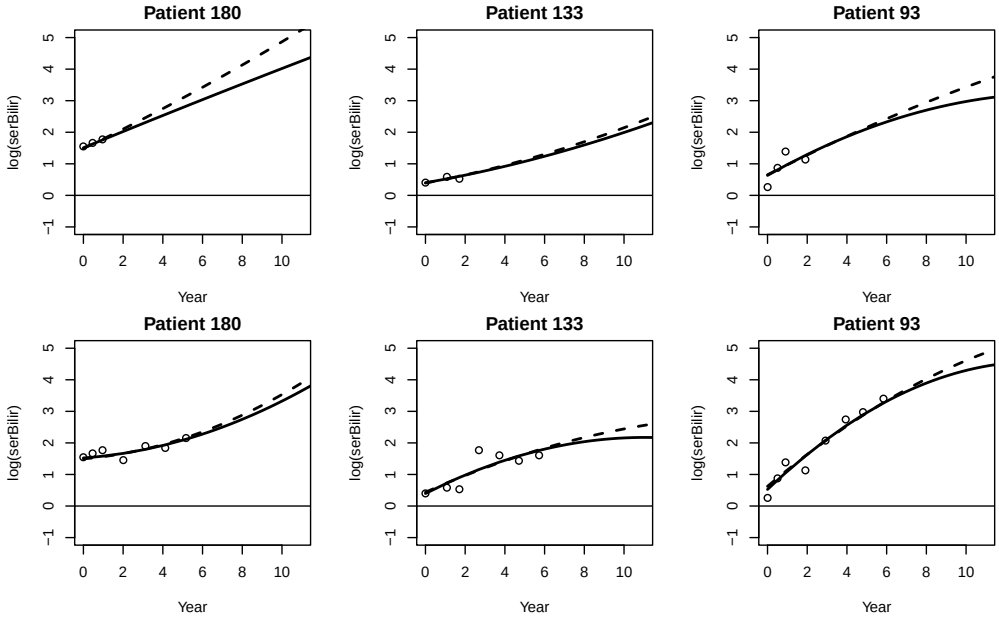


Figure 5.1: Observed log($serBilir$) values (black circles) and longitudinal trajectories predicted by model J1 (solid line) and model J2 (dashed line). The trajectories of the first row are based on the measurements collected in the first 2 years of study whereas those in the second row on the measurements collected in the first 6 years.

patients are usually characterized by high serum bilirubin levels at baseline and low slopes. Thus, for model J1, these patients are high risk patients while for model J2 are low risk patients. A clear example is given in Figure 5.3 and Table 5.4. Patient 130 presents high *covariate* levels from the beginning of the study whereas patients 205 and 51 have much lower initial values. The slopes of patients 130 and 205 are slightly increasing over time whereas the slope of patient 50 is markedly decreasing. Since model J2 considers only the trajectory's slope, the ESP2 of patient 130 is very high in comparison to the corresponding ESP1. On the other hand, model J1 and model J2 agree on patient 205: the covariate initial levels are above 1 and the slope is not ignorable so that both models predict a slight worsening of patient's health conditions between year 2 and year 4. Finally, patient 50 is characterized by

rather low initial *covariate* values and this strongly affects model J1's judgement: there is a 90% chance of surviving up to year 4, given that patient 50 survived year 2.

If ESP1 seems not to be so influenced by the rapid increment of the *covariate* over time, ESP2 does not ignore it. In this case, model J2 is more pessimist than model J1 because of the high, although decreasing, slope. The histogram in Figure 1 shows that most of the differences (66.5%) between ESP1 and ESP2 are contained in the interval $[0, 1)$ and the 88.6% of them are smaller than 0.05. The scatter plot nearby displays a tenuous curvature of the loess curve towards positive values for ESPs closed to 1. These little positive differences belong to patients with low covariate values, mostly below 1, and a weak increasing trend. This last feature causes model J2 to be a little more pessimist than model J1.

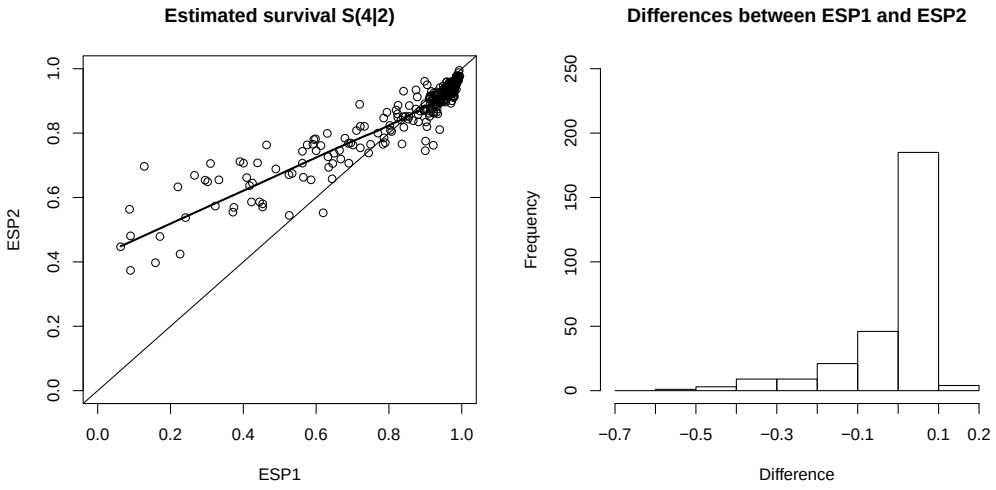


Figure 5.2: Scatter plot of ESP1 and ESP2 and histogram of ESP1-ESP2

Subject	ESP1	ESP2	ESP1-ESP2
130	0.31	0.70	0.39
205	0.83	0.84	0.01
51	0.90	0.76	0.14

Table 5.4: Comparison of ESP1 and ESP2 for three patients

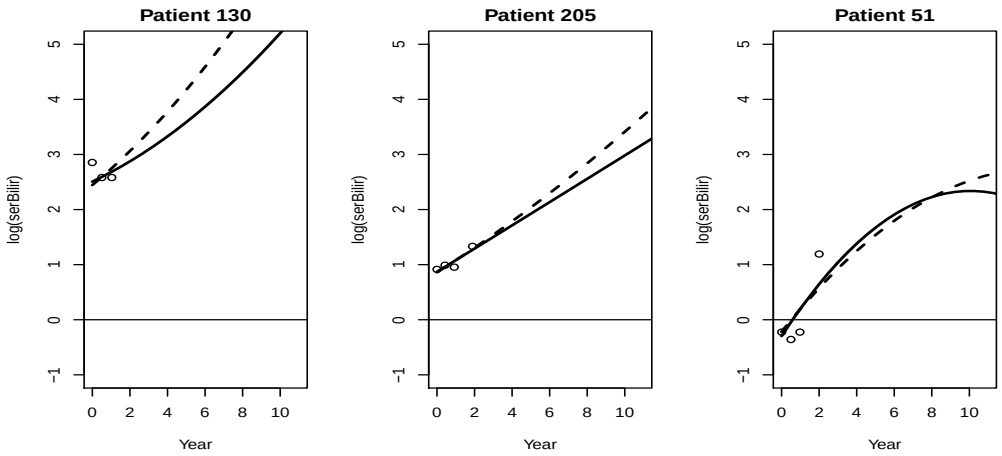


Figure 5.3: Longitudinal trajectories predicted by model J1 (solid line) and model J2 (dashed line), and observed log($serBilir$) (black circles)

ESP1 vs ESP3

The 95% of the differences between ESP1 and ESP3 belong to the interval $[-0.10, 0.10)$ meaning that model J1 and model J3 produce very close survival predictions (see Figure 5.4). Patients 156, 260 and 90 can help us to understand what happens when the information from the slope is added to that from the current value. The results are presented in Table 5.5 and Figure 5.5. The *covariate* trajectory of patient 156 is characterized by *covariate* values which are already high at the beginning of the study and slightly increase over time. The very optimistic prediction made by model J2 is corrected by model J3. Patient 260 presents a better health condition in comparison to patient 156, however, the reasoning is the same. On the contrary, model J3 is more pessimist about the future of patient 90 than are models J1 and J2. Model J3, in fact, combines the effect of the *covariate* value, which becomes higher than 2 between year 2 and 4, with that of a slope which, although decreasing, is still highly positive.

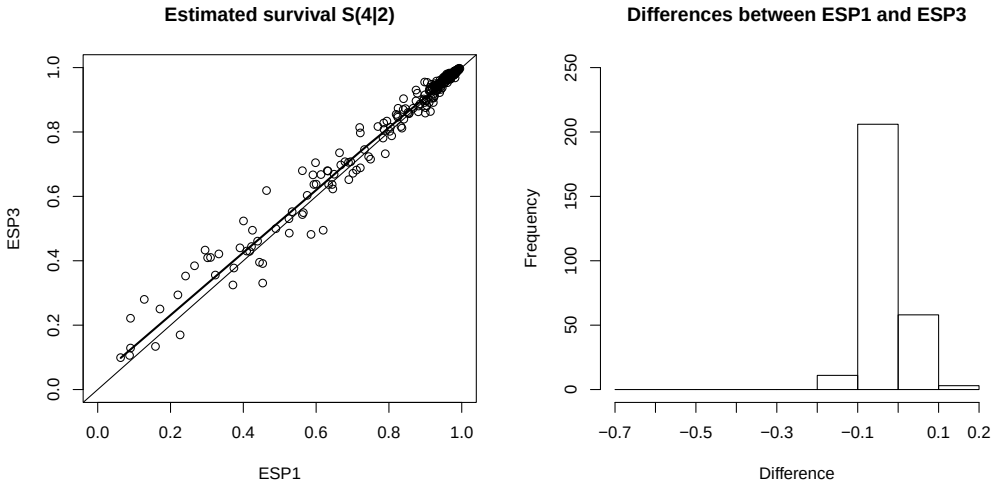


Figure 5.4: Scatter plot of ESP1 and ESP3 and histogram of ESP1-ESP3

Subject	ESP1	ESP2	ESP3	ESP1-ESP3
156	0.13	0.70	0.28	-0.15
260	0.60	0.75	0.64	-0.04
90	0.62	0.55	0.49	0.13

Table 5.5: Comparison of ESP1, ESP2 and ESP3 for three patients

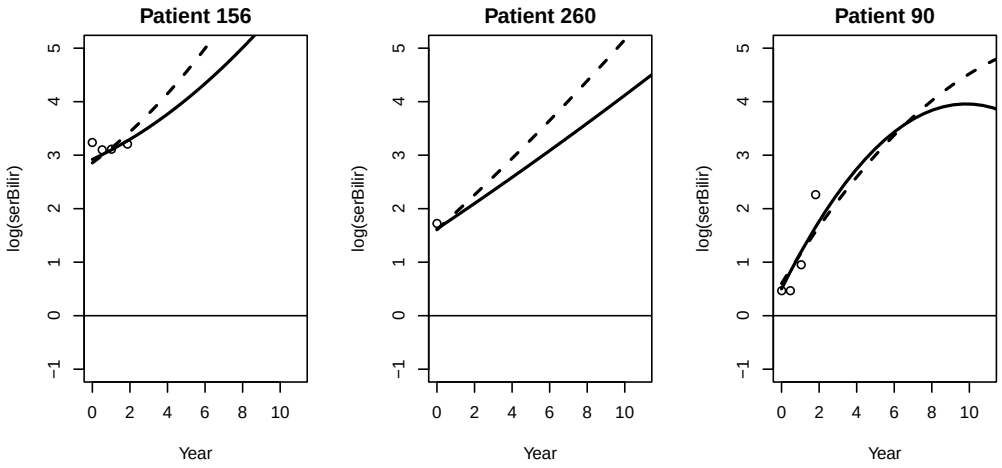


Figure 5.5: Longitudinal trajectories predicted by model J1 (solid line) and model J3 (dashed line), and observed log($serBilir$) (black circles)

ESP1 vs ESP4

74% of the differences between ESP1 and ESP4 belong to the interval $[0, 0.10)$ while the remaining 26% is below zero. Again, the more evident differences belong to patients with the lowest survival probabilities, but, as it can be seen in the scatter plot of Figure 5.6, the dispersion around the 1:1 is rather small. In this case, two may be the reasons behind this behaviour. First, all the integrals start from zero independently of the *covariate* trajectory's features and the integral tends to homogenize the trajectory differences, especially in the very first years of study when trajectories are more similar. Second, the value of association parameter α_4 is not that high ($\alpha_4 = 0.181$) so that a substantial increase of the integral value is needed in order to significantly augment the value of the risk function. For example, in order to obtain the same risk's increment given by one unit increase of the *covariate* value, the integral increment should be equal to 7.141, which is a lot for the types of longitudinal trajectories considered here. The consequences are not univocal for all patients. More specifically, they can be divided into two groups: to the first group belong 207 patients with a EPS1 higher than or equal to 0.794, all the other 71 patients belong to the second group, where ESP1 is lower than 0.794. The first group generates the tenuous deviation on the right of the loess curve from the 1:1 line in the scatter plot of Figure 5.6. Similarly, the negative differences of the second group generate the prominent deviation towards the left and this deviation increases as ESP1 decreases. The maximum difference between ESP1 and ESP4 is equal to -0.489.

We use patients 25, 54, and 80 as examples (see Table 5.6 and Figure 5.7). ESP1 is slightly more optimist than model J4. This is probably due to the fact that the baseline hazard function $h_0(t)$ of model J4 is slightly higher than that of model J1 and to the small value of the parameter α_4 which makes the negative integral to be less 'protective' than the negative *covariate* value. In patient 54, a consistent increasing integral value and a *covariate* value which starts around 0.5 and do not reach 2 within year 4 makes the two ESPs to practically agree. On the contrary, the rather high

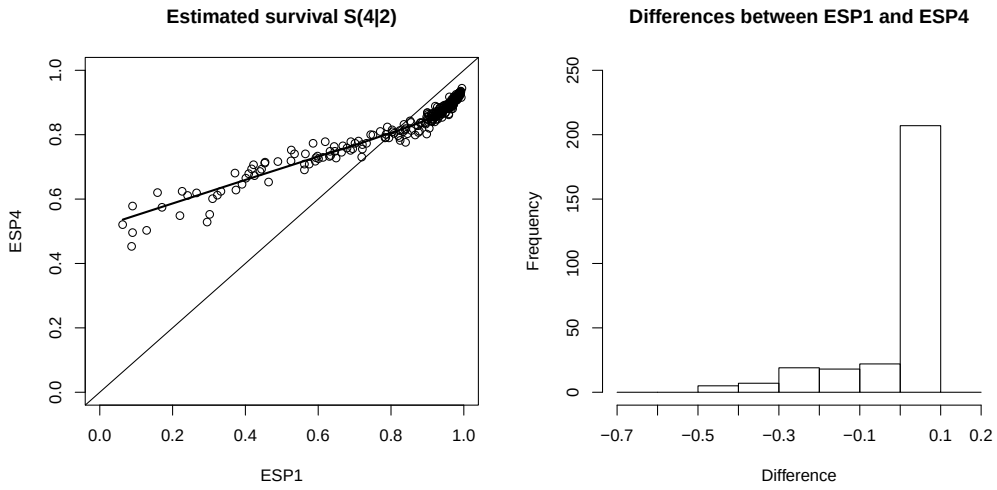


Figure 5.6: Scatter plot of ESP1 and ESP4 and histogram of ESP1-ESP4

initial *covariate* value of patient 80 and its increment over time makes ESP1 decrease more quickly than does the corresponding increment of the integral with respect to ESP4.

Subject	ESP1	ESP4	ESP1-ESP4
25	0.98	0.92	0.05
54	0.84	0.83	0.01
80	0.37	0.63	-0.26

Table 5.6: Comparison of ESP1 and ESP4 for three patients

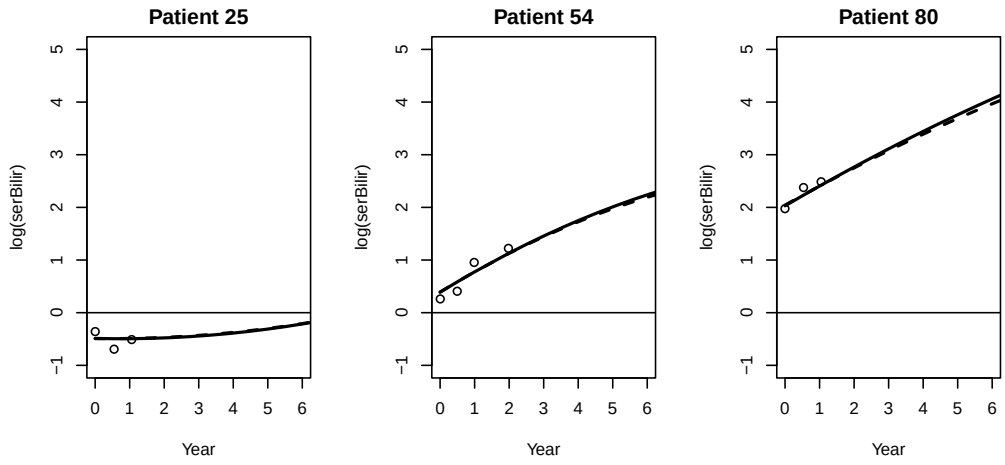


Figure 5.7: Longitudinal trajectories predicted by model J1 (solid line) and model J4 (dashed line), and observed log($serBilir$) (black circles)

ESP1 vs ESP5

Maybe surprisingly, using a weighted cumulative effect greatly reduces the observed gap between ESP1 and ESP4. All the differences belong to the interval $[-0.10, 0.10]$ (see Figure 5.8). Evidently, model J1 and model J5, although using a different version of the information contained in the *covariate*, reach very close conclusions.

The weighted integral has a much greater influence on the risk function of model J5 that does the normal integral on that of model J4: while α_4 is equal to 0.184, α_5 is equal to 2.435. This means that a one unit increase of the normal integral is associated with a $\exp(0.184) = 1.20$-fold increase in the patient's risk, whereas a one unit increase of the weighted integral produces a $\exp(2.435) = 11.42$-fold increase in the patient's risk.

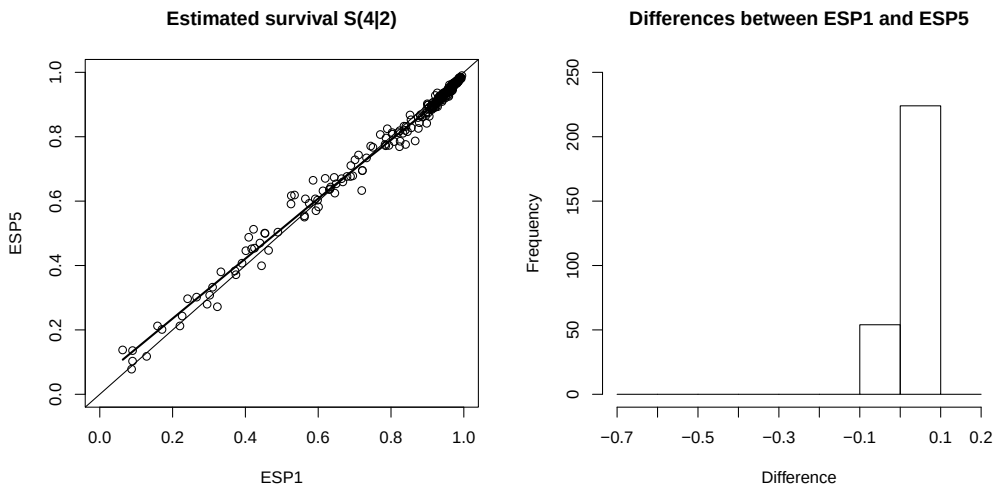


Figure 5.8: Scatter plot of ESP1 and ESP5 and histogram of ESP1-ESP5

ESP1 vs ESP6

253 (91%) of the differences between ESP1 and ESP6 belong to the interval $[-0.10, 0.10)$ and 226 (81.3%) are contained in the interval $[-0.05, 0.05)$. Only 25 (9%) patients have an absolute difference higher than 0.10. This time, there is not a very evident pattern that could explain a negative instead of a positive difference. We see that model J1 tends to be slightly

more optimist than model J2 with low risk patients (Figure 5.9). In fact, considering the 188 individuals with a difference in the interval [0, 0.05), 165 (87.8%) of them have a ESP1 higher than 0.90, 13 (6.9%) have a ESP1 between 0.80 and 0.90 and the other 10 (5.3%) have a ESP1 lower than 0.80. In general, differences belonging to the interval [-0.10, 0.10] can be ascribable to the fact that the quantities ESP1 and ESP6 are based on models using slightly different baseline hazard functions (the estimated baseline hazard of model J6 is higher than that of model J1) and longitudinal models, and only marginally to trajectories' characteristics which are practically constant in the first four years of study.

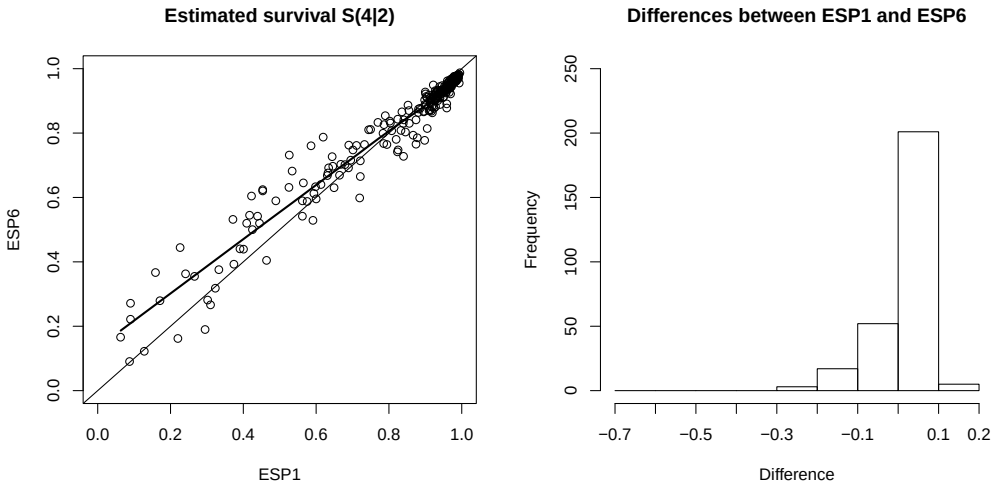


Figure 5.9: Scatter plot of ESP1 and ESP6 and histogram of ESP1-ESP6

For those differences not belonging to that interval, we notice a possible stronger influence of the different use of the *covariate* information made by the two models. For example, patients 254 and 269 have the biggest positive and negative differences (Figure 5.10 and Table 5.7). Their longitudinal trajectories are characterized by moderate initial values of the *covariate* and quite high slopes. Since, with high and increasing *covariate* values, the risk increases more quickly and model J6 considers lagged values, ESP1 is lower than ESP2. On the contrary, patients 93 and 130 do present a *covariate*

increment which is smaller than that of the previous two patients so that both models predict similar ESPs.

Subject	ESP1	ESP6	ESP1-ESP6
254	0.23	0.44	-0.21
269	0.16	0.37	-0.21
93	0.80	0.84	-0.84
130	0.31	0.27	0.04

Table 5.7: Comparison of ESP1 and ESP6 for four patients

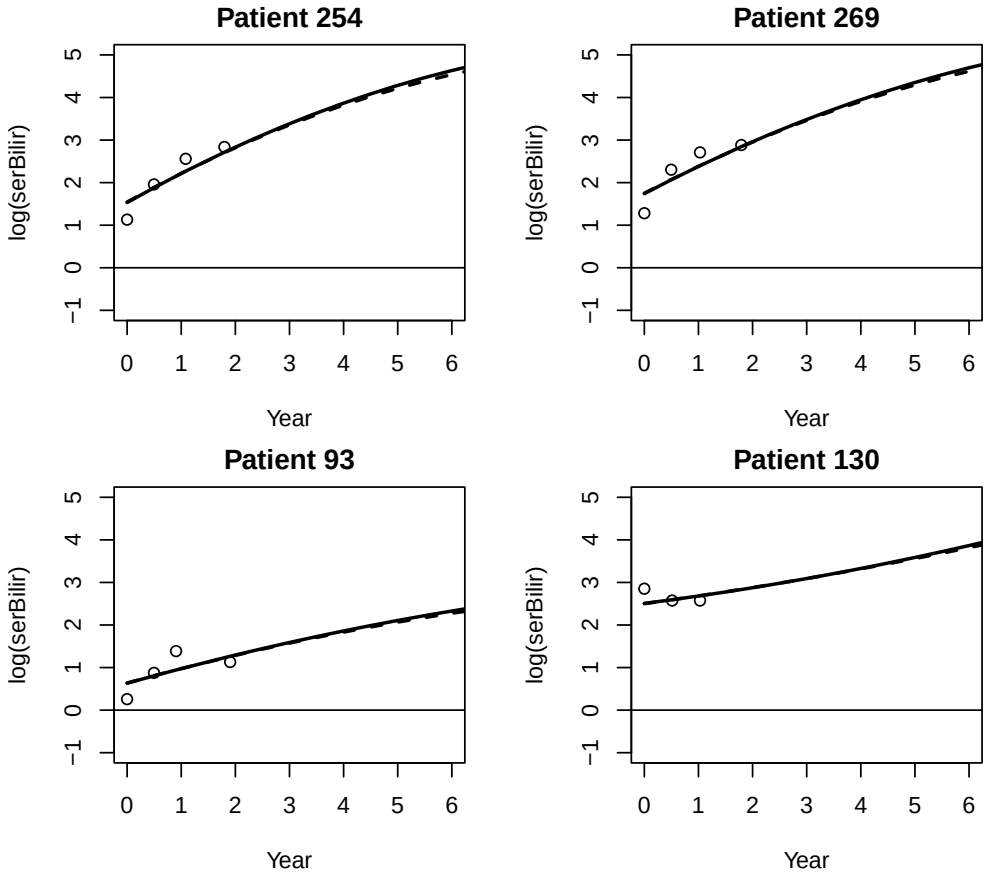


Figure 5.10: Longitudinal trajectories predicted by model J1 (solid line) and model J4 (dashed line), and observed log($serBilir$) (black circles)

ESP1 vs ESP7

225 (80.9%) are the patients with a difference between ESP1 and ESP7 in the interval [-0.10, 0.10) whereas 43 (15.5%) patients have differences belonging to the interval (-0.70, -0.20). These patients are the same individuals which were pointed out in the comparison between the ESPs of model J1 and model J2: their trajectories are mostly characterized by high *covariate* initial values and slopes, which are constant or slightly decreasing. However, the differences observed in this case tend to be more accentuated. In fact, comparing the scatter plots in Figure 5.2 and Figure 5.11 we can see that the loess function in the latter plot tends to stay higher.

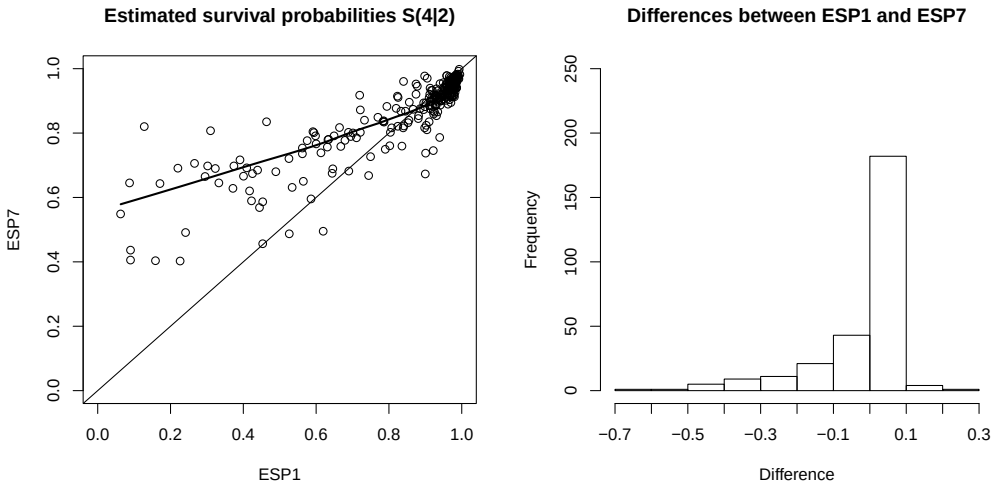


Figure 5.11: Scatter plot of ESP1 and ESP7 and histogram of ESP1-ESP7

As we have noticed before, the information given by the slope, and in this case the lagged slope, can result in quite different expected survival probabilities with respect to the information given by the current value. Table 5.8 and Figure 5.12 give a further example. According to model J7, patients 156, 130, 241, and 214 have practically the same probability of surviving year 4 given that they were alive at year 2: ESP7 is equal to 0.82 for patients 156 and 241, while it is equal to 0.81 for patients 130 and 214. However, model J1 rather disagrees: the ESP1s are equal to 0.13, 0.66, 0.31 and 0.90,

respectively. The reason is made clearer looking at the *covariate* trajectories in Figure 5.12 fitted according to model J7. Indeed, patients 156, 241, and 214 have the same (lagged) slope in practice, and this explains the practically equal ESP7s. On the contrary, their trajectories are quite different in the starting values and this strongly affects the ESPs of model J1: since patient 156 has the highest starting value, his/her survival probability is very low, whereas patient 214, with a starting value close to zero, has the highest survival probability.

Subject	ESP1	ESP7	ESP1-ESP7
156	0.13	0.82	-0.69
130	0.31	0.81	-0.50
241	0.66	0.82	-0.16
214	0.90	0.81	0.09

Table 5.8: Comparison of ESP1 and ESP7 for four patients

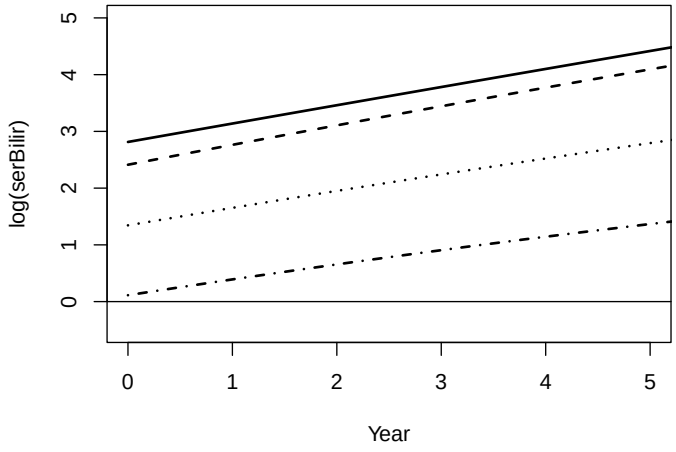


Figure 5.12: Longitudinal trajectories predicted by model J7 of patients 156 (solid line), 130 (dashed line), 241 (dotted line), and 214 (dashed and dotted line)

ESP1 vs ESP8

Model J8 uses the information from both the lagged value and the lagged slope so that the discrepancies seen in the previous comparison are greatly reduced (see Figure 5.13). 262 (94.2%) differences are in the interval [-0.10, 0.10), with 241 (86.7%) being in the interval [-0.05, 0.05). The 16 (5.8%) absolute differences which are higher than 0.10, belong to the same high risk patients pointed out in the previous comparisons. The information brought into the model by the lagged covariate value makes model J8 to be less optimist than model J7 for those patients.

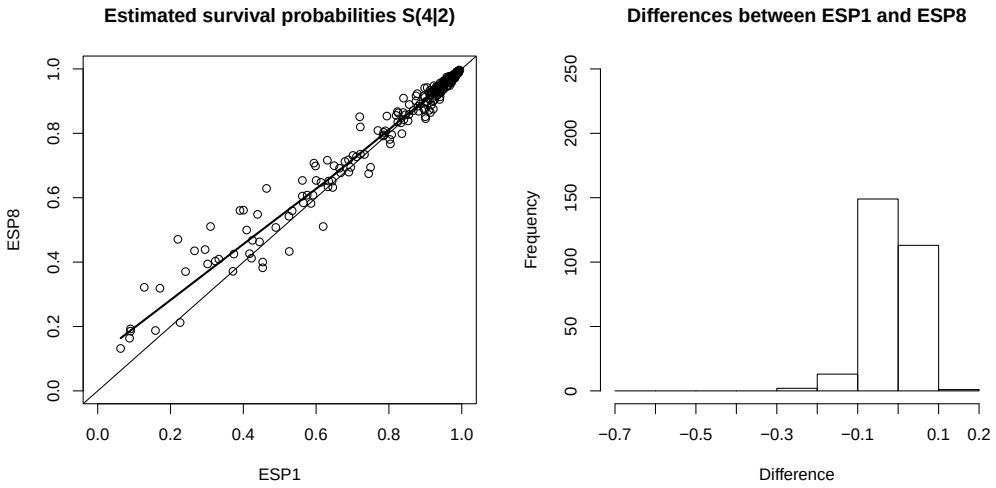


Figure 5.13: Scatter plot of ESP1 and ESP8 and histogram of their difference

Table 5.9 reports the ESPs of model J1 and model J8 for the four patients already considered in Table 5.8. In particular, the difference between ESP1 and ESP8 is about -0.20 for both patients 156 and 130, whereas the difference between ESP1 and ESP7 was -0.69 for patient 156 and -0.50 for patient 130.

Subject	ESP1	ESP8	ESP1-ESP8
156	0.13	0.32	-0.19
130	0.31	0.51	-0.20
241	0.66	0.69	-0.03
214	0.90	0.89	0.01

Table 5.9: Comparison of ESP1 and ESP8 for four patients

5.4.2 The time interval effect on ESPs differences

In this part of the analysis, we wish to monitor the changes in the ESP differences when Δt increases. This time, we are going to use only those patients who survived at least 6 years. In fact, for these patients the number of serum bilirubin measurements used to estimate the longitudinal part of the model is higher so that the longitudinal trajectories will be closer to the 'true' final trajectories, which are those obtained using all the measurement values actually available at the end of the study. For example, in Figure 5.14, the longitudinal trajectories of patients 96 and 142 are shown. The trajectories of the first and the second row are based on the measurements available at year 2 and 6, respectively. Since the measurements of patient 96 are nearly constant between year 2 and year 6, the trajectories are practically the same. On the contrary, patient 142 presents an increasing trend which makes his/her trajectory at year 6 steeper than that at year 2. The effect on the predicted trajectory when new measurent are available can be also seen in Figure 5.1.

The number of patients involved in this analysis are 166 (53%). It should always be kept in mind that these patients are a selected group of all the 312 patient participating to the study. More specifically, the major part of those patients who were already at high risk at year 2 (according to model J1) did not survived until year 6. Consequently, most of the 166 patients considered now had a low or moderate risk of death at year 2. Their risk level at year 6 depends on the serum bilirubin evolution between year 2 and year 6.

In the following part, we will show how differences between the ESPs predicted by model J1 and the other models as the time interval Δt increases. For each comparison, we will report a figure containing the ESPs' scatter plots for $S(7 \mid 6)$, $S(8 \mid 6)$ and $S(10 \mid 6)$, a table containing the ESPs for $S(7 \mid 6)$, $S(8 \mid 6)$ and $S(10 \mid 6)$ and the respective differences, and a figure which will represent these differences. In order to give more specific examples, we will use 20 patients whose *covariate* trajectories are considered 'representative' of the different trajectory types characterizing the 166 pa-

tients involved in the analysis. The identification numbers of the selected patients are: 135, 218, 19, 96, 34, 48, 136, 99, 258, 166, 39, 161, 133, 85, 180, 104, 46, 142, 51, 93. Patients are listed according to their ESP1 value for $S(7 \mid 6)$ in a descending order. Figures from 5.15 to 5.19 contain the predicted trajectories of the 20 selected patients fitted according to model J1 (solid line) and model J2 (dashed line).

For practical reasons, the differences between ESP1 and ESPj, $j = 2, ..., 8$, will be indicated with DIFF1j. Moreover, in general, when we will describe DIFF1j, we will mean its absolute value. For example, if we say

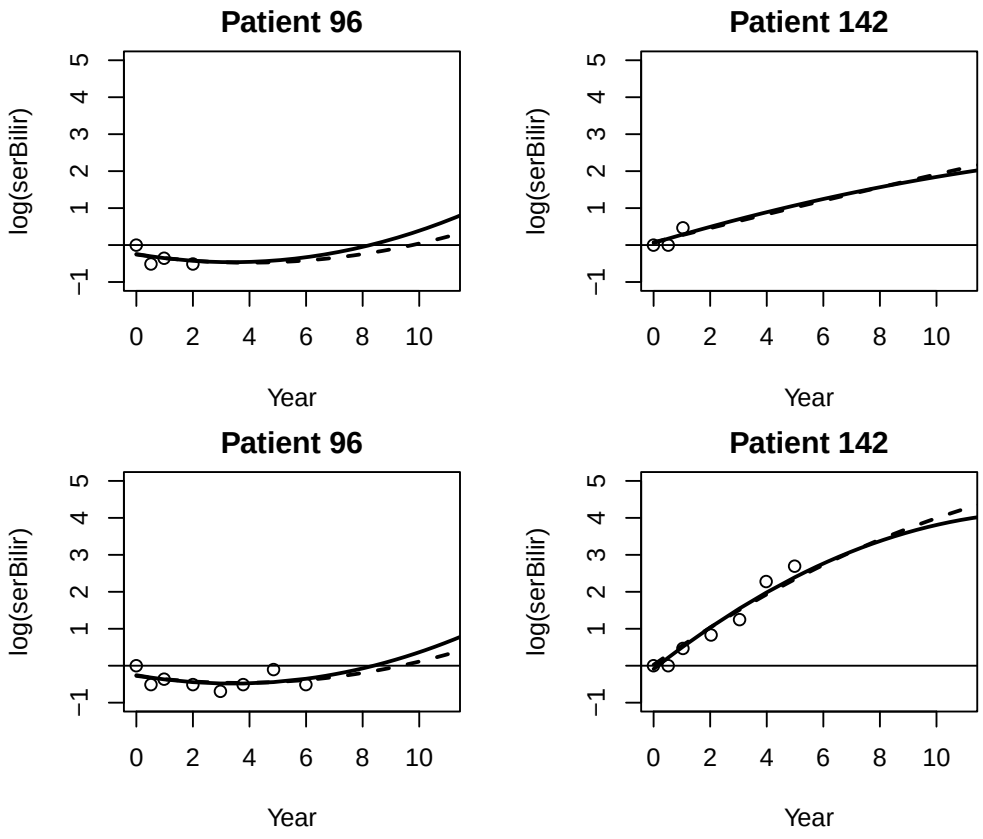


Figure 5.14: Longitudinal trajectories predicted by model J1 (solid line) and model J2 (dashed line) at year 2 (first row). Longitudinal trajectories predicted by model J1 (solid line) and model J2 (dashed line) at year 6 (second row)

that DIFF1j is increasing over time it could either be that DIFF1j was -0.05 and becomes -0.20 or that diff1j was 0.05 and becomes 0.20. The context or explicit indications will clarify in which case we are.

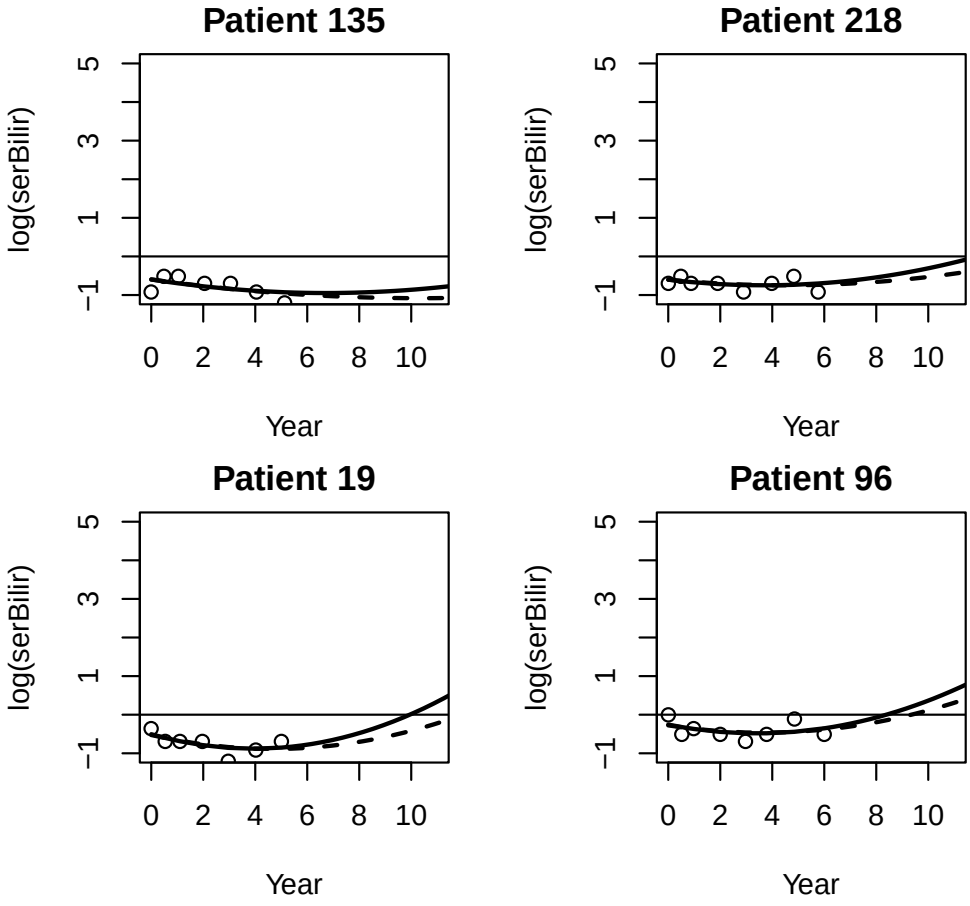


Figure 5.15: Longitudinal trajectories predicted by model J1 (solid line) and model J2 (dashed line), and observed log($serBilir$) (black circles)

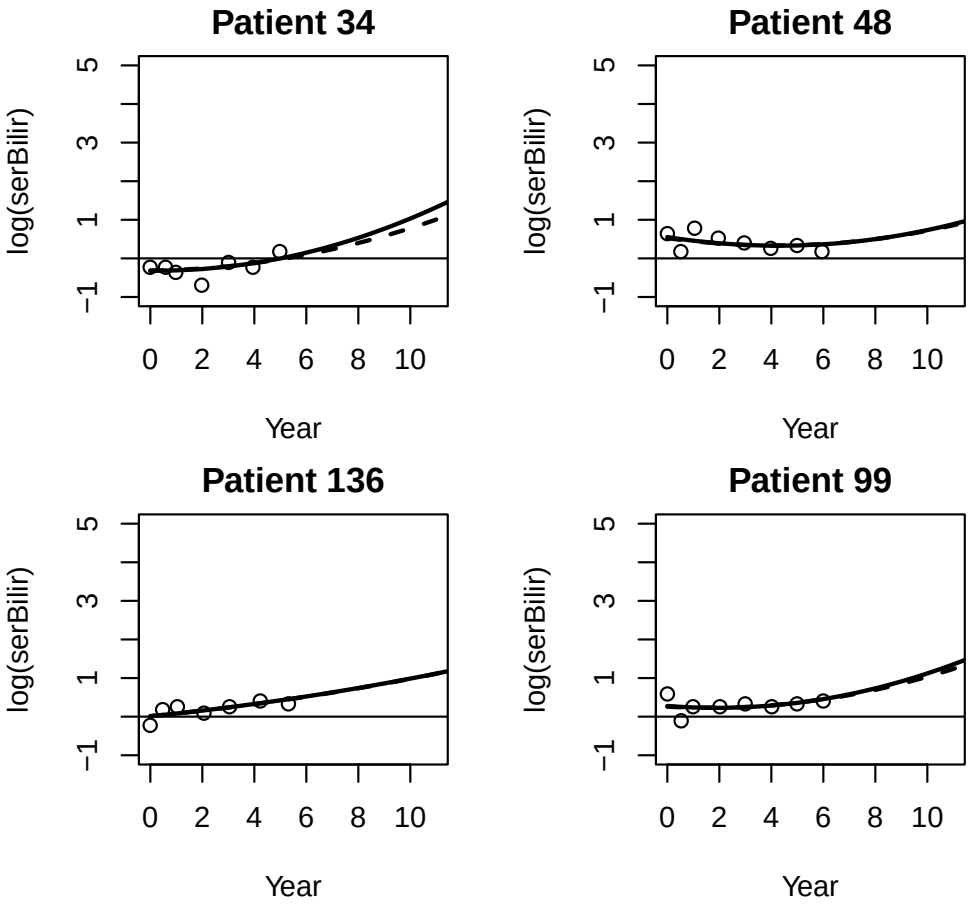


Figure 5.16: Longitudinal trajectories predicted by model J1 (solid line) and model J2 (dashed line), and observed log($serBilir$) (black circles)

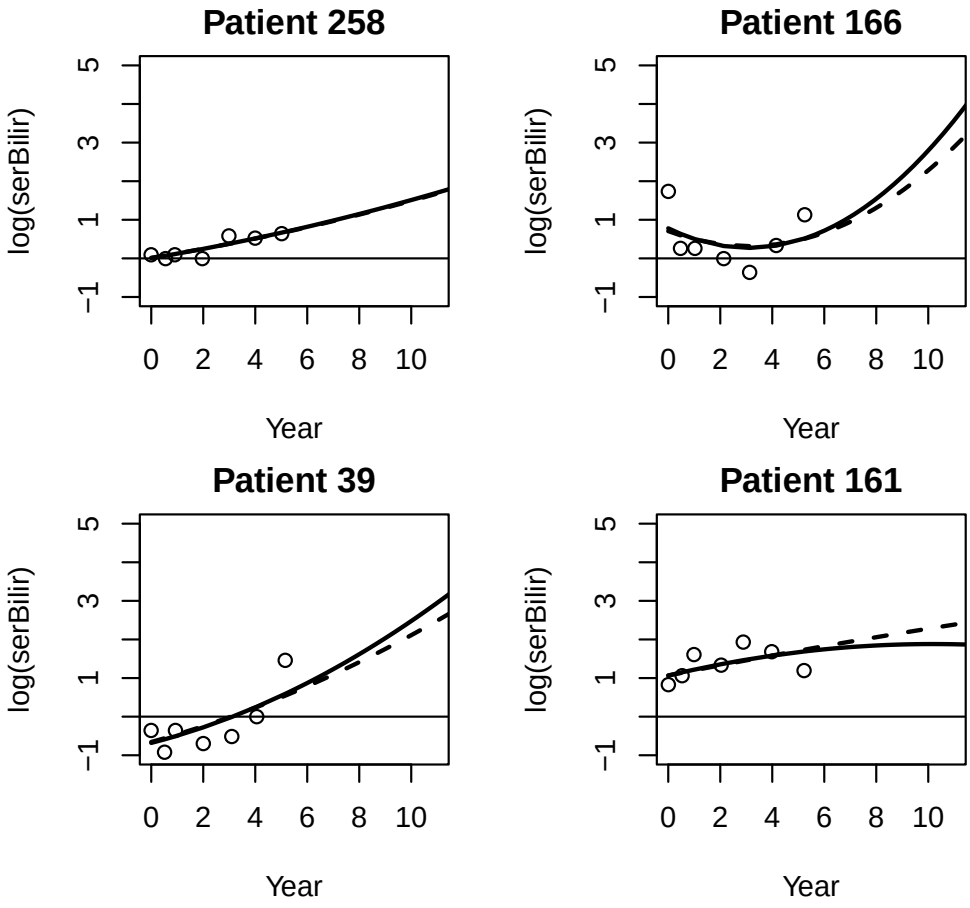


Figure 5.17: Longitudinal trajectories predicted by model J1 (solid line) and model J2 (dashed line), and observed log($serBilir$) (black circles)

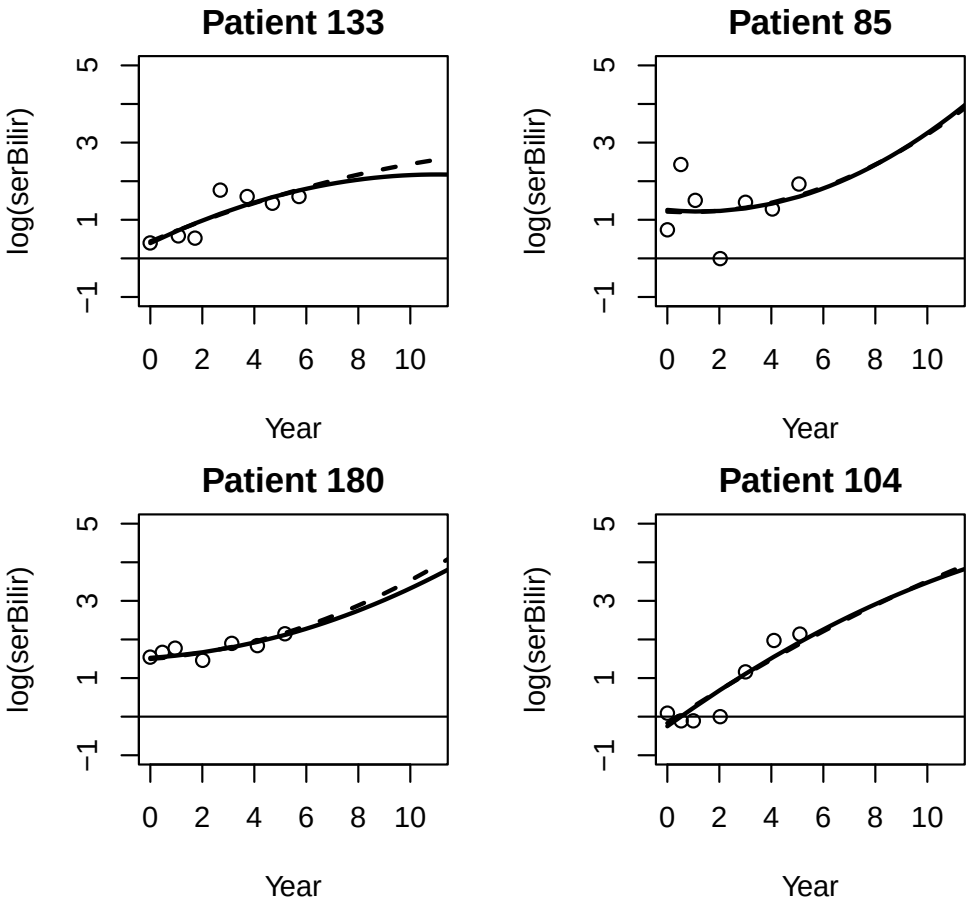


Figure 5.18: Longitudinal trajectories predicted by model J1 (solid line) and model J2 (dashed line), and observed log($serBilir$) (black circles)

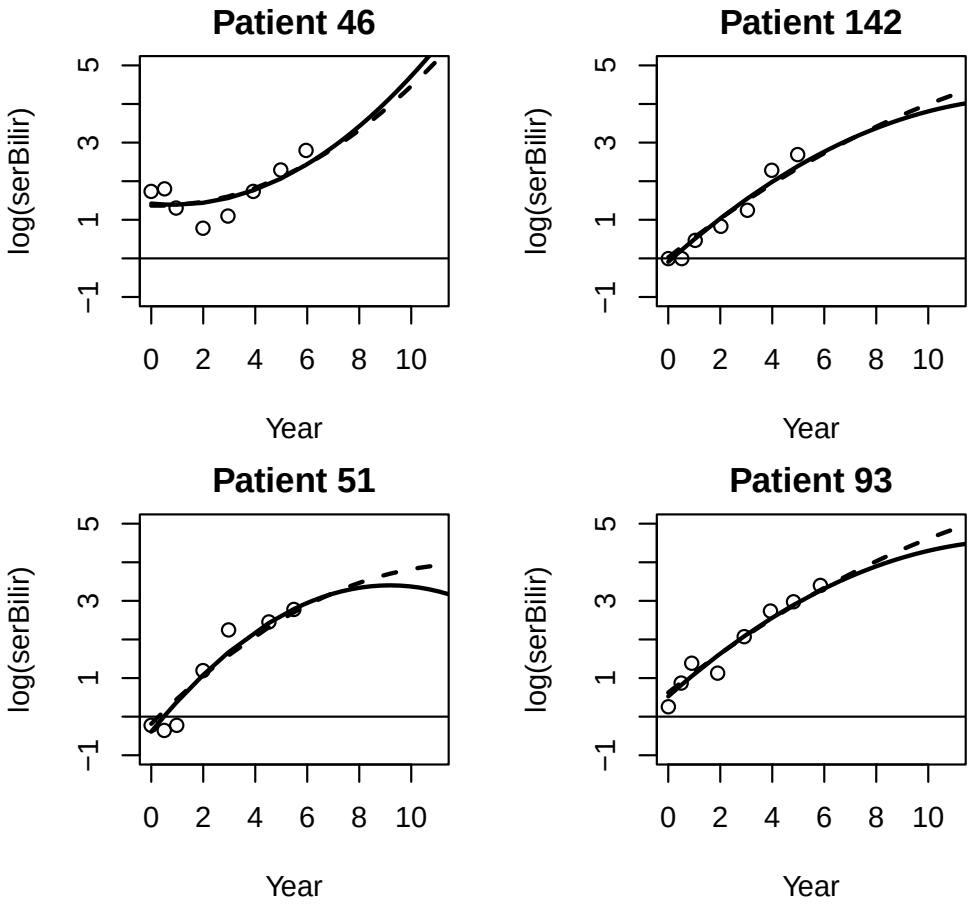


Figure 5.19: Longitudinal trajectories predicted by model J1 (solid line) and model J2 (dashed line), and observed log($serBilir$) (black circles)

ESP1 vs ESP2

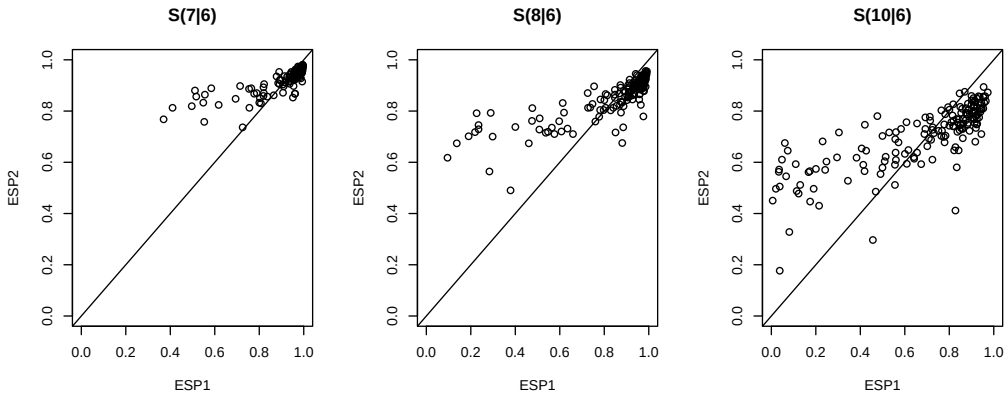


Figure 5.20: Scatter plots of ESP1 and ESP2 for $S(7 \mid 6)$, $S(8 \mid 6)$, and $S(10 \mid 6)$

As it can be seen in Figure 5.20, as Δt increases, the differences between ESP1 and ESP2 tends to increase, however the pattern is not equal for all patients. In fact, Figure 5.21 and Table 5.10 show that model J1 tends to be more optimist than model J2 for patients with the highest $S(7 \mid 6)$ values (i.e. patients on the left of Figure 5.21). On the contrary, model J1 is more pessimist than model J2 for those patients with the lowest $S(7 \mid 6)$ (i.e. patients on the right of Figure 5.21). We give some examples. Patients 19, 96 and 34 have the biggest (positive) differences since their *covariate* values tends to stay below 0 for several years after the study entry, and then start to slowly increase. Model J2 is particularly sensitive to this increment while model J1, since the current *covariate* values are still low, is more positive. Models J1 and J2 substantially agree on the ESPs of patients 99 and 258: their slope is not ignorable and their *covariate* values are above zero. Patients 93, 51, and 142 have the biggest (negative) differences: ESP2 does not decrease as fast as ESP1 so that the DIFF12 increases over time. The fast decrement of ESP1 is attributable to the quite high *covariate* values which characterize patients' trajectories, especially between year 6 and year 10. On the other

hand, model J2 sees a current slope which is decreasing in that interval. On the contrary, patients 46 have a high and increasing current slopes which makes DIFF12 decrease over time.

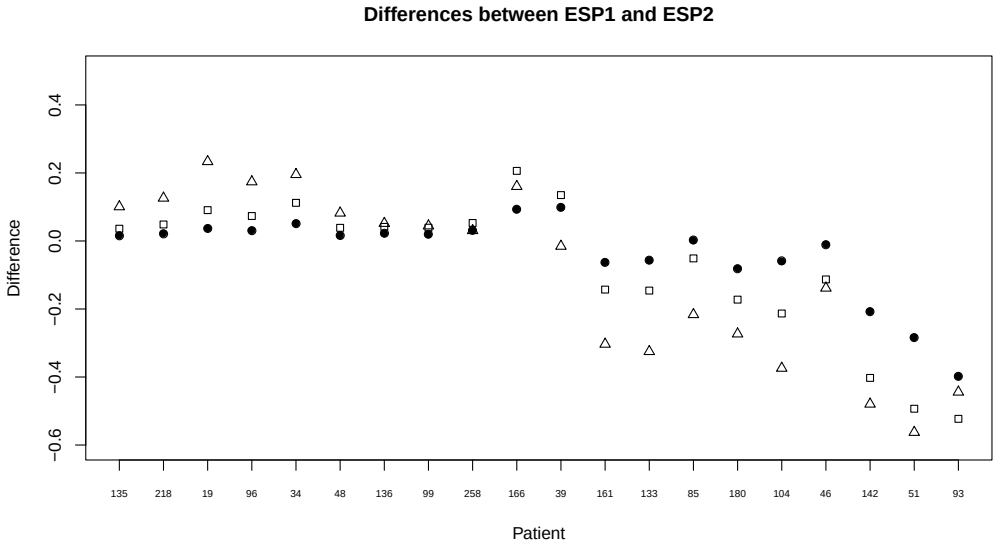


Figure 5.21: DIFF12 for $S(7 \mid 6)$ [circle], $S(8 \mid 6)$ [square], and $S(10 \mid 6)$ [triangle]

Now, we wish to focus our attention on two groups of patients. In the first group we have patients 161, 133, 85, 180, and 104. In the second we have patients 142, 51, 93. We see that DIFF12 at year 8 and DIFF12 at year 10 are more far apart in the first group than in the second (see Figure 5.21). In particular, what happen to patient 93 is emblematic: DIFF12 decreases between year 8 and 10. We call this the *low-ESP effect* and it can be well understood looking at Figure 5.22 and Table 5.10. The ESP1 of patient 93 at year 8 is already very low (i.e. ESP1 = 0.095). The consequence is that ESP1 will not decrease much further during the subsequent 2 years since ESPs have zero as lower boundary. On the contrary, ESP2 at year 8 is equal to 0.618 so that it can significantly decrease, shortening the gap with ESP1. Conversely, patient 104 has a higher ESP1 at year 8 than patient 93 and thus his/her DIFF12 is immune to the *low-ESP effect*.

	$S(7 \mid 6)$			$S(8 \mid 6)$			$S(10 \mid 6)$		
Patient	ESP1	ESP2	DIFF12	ESP1	ESP2	DIFF12	ESP1	ESP2	DIFF12
135	0.996	0.981	0.015	0.991	0.955	0.036	0.973	0.872	0.101
218	0.995	0.974	0.021	0.989	0.940	0.049	0.965	0.839	0.126
19	0.995	0.958	0.037	0.986	0.895	0.091	0.944	0.711	0.233
96	0.992	0.962	0.030	0.980	0.907	0.073	0.925	0.751	0.174
34	0.984	0.933	0.051	0.958	0.846	0.112	0.842	0.647	0.195
48	0.980	0.963	0.017	0.950	0.911	0.039	0.838	0.756	0.082
136	0.975	0.953	0.022	0.938	0.897	0.041	0.809	0.757	0.052
99	0.975	0.955	0.020	0.935	0.896	0.039	0.767	0.722	0.045
258	0.963	0.931	0.032	0.906	0.853	0.053	0.718	0.688	0.030
166	0.962	0.868	0.094	0.881	0.675	0.206	0.457	0.297	0.160
39	0.951	0.853	0.098	0.851	0.716	0.135	0.470	0.485	-0.015
161	0.890	0.952	-0.062	0.754	0.897	-0.143	0.477	0.780	-0.303
133	0.878	0.935	-0.057	0.725	0.871	-0.146	0.421	0.746	-0.325
85	0.866	0.862	0.004	0.659	0.711	-0.052	0.214	0.431	-0.217
180	0.789	0.870	-0.081	0.547	0.720	-0.173	0.174	0.447	-0.273
104	0.755	0.813	-0.058	0.461	0.674	-0.213	0.114	0.488	-0.374
46	0.726	0.737	-0.011	0.378	0.491	-0.113	0.038	0.177	-0.139
142	0.617	0.824	-0.207	0.298	0.700	-0.402	0.066	0.546	-0.480
51	0.548	0.833	-0.285	0.235	0.729	-0.494	0.048	0.611	-0.563
93	0.370	0.768	-0.398	0.095	0.618	-0.523	0.006	0.450	-0.444

Table 5.10: Comparison of ESP1, ESP2, and DIFF12 over time

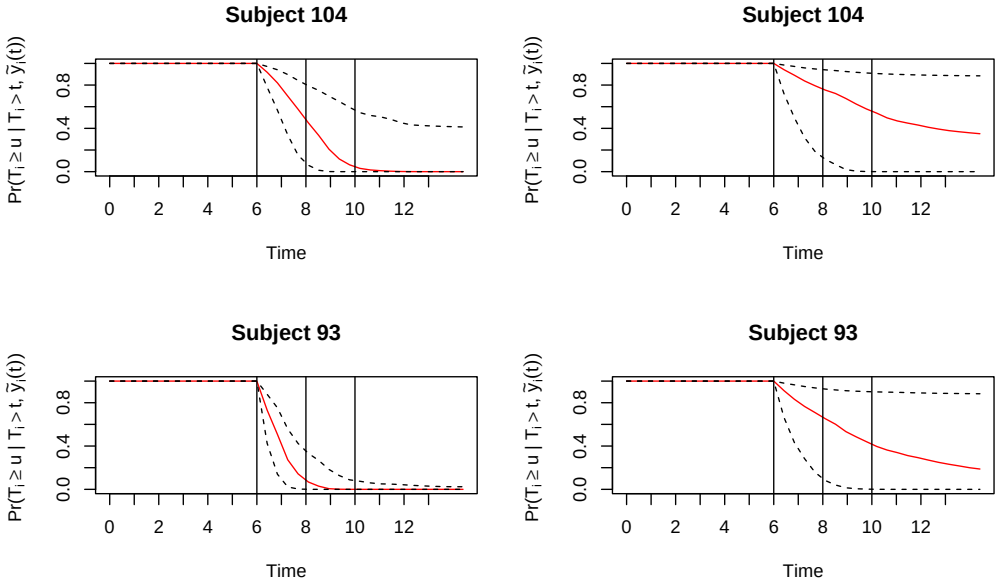


Figure 5.22: Comparison of ESP1 (first column) and ESP2 (second column) of two patients. The solid line depicts the mean of $\pi_i(u \mid t)$ over the Monte Carlo samples. The dashed line depicts a 95% pointwise confidence intervals based on the quantiles of $\pi_i(u \mid t)$ over the Monte Carlo samples.

ESP1 vs ESP3

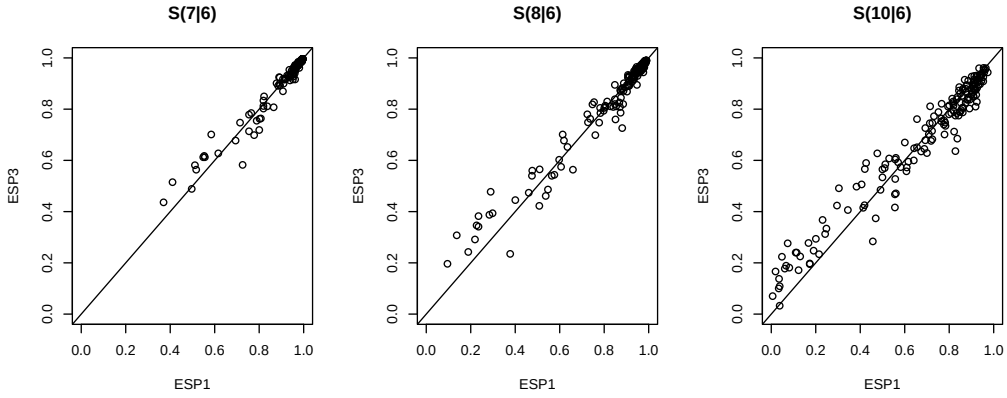


Figure 5.23: Scatter plots of ESP1 and ESP3 for $S(7 \mid 6)$, $S(8 \mid 6)$, and $S(10 \mid 6)$

Model J1 and J3 tend to give closer ESPs than model J2 as we can see from Figure 5.23 and Figure 5.24. In general, from Figure 5.24, we can see that differences start to reach consistent values when $\Delta t = 4$. In fact, when $\Delta = 1$, $\Delta t = 2$, and $\Delta t = 4$ there are , respectively, 3, 9, and 24 differences which are bigger than 0.10 (but smaller than 0.20). The highest negative differences belong to patients with a longitudinal trajectory similar to that of patients 161 and 51: their slopes are small and decreasing between year 6 and 10 so that model J3 tends to be more optimist. The influence of an increasing slope on ESP3 is made quite evident if we consider patient 46 and 166: both their trajectories have a strongly increasing slope after year 4 but patient 46 has much higher initial *covariate* values. First we analyse patient 46. DIFF13 at year 7 is equal to 0.143 so that model J1 is more optimist than model J3 which is influenced by the high value of the current slope. Between year 7 and 8, DIFF13 stays constant in practice but between year 8 and 10 strongly decreases so that, at year 10, it is nearly zero. We remind that in model J1, the current *covariate* value has more weight than in model J3, thus the high *covariate* values between year 8 and 10 allow ESP1 to decrease

more quickly than ESP3 and reach similar values. Now we analyse patient 166. At year 7, ESP3 is only slightly lower than ESP1. However, in this case, ESP3 will decrease more rapidly than ESP1 since patient's *covariate* value exceeds 2 only after year 8.

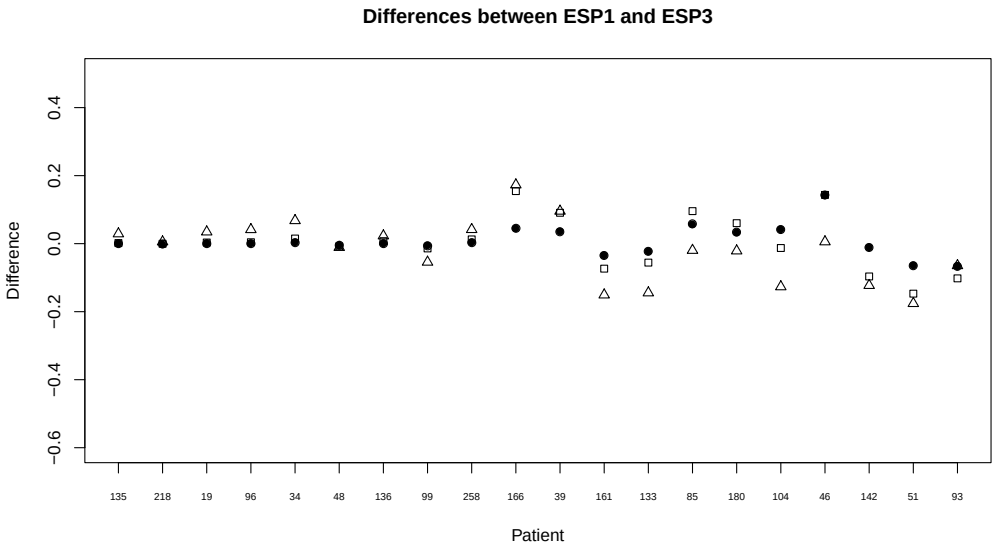


Figure 5.24: DIFF13 for $S(7 \mid 6)$ [circle], $S(8 \mid 6)$ [square], and $S(10 \mid 6)$ [triangle]

	$S(7 \mid 6)$			$S(8 \mid 6)$			$S(10 \mid 6)$		
Patient	ESP1	ESP3	DIFF13	ESP1	ESP3	DIFF13	ESP1	ESP3	DIFF13
135	0.996	0.996	0	0.991	0.989	0.002	0.973	0.944	0.029
218	0.995	0.996	-0.001	0.989	0.989	0	0.965	0.960	0.005
19	0.995	0.994	0.001	0.986	0.983	0.003	0.944	0.910	0.034
96	0.992	0.992	0	0.980	0.976	0.004	0.925	0.884	0.041
34	0.984	0.980	0.004	0.958	0.943	0.015	0.842	0.774	0.068
48	0.980	0.984	-0.004	0.950	0.958	-0.008	0.838	0.849	-0.011
136	0.975	0.975	0	0.938	0.931	0.007	0.809	0.786	0.023
99	0.975	0.981	-0.006	0.935	0.949	-0.014	0.767	0.821	-0.054
258	0.963	0.960	0.003	0.906	0.894	0.012	0.718	0.677	0.041
166	0.962	0.917	0.045	0.881	0.726	0.155	0.457	0.284	0.173
39	0.951	0.916	0.035	0.851	0.760	0.091	0.470	0.374	0.096
161	0.890	0.924	-0.034	0.754	0.828	-0.074	0.477	0.627	-0.15
133	0.878	0.901	-0.023	0.725	0.781	-0.056	0.421	0.566	-0.145
85	0.866	0.807	0.059	0.659	0.564	0.095	0.214	0.234	-0.020
180	0.789	0.755	0.034	0.547	0.487	0.060	0.174	0.195	-0.021
104	0.755	0.713	0.042	0.461	0.474	-0.013	0.114	0.241	-0.127
46	0.726	0.583	0.143	0.378	0.235	0.143	0.038	0.033	0.005
142	0.617	0.628	-0.011	0.298	0.394	-0.096	0.066	0.189	-0.123
51	0.548	0.613	-0.065	0.235	0.382	-0.147	0.048	0.224	-0.176
93	0.370	0.436	-0.066	0.095	0.197	-0.102	0.006	0.071	-0.065

Table 5.11: Comparison of ESP1, ESP3, and DIFF13 over time

ESP1 vs ESP4

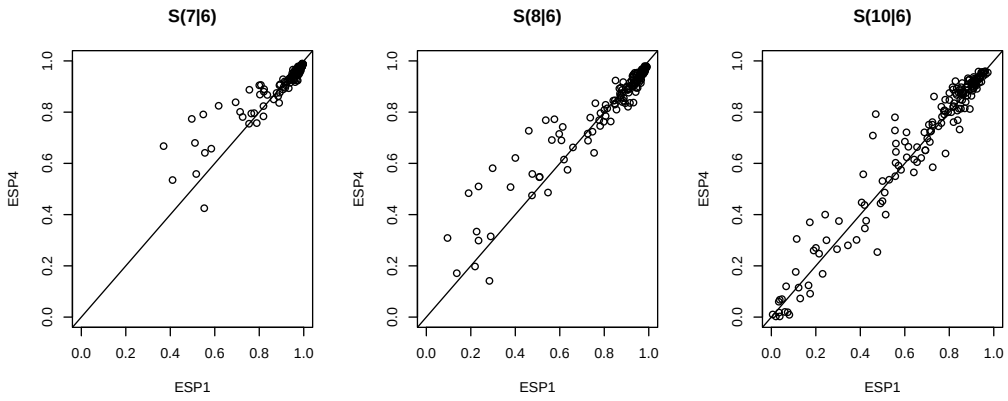


Figure 5.25: Scatter plots of ESP1 and ESP4 for $S(7 \mid 6)$, $S(8 \mid 6)$, and $S(10 \mid 6)$

From Figure 5.25 we can see that, also in this case, the biggest differences belong to medium and high risk patients, with model J1 being, in general, more pessimist than model J4.

Figure 5.26 and Table 5.12, however, show that high differences at year 7 tend to decrease as time passes and to be particularly close to zero at year 10. On the other hand, several differences which were small at year 7, tend to increase as time passes. To the first group belong patients 51, 93, 142, and 104. Their trajectories are characterized by *covariate* values close to zero at the beginning of the study and a rapid increase over time which makes them reach very high *covariate* values, especially between year 6 and 10. A consequence is that the value of the integral increases a lot in four years thus making ESP4 decrease rapidly. To the second group belong patients 39, 166, and 161. Patients 39 and 166 have *covariate* values which stay below one at least until year 6 and then start to rapidly increase. On the contrary, patient 161 has *covariate* values which are slightly above one at the beginning of the study and are not expected to exceed two in the following years. The reduction of the gap between ESP1 and ESP4 among

patients of the first group is attributable to both the rapid increment of the integral after year 6 and to the *low-ESP effect*. Figure 5.27 shows it quite clearly. Indeed, considering patient 104, DIFF16 decreases over time but is still bigger than 0.10 since his/her ESP1 at year 7 and 8 is still high (i.e. ESP1 = 0.755 and ESP1 = 0.461 respectively, see Table 5.12). Thus, ESP1 can significantly decrease and be still lower than ESP6. On the contrary, ESP1 at year 7, but especially at year 8, are already low (i.e. ESP1 = 0.37 and ESP1 = 0.095, respectively) so that it cannot decrease much further. Therefore, ESP6 has the opportunity to reach ESP1. For patients 39 and 166, DIFF14 is increasing over time since their integrals are still small at year 6 and, although increasing, ESP6 does not decrease as fast as ESP1. On the other hand, the reasoning behind patient 161 is slightly different. ESP6 decreases more quickly than ESP1 since integral at year 6 is already consistent and keeps increasing significantly. On the contrary, ESP1 tends to decrease slowly since the *covariate* value is not that high and the increment between year 6 and 10 is truly small.

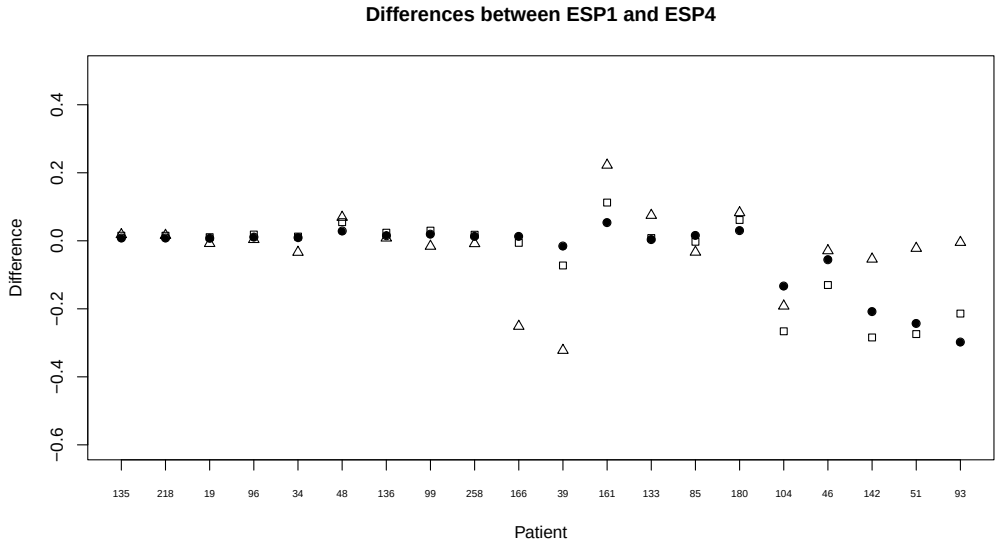


Figure 5.26: DIFF14 for $S(7 \mid 6)$ [circle], $S(8 \mid 6)$ [square], and $S(10 \mid 6)$ [triangle]

	$S(7 \mid 6)$			$S(8 \mid 6)$			$S(10 \mid 6)$		
Patient	ESP1	ESP4	DIFF14	ESP1	ESP4	DIFF14	ESP1	ESP4	DIFF14
135	0.996	0.988	0.008	0.991	0.976	0.015	0.973	0.955	0.018
218	0.995	0.987	0.008	0.989	0.974	0.015	0.965	0.949	0.016
19	0.995	0.987	0.008	0.986	0.975	0.011	0.944	0.952	-0.008
96	0.992	0.981	0.011	0.980	0.961	0.019	0.925	0.921	0.004
34	0.984	0.974	0.010	0.958	0.945	0.013	0.842	0.876	-0.034
48	0.980	0.950	0.030	0.950	0.895	0.055	0.838	0.769	0.069
136	0.975	0.960	0.015	0.938	0.914	0.024	0.809	0.800	0.009
99	0.975	0.956	0.019	0.935	0.905	0.030	0.767	0.783	-0.016
258	0.963	0.950	0.013	0.906	0.888	0.018	0.718	0.727	-0.009
166	0.962	0.949	0.013	0.881	0.887	-0.006	0.457	0.709	-0.252
39	0.951	0.966	-0.015	0.851	0.924	-0.073	0.470	0.792	-0.322
161	0.890	0.837	0.053	0.754	0.642	0.112	0.477	0.254	0.223
133	0.878	0.875	0.003	0.725	0.717	0.008	0.421	0.346	0.075
85	0.866	0.850	0.016	0.659	0.662	-0.003	0.214	0.248	-0.034
180	0.789	0.758	0.031	0.547	0.486	0.061	0.174	0.091	0.083
104	0.755	0.887	-0.132	0.461	0.727	-0.266	0.114	0.306	-0.192
46	0.726	0.781	-0.055	0.378	0.508	-0.130	0.038	0.067	-0.029
142	0.617	0.825	-0.208	0.298	0.582	-0.284	0.066	0.120	-0.054
51	0.548	0.791	-0.243	0.235	0.510	-0.275	0.048	0.070	-0.022
93	0.370	0.668	-0.298	0.095	0.309	-0.214	0.006	0.011	-0.005

Table 5.12: Comparison of ESP1, ESP4, and DIFF14 over time

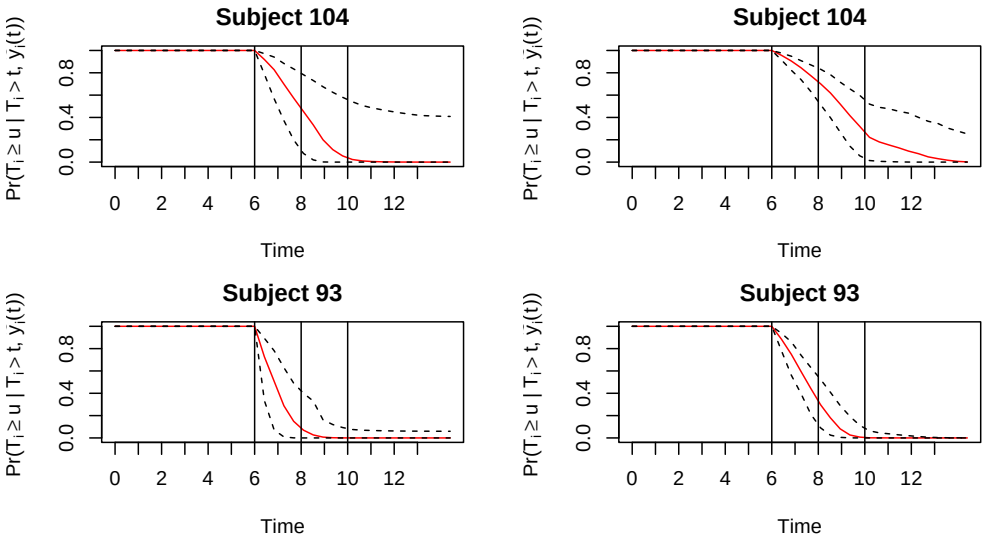


Figure 5.27: Comparison of ESP1 (first column) and ESP4 (second column) of two patients. The solid line depicts the mean of $\pi_i(u \mid t)$ over the Monte Carlo samples. The dashed line depicts a 95% pointwise confidence intervals based on the quantiles of $\pi_i(u \mid t)$ over the Monte Carlo samples.

ESP1 vs ESP5

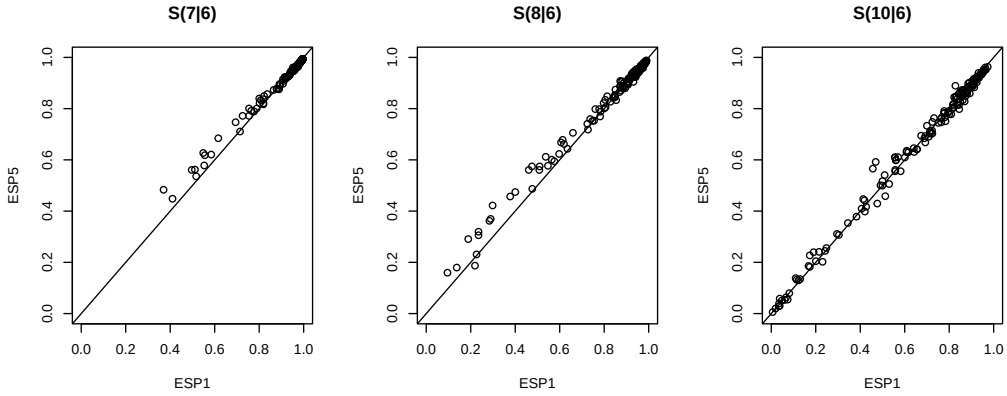


Figure 5.28: Scatter plots of ESP1 and ESP2 for $S(7 \mid 6)$, $S(8 \mid 6)$, and $S(10 \mid 6)$

Figure 5.28 shows that model J1 and model J5, although using the *covariate* information in a different way, they do predict very similar ESP values. We report Figure 5.29 and Table 5.13 for completeness.

	$S(7 \mid 6)$			$S(8 \mid 6)$			$S(10 \mid 6)$		
Patient	ESP1	ESP5	DIFF15	ESP1	ESP5	DIFF15	ESP1	ESP5	DIFF15
135	0.996	0.994	0.002	0.991	0.988	0.003	0.973	0.963	0.010
218	0.995	0.993	0.002	0.989	0.982	0.007	0.965	0.953	0.012
19	0.995	0.993	0.002	0.986	0.984	0.002	0.944	0.937	0.007
96	0.992	0.989	0.003	0.980	0.974	0.006	0.925	0.918	0.007
34	0.984	0.982	0.002	0.958	0.950	0.008	0.842	0.858	-0.016
48	0.980	0.974	0.006	0.950	0.938	0.012	0.838	0.826	0.012
136	0.975	0.969	0.006	0.938	0.928	0.010	0.809	0.778	0.031
99	0.975	0.970	0.005	0.935	0.937	-0.002	0.767	0.766	0.001
258	0.963	0.958	0.005	0.906	0.901	0.005	0.718	0.703	0.015
166	0.962	0.962	0	0.881	0.907	-0.026	0.457	0.566	-0.109
39	0.951	0.961	-0.01	0.851	0.875	-0.024	0.470	0.592	-0.122
161	0.890	0.876	0.014	0.754	0.753	0.001	0.477	0.429	0.048
133	0.878	0.878	0	0.725	0.740	-0.015	0.421	0.399	0.022
85	0.866	0.873	-0.007	0.659	0.705	-0.046	0.214	0.241	-0.027
180	0.789	0.801	-0.012	0.547	0.577	-0.003	0.174	0.182	-0.008
104	0.755	0.801	-0.046	0.461	0.561	-0.100	0.114	0.132	-0.018
46	0.726	0.772	-0.046	0.378	0.457	-0.079	0.038	0.058	-0.02
142	0.617	0.685	-0.068	0.298	0.422	-0.124	0.066	0.064	0.002
51	0.548	0.627	-0.079	0.235	0.320	-0.085	0.048	0.051	-0.003
93	0.370	0.483	-0.113	0.095	0.159	-0.064	0.006	0.007	-0.001

Table 5.13: Comparison of ESP1, ESP5, and DIFF15 over time

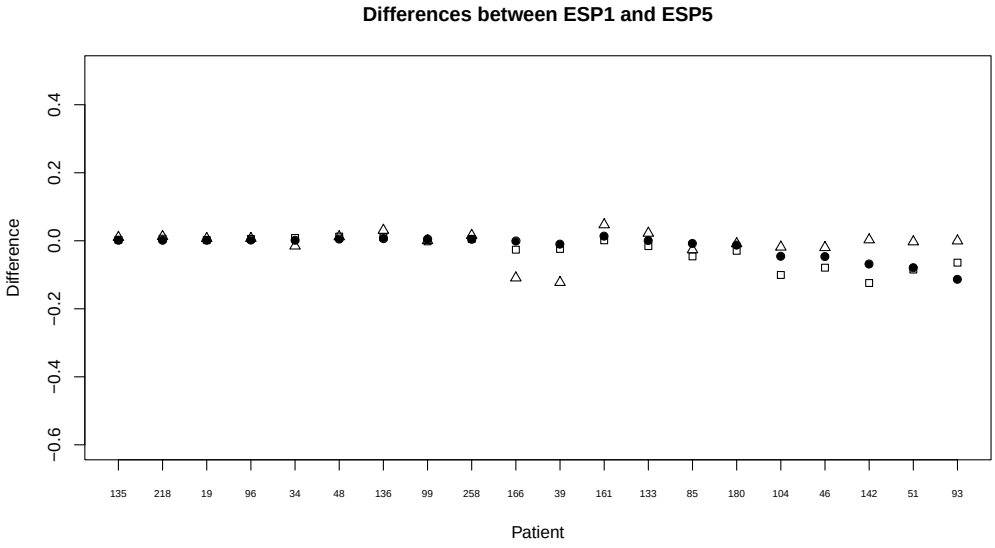


Figure 5.29: DIFF15 for $S(7 \mid 6)$ [circle], $S(8 \mid 6)$ [square], and $S(10 \mid 6)$ [triangle]

ESP1 vs ESP6

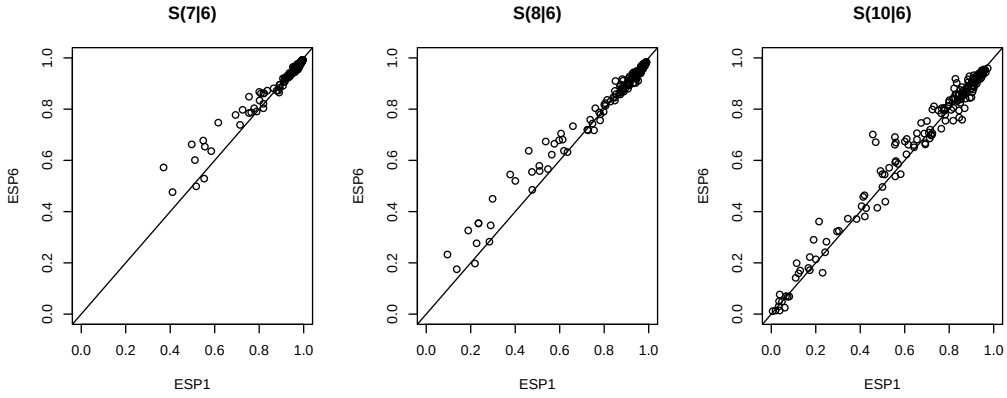


Figure 5.30: Scatter plots of ESP1 and ESP2 for $S(7 \mid 6)$, $S(8 \mid 6)$, and $S(10 \mid 6)$

Looking at Figure 5.30 we see that there are very few differences between ESP1 and ESP6 which are lower than -0.10: we have 4, 9, and 7 patients for $S(7 \mid 6)$, $S(8 \mid 6)$, and $S(10 \mid 6)$, respectively. As happened before, those patients showing high differences at year 7 and 8 do not coincide with those showing high differences at year 10. For example, the four high DIFF16s at year 7 tend to decrease with time, whereas those DIFF16s which which are lower that -0.10 at year 10 where smaller at year 7 and 8. This behaviour is strictly linked to the different longitudinal trajectories. Patients with decreasing DIFF16s have longitudinal trajectories similar to those of patients 93 and 142, which are characterized by high *covariate* values, especially after year 6. Since model J6 considers at year t the *covariate* values at year $t-2$, ESP6 decrease more slowly than ESP1 with this type of trajectories. The decrement of DIFF16 with time is mostly due to the *low-ESP effect*. As exampe, this time we consider patient 142. DIFF17 at year 7, 8, and 10 is equal to -0.130, -0.152, and -0.004, respectively. Thus, it slowly increases between year 7 and 8 and then decreases until year 10. In Figure 5.32 we see that ESP1 decreases faster than ESP6. However, the ESP1 decrement

rate is slowing after year 9 and this gives to ESP6 the opportunity to reach ESP1 at year 10. Apart from the lower boundary issue, the same reasoning is applicable when trajectories similar to those of patients 39 and 166 are involved: since model J6 uses lagged covariate values, which are lower than the current values, ESP6 decrease more slowly than ESP1 and this become more evident at year 10 (see Figure 5.32).

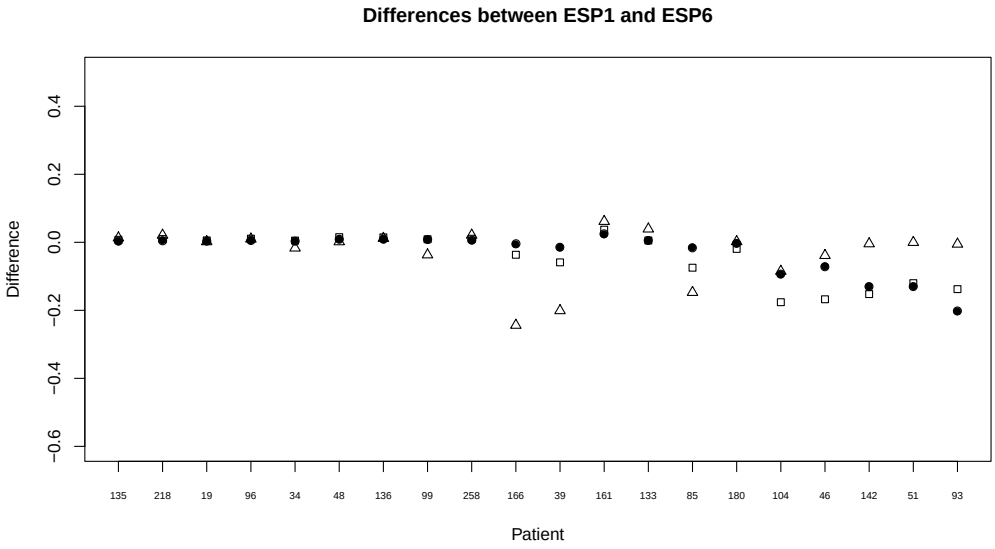


Figure 5.31: DIFF16 for $S(7 \mid 6)$ [circle], $S(8 \mid 6)$ [square], and $S(10 \mid 6)$ [triangle]

	$S(7 \mid 6)$			$S(8 \mid 6)$			$S(10 \mid 6)$		
Patient	ESP1	ESP6	DIFF16	ESP1	ESP6	DIFF16	ESP1	ESP6	DIFF16
135	0.996	0.992	0.004	0.991	0.984	0.007	0.973	0.960	0.013
218	0.995	0.991	0.004	0.989	0.979	0.010	0.965	0.944	0.021
19	0.995	0.991	0.004	0.986	0.980	0.006	0.944	0.942	0.002
96	0.992	0.986	0.006	0.980	0.969	0.011	0.925	0.916	0.009
34	0.984	0.980	0.004	0.958	0.953	0.005	0.842	0.859	-0.017
48	0.980	0.970	0.010	0.950	0.935	0.015	0.838	0.836	0.002
136	0.975	0.966	0.009	0.938	0.924	0.014	0.809	0.797	0.012
99	0.975	0.967	0.008	0.935	0.926	0.009	0.767	0.804	-0.037
258	0.963	0.956	0.007	0.906	0.896	0.010	0.718	0.697	0.021
166	0.962	0.966	-0.004	0.881	0.918	-0.037	0.457	0.701	-0.244
39	0.951	0.965	-0.014	0.851	0.910	-0.059	0.470	0.671	-0.201
161	0.890	0.865	0.025	0.754	0.717	0.037	0.477	0.415	0.062
133	0.878	0.873	0.005	0.725	0.719	0.006	0.421	0.382	0.039
85	0.866	0.882	-0.016	0.659	0.734	-0.075	0.214	0.362	-0.148
180	0.789	0.791	-0.002	0.547	0.566	-0.019	0.174	0.172	0.002
104	0.755	0.849	-0.094	0.461	0.637	-0.176	0.114	0.199	-0.085
46	0.726	0.797	-0.071	0.378	0.546	-0.168	0.038	0.077	-0.039
142	0.617	0.747	-0.13	0.298	0.450	-0.152	0.066	0.071	-0.005
51	0.548	0.678	-0.13	0.235	0.356	-0.121	0.048	0.048	0
93	0.370	0.572	-0.202	0.095	0.233	-0.138	0.006	0.012	-0.006

Table 5.14: Comparison of ESP1, ESP6, and DIFF16 over time

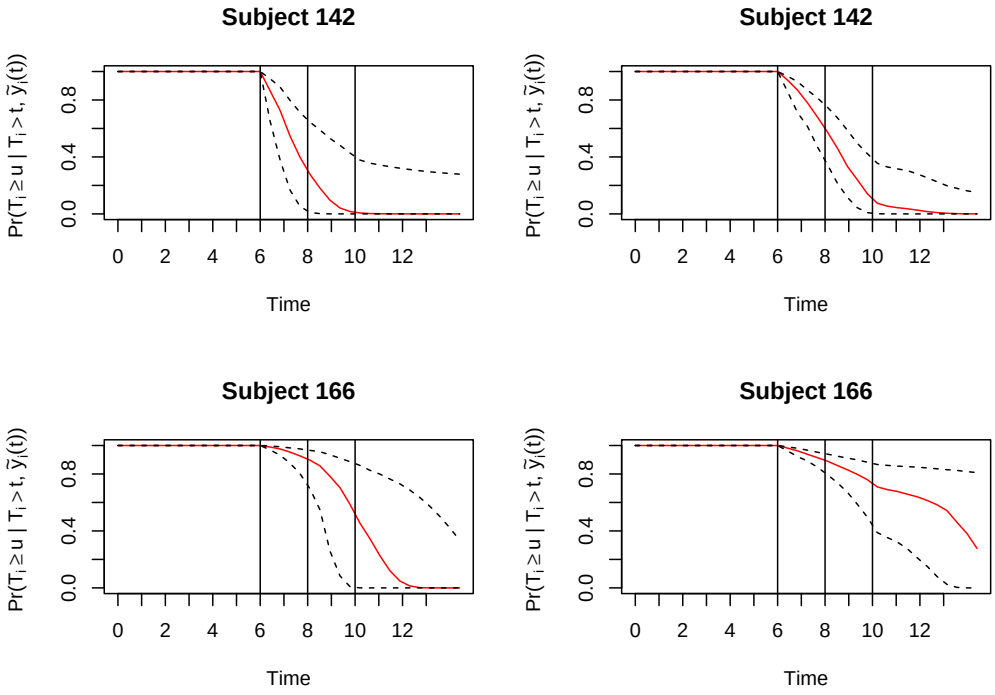


Figure 5.32: Comparison of ESP1 (first column) and ESP6 (second column) of two patients. The solid line depicts the mean of $\pi_i(u \mid t)$ over the Monte Carlo samples. The dashed line depicts a 95% pointwise confidence intervals based on the quantiles of $\pi_i(u \mid t)$ over the Monte Carlo samples.

ESP1 vs ESP7

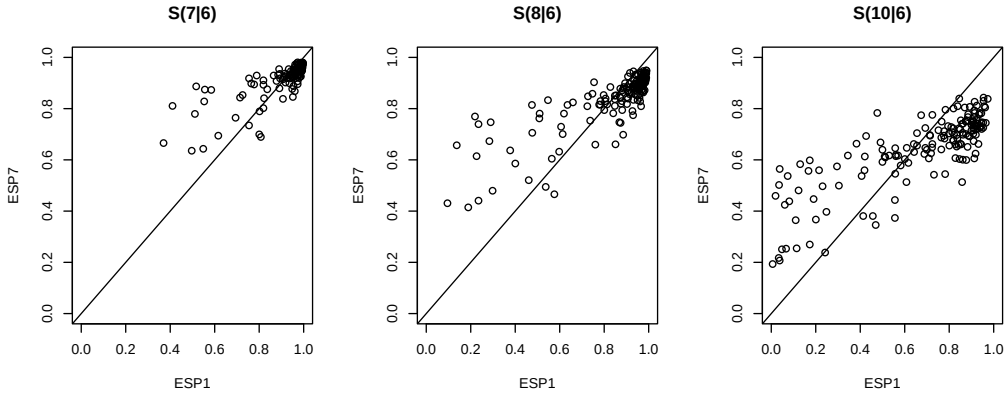


Figure 5.33: Scatter plots of ESP1 and ESP7 for $S(7 \mid 6)$, $S(8 \mid 6)$, and $S(10 \mid 6)$

Comparing Figure 5.20 with Figure 5.33 and Figure 5.21 with Figure 5.34 we see that the patterns are the same. However, comparing ESP7s with ESP2s, we notice that model J7 is slightly more optimist or more pessimist than model J2. Again, this behaviour is strictly linked to the patient's longitudinal trajectory. Since model J7 uses lagged slope values, ESP7s are higher than ESP2s if the current slope is increasing after year 6. In fact, model J7 uses smaller slope values. On the contrary, if the current slope is decreasing, ESP7 will be lower than ESP2 because ESP7 uses higher slope values. Comparing the results in Table 5.15 and Table 5.10, we see that, in the former case, DIFF17 tends to be bigger than DIFF12 (see, for instance, patients 135, 96, 218, 46), whereas, in the latter case, it will be smaller (see, for instance, patients 166, 133, 142, 51, 93).

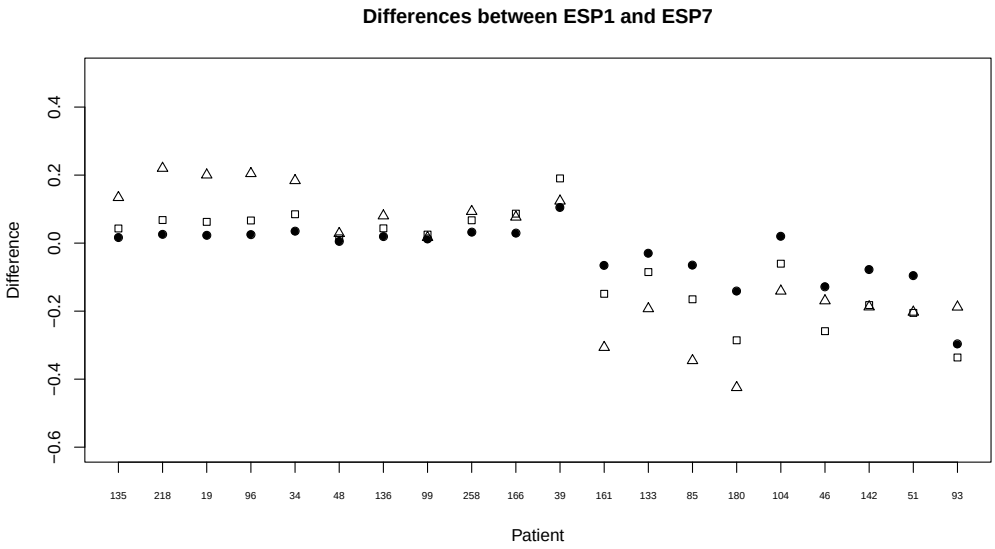


Figure 5.34: DIFF17 for $S(7 \mid 6)$ [circle], $S(8 \mid 6)$ [square], and $S(10 \mid 6)$ [triangle]

	$S(7 \mid 6)$			$S(8 \mid 6)$			$S(10 \mid 6)$		
Patient	ESP1	ESP7	DIFF17	ESP1	ESP7	DIFF17	ESP1	ESP7	DIFF17
135	0.996	0.979	0.017	0.991	0.948	0.043	0.973	0.839	0.134
218	0.995	0.970	0.025	0.989	0.921	0.068	0.965	0.745	0.220
19	0.995	0.971	0.024	0.986	0.924	0.062	0.944	0.744	0.200
96	0.992	0.967	0.025	0.980	0.913	0.067	0.925	0.721	0.204
34	0.984	0.948	0.036	0.958	0.873	0.085	0.842	0.658	0.184
48	0.980	0.974	0.006	0.950	0.936	0.014	0.838	0.809	0.029
136	0.975	0.956	0.019	0.938	0.895	0.043	0.809	0.728	0.081
99	0.975	0.963	0.012	0.935	0.910	0.025	0.767	0.750	0.017
258	0.963	0.930	0.033	0.906	0.839	0.067	0.718	0.625	0.093
166	0.962	0.932	0.030	0.881	0.794	0.087	0.457	0.381	0.076
39	0.951	0.846	0.105	0.851	0.661	0.190	0.470	0.346	0.124
161	0.890	0.955	-0.065	0.754	0.903	-0.149	0.477	0.783	-0.306
133	0.878	0.908	-0.030	0.725	0.810	-0.085	0.421	0.614	-0.193
85	0.866	0.930	-0.064	0.659	0.825	-0.166	0.214	0.560	-0.346
180	0.789	0.929	-0.140	0.547	0.833	-0.286	0.174	0.599	-0.425
104	0.755	0.735	0.020	0.461	0.521	-0.060	0.114	0.255	-0.141
46	0.726	0.854	-0.128	0.378	0.637	-0.259	0.038	0.208	-0.170
142	0.617	0.695	-0.078	0.298	0.480	-0.182	0.066	0.254	-0.188
51	0.548	0.643	-0.095	0.235	0.441	-0.206	0.048	0.251	-0.203
93	0.370	0.666	-0.296	0.095	0.431	-0.336	0.006	0.194	-0.188

Table 5.15: Comparison of ESP1, ESP7, and DIFF17 over time

ESP1 vs ESP8

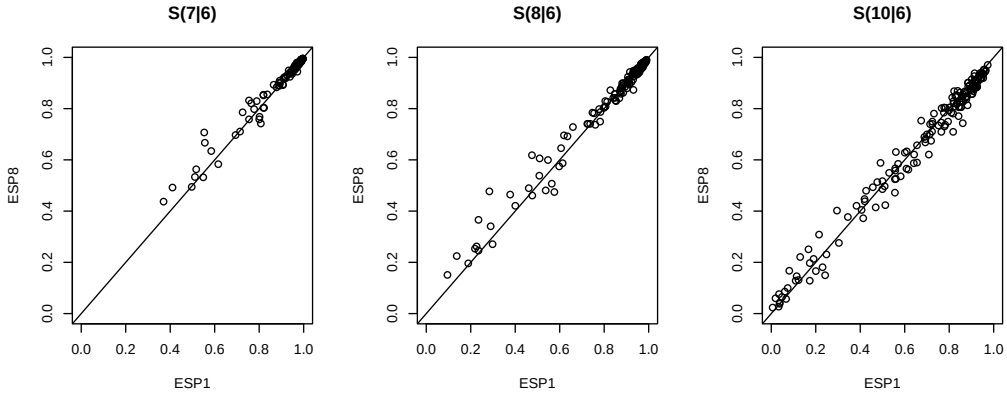


Figure 5.35: Scatter plots of ESP1 and ESP8 for $S(7 \mid 6)$, $S(8 \mid 6)$, and $S(10 \mid 6)$

Also in this case, as already happened in the comparison between ESP1s and ESP3s, model J8 give results which are very close to those of model J1 (see Figure 5.35). We report Figure 5.36 and Table 5.16 for completeness.

	$S(7 \mid 6)$			$S(8 \mid 6)$			$S(10 \mid 6)$		
Patient	ESP1	ESP8	DIFF18	ESP1	ESP8	DIFF18	ESP1	ESP8	DIFF18
135	0.996	0.996	0	0.991	0.991	0	0.973	0.970	0.003
218	0.995	0.994	0.001	0.989	0.986	0.003	0.965	0.949	0.016
19	0.995	0.995	0	0.986	0.986	0	0.944	0.936	0.008
96	0.992	0.991	0.001	0.980	0.974	0.006	0.925	0.884	0.041
34	0.984	0.982	0.002	0.958	0.948	0.010	0.842	0.771	0.071
48	0.980	0.983	-0.003	0.950	0.958	-0.008	0.838	0.870	-0.032
136	0.975	0.974	0.001	0.938	0.934	0.004	0.809	0.785	0.024
99	0.975	0.979	-0.004	0.935	0.946	-0.011	0.767	0.803	-0.036
258	0.963	0.962	0.001	0.906	0.900	0.006	0.718	0.675	0.043
166	0.962	0.968	-0.006	0.881	0.900	-0.019	0.457	0.494	-0.037
39	0.951	0.943	0.008	0.851	0.832	0.019	0.470	0.415	0.055
161	0.890	0.902	-0.012	0.754	0.782	-0.028	0.477	0.513	-0.036
133	0.878	0.883	-0.005	0.725	0.741	-0.016	0.421	0.438	-0.017
85	0.866	0.893	-0.027	0.659	0.728	-0.069	0.214	0.308	-0.094
180	0.789	0.830	-0.041	0.547	0.600	-0.053	0.174	0.198	-0.024
104	0.755	0.758	-0.003	0.461	0.489	-0.028	0.114	0.146	-0.032
46	0.726	0.786	-0.060	0.378	0.465	-0.087	0.038	0.039	-0.001
142	0.617	0.584	0.033	0.298	0.271	0.027	0.066	0.058	0.008
51	0.548	0.532	0.016	0.235	0.246	-0.011	0.048	0.064	-0.016
93	0.370	0.437	-0.067	0.095	0.152	-0.057	0.006	0.023	-0.017

Table 5.16: Comparison of ESP1, ESP8, and DIFF18 over time

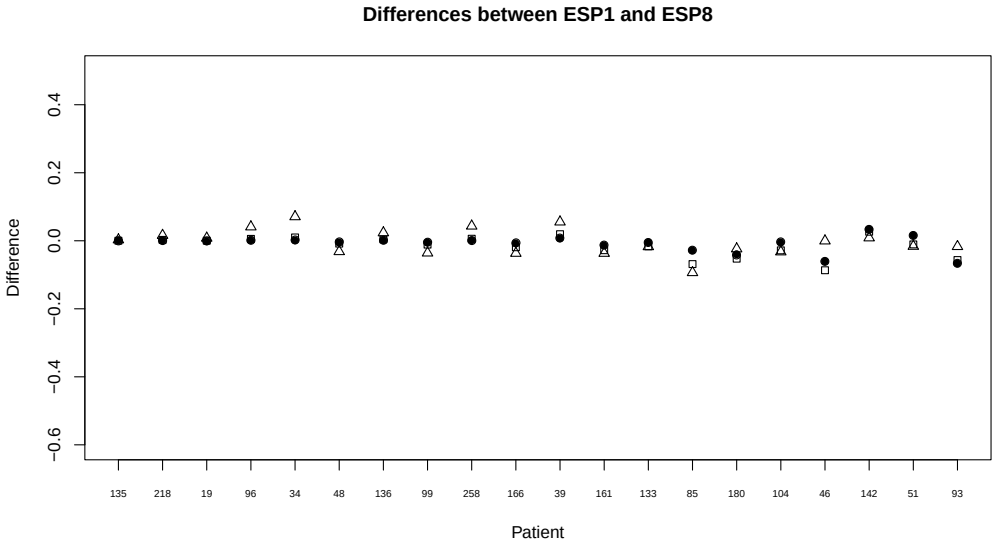


Figure 5.36: DIFF18 for $S(7 \mid 6)$ [circle], $S(8 \mid 6)$ [square], and $S(10 \mid 6)$ [triangle]

Conclusions

Longitudinal studies, such as clinical trials or observational studies, often produce two types of outcome: one or more repeatedly measured biomarkers, along with baseline covariates, and the elapsed time to an event of interest. These markers are frequently important heath indicators since they can be used to monitor the disease progression. In order to optimally use the information contained in the collected data and significantly reduce the bias due to the traditional separate analyses of these types of data, a joint modeling approach has been recently proposed and is currently under development. In this approach, the longitudinal process and the time-to-event process are modeled simultaneously.

During the last decade, there has been an increasing interest in the use of joint models to obtain subject-specific predictions for the survival outcome. Survival predictions based on a joint model have the advantage that they can be updated as soon as new information is available.

The association structure represents the way the longitudinal outcome is related to the risk for an event. In the standard joint model, it is usually assumed that the risk for an event at a particular time point t depends on the current value of the longitudinal biomarker. However, this type of association may not always be the most appropriate in expressing the correct relationship between the longitudinal and the survival processes so that other types of association structures should be considered. In fact, choosing one association structure instead of another may substantially influence the derived results.

The aim of this work was indeed to investigate how sensitive are survival

predictions with respect to the assumed association structure of the survival sub-model. Seven where the alternative association structures considered in this work which were formalized through seven different parameterizations.

Our analysis was based on data collected during a study on primary biliary cirrhosis (PBC). The serum bilirubin level is the longitudinal covariate considered while death is the event of interest. After having estimated the eight joint models, for each of them, we computed the patients' expected survival probabilities (ESPs) at different time points. We remind that the expected survival probability of a patient is the estimate of the conditional probability of surviving time $u > t$ given survival up to t, i.e. $S(t + \Delta t \mid t)$. Subsequently, we compare the ESPs of the standard model with those of the alternative models.

Not surprisingly, those parameterizations which were more similar to the standard one were more likely to give analogous ESPs. However, independently of the alternative parameterization used, the differences between ESPs where more accentuate for the high risk patients. Their longitudinal trajectories, in fact, are characterized by high serum bilirubin values and an increasing trend (i.e. positive slope) which may be more or less pronounced. With this type of trajectories, in comparison with the standard parameterization, the different interpretation of the trajectory features given by the alternative parameterizations shows its maximum effect. On the contrary, the discrepancies between ESPs of low risk patients are usually small. In fact, their longitudinal trajectories are characterized by small serum bilirubin values, especially at the beginning of the study, which tend to stay constant or slightly increase over time. These trajectory features guarantee similar ESPs. Thus, in general, for low risk patients, using one association structure instead of another does not have a strong influence on the obtained ESPs. On the contrary, for high risk patients the choice of the association structure has a much stronger impact.

Furthermore, in this type of analysis, the evolution in time of the longitudinal covariate should not be forgotten. In fact, the ESPs comparisons

have been done considering different values for both t and Δt. In particular, the most interesting results were found keeping t fixed while increasing the prediction interval Δt. In our analysis we have chosen $\Delta t = 1, 2, 4$ years. The fact that for $\Delta t = 1$, the standard parameterization and an alternative one give similar or, on the contrary, very different ESPs does not guarantee that it will be the same for $\Delta t = 2$ or $\Delta t = 4$. Differences can either stay constant or increase or decrease when Δt increases. Again, everything depends on the evolution of the patient's longitudinal trajectory whose features are differently interpreted by the eight considered models.

Besides, monitoring the value of the differences between ESPs as Δt increases, we noticed that, for high-risk patients, it was decreasing. We called this phenomenon the *low-ESP effect*. In comparison with the other alternative parameterizations, the standard parameterization tends to be more pessimist with high risk patients. This means that, for these patients, the ESPs of the standard joint model have a higher decreasing rate so that they already reach values close to zero when $\Delta t = 2$. Since zero is the lower boundary, the ESP of the standard joint model will not significantly decrease any further while the ESP of the alternative model, being much higher when $\Delta t = 2$, will still have space to decrease. Thus, as Δt augments, the two ESP values will become closer.

We are aware of the fact that our work may be questionable since the differences in ESPs are not only due to the assumed association structure but also to the fact that each model has its own estimates of the parameters defining the longitudinal sub-model and the baseline hazard function of the survival sub-model. However, it can easily be noticed that it is the association structure to have the greatest influence on the computation of the ESPs. If this wasn't true, we would not have individuate the strong link between the ESPs and the features of the longitudinal trajectory.

We believe that joint models, thanks to their capability to maximize the use of the information contained in prognostic factors such as biomarkers, will constitute a precious tool in the future clinical practice. The physician expe-

rience combined with a subject-specific survival probability would hopefully help decision making and improve clinical outcome. Due to their important application in patients' management, it is important to obtain survival probabilities which are as accurate as possible. Our analysis has demonstrated how strong can be the influence of the assumed association structure on individual expected survival probabilities, especially for high risk patients. Therefore, in the part of the data analysis dedicated to the choice of the more appropriate prognostic model, it is fundamental to consider and investigate alternative associations between the longitudinal covariate and the risk for an event and not solely rely on the standard one. The availability of statistical softwares within the joint modeling framework has recently made this investigation rather feasible.

Bibliography

A. Abu-Hanna and P. J. F. Lucas (2001). Prognostic models in medicine: AI and statistical approaches. *Methods of Information in Medicine* **40**, 1–5.

Ali, K., Gunnar, A., Jan-Erik, D., Hans, L., Pr, L., and Jonas, H. (2006). PSA doubling time predicts the outcome after active surveillance in screening-detected prostate cancer: results from the European randomized study of screening for prostate cancer, Sweden section. *International Journal of Cancer* **120**, 170–174.

Aronson, J. K. (2005). Biomarkers and surrogate endpoints. *British Journal of Clinical Pharmacology* **59**, 491–494.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

Brown, E., Ibrahim, J. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* **59**, 221–228.

Brown, E., Ibrahim, J., and DeGruttola, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* **61**, 64–73.

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., and Geys, H. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* **1**, 49–67.

Bycott, P., and Taylor, J. (1998). A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Statistics in Medicine* **17**, 2061–2077.

Carter, H. B., and Pearson, J. D. (1993). PSA velocity for the diagnosis of early prostate cancer. A new concept. *The Urologic Clinics of North America* **20**, 665–670.

Cavender, J. B., Rogers, W. J., Fisher, L.D., Gersh, B. J., Coggin, J. C., and Myers, W.O. (1992). Effects of smoking on survival and morbidity in patients randomized to medical or surgical therapy in the Coronary Artery Surgery Study (CASS): 10-year follow-up. *Journal of the American College of Cardiology* **20**, 287–294.

Chi, Y.-Y., and Ibrahim, J. (2006). Joint models for multivariate longitudinal and multivariate survival data.*Biometrics* **62**, 432–445.

Chib, S., and Edward Greenberg (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician* **49**, 327–335.

Christakis, N. A., and Iwashyna, T. J. (1998). Attitude and self-reported practice regarding prognostication in a National sample of internists. *Archives of internal medicine* **158**, 2389–2395.

Christakis, N. A., and Lamont, E. B. (2000). Extent and determinants of error in physicians' prognoses in terminally ill patients: prospective cohort study. *British medical journal* **320**, 469-473.

Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **34**, 187–220.

Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

De Gruttola, V., Wulfsohn, M., Fischl, M. A., and Tsiatis, A. (1993). Modeling the relationship between survival and CD4 lym-

phocytes in patients with AIDS and AIDS-related complex. *Journal of Acquired Immune Deficiency Syndromes* **6**, 359–365.

Efron, B. (1974). The efficiency of Coxs likelihood function for censored data. *Journal of the American Statistical Association* **72**, 557–565.

Fisher, L. D., and Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health* **20**, 145–57.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2008). *Longitudinal data analysis*, Chapman and Hall/CRC, New York.

Fleming, T. R., and Harrington, D. P. (1991). *Counting processes and survival analysis*, Wiley, New York.

Glare, P., Sinclair, C., Downinge, M., Stoneg, P., Maltonih, M. and Vigano, A. (2008). Predicting survival in patients with advanced disease. *European journal of cancer* **44**, 1146–1156.

Gould, S. J. (1986). The median isn't the message. Available at: `http://cancerguide.org/median_not_msg.html`. Last access: 05/03/2012

Hagerty, R. G., Butow, P. N., Ellis, P. M., Lobb, E. A., Pendlebury, S. C., Leighl, N., Mac Leod, C. and Tattersall, M. H. N. (2005). Communicating with realism and hope: incurable cancer patients views on the disclosure of prognosis. *Journal of Clinical Oncology* **23**, 1278–1288.

Hannan, E. L. (2008). Randomized clinical trials and observational studies. Guidelines for assessing respective strengths and limitations. *Journal of the American College of Cardiology* **1**, 211–217.

Hanson, T.E., Branscum, A.J. and Johnson W.O. (2011). Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. *Lifetime data analysis* **17**, 3–28.

Henderson, R., Diggle, P., Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data.*Biostatistics* **1**, 465–480.

Henderson, R., Jones, M., and Stare, J. (2001). Accuracy of point predictions in survival analysis. *Statistics in medicine* **20**, 3083–3096.

Henderson, R., and Keiding, N. (2005). Individual survival time prediction using statistical models. *Journal of medical ethics* **31**, 703–706.

Hsieh, F., Tseng, Y. K., and Wang, J. L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* **62**, 1037–1043.

Hollnagel, H. (1999). Explaining risk factors to patients during a general practice consultation. Conveying group-based epidemiological knowledge to individual patients. *Scandinavian Journal of Primary Health Care* **17**, 3–5.

Kalbfleisch, J., and Prentice, R. (2002). *The statistical analysis of failure time data*. 2nd edition. John Wiley & Sons, New York.

Klein, J. P., and Moeschberger, M. L. (2003). *The statistical analysis of failure time data*. 2nd edition. Springer-Verlag, New York.

Krzeski, P., Zych, W., Kraszewska, E., Milewski, B., Butruk, E., and Habior, A. (1999). Is serum bilirubin concentration the only valid prognostic marker in primary biliary cirrhosis? *Hepatology* **30**, 865–869.

Mackillop, W. J., and Quirt, C. F. (1997). Measuring the accuracy of prognostic judgments in oncology. *Journal of clinical epidemiology* **50**, 21–29.

Mackillop, W. J. (2006). The importance of prognosis in cancer medicine. TNM Online.

Markus B.H., Dickson, E.R., Grambsch, P.M., Fleming, T.R., Mazzaferro, V., Klintmalm, G.B., Wiesner, R.H., Van Thiel, D.H., and Starzl, T.E. (1989). Efficacy of liver transplantation in patients with primary biliary cirrhosis. *The New England Journal of Medicine* **320**, 1709–1713.

Mayo Clinic Primary biliary cirrhosis. Available at: `http://www.mayoclinic.com/health/primary-biliary-cirrhosis/DS00604`. Last access: 22/03/2012

Mayo, M. J. (2008). Natural history of primary biliary cirrhosis. *Clinics in Liver Disease* **12**, 277–288.

Medicare Medicare Hospice Benefits. Available at: `http://www.medicare.gov/Publications/Pubs/pdf/02154.pdf`. Last access: 28/03/2012.

Molenberghs, G., and Kenward, M. G. (2007). *Missing data in clinical studies*. John Wiley & Sons, Chichester.

Muers, M. F., Shevlin, P., and Brown, J. (1996). Prognosis in lung cancer: Physicians opinions compared with outcome and a predictive model. *Thorax* **51**, 894–902.

Murtaugh, P. A., Dickson, E. R., Van Dam, G. M., Malinchoc, M., Grambsch, P.M., Langworthy, A.L., and Gips, C.H. (1994). Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology* **20**, 126–34.

NIH Definitions Working Group (2001). Biomarkers and surrogate endpoints: definitions and conceptual framework. *Clinical Pharmacology & Therapeutics* **69**, 89-95.

Oesterling, J. E. (1991). Prostate specific antigen: a critical assessment of the most useful tumor marker for adenocarcinoma of the prostate. *The Journal of Urology* **145**, 907–923.

Parkes, C. M. (1972). Accuracy of predictions of survival in the later stages of cancer. *British medical journal* **2**, 29-31.

Pares, A. and Rodes, J. (2003). Natural history of primary biliary cirrhosis. *Clinics in Liver Disease* **7**, 779–794.

Peace, K. E. and Chen, D. (2011). *Clinical Trial Methodology*, Chapman and Hall/CRC Press, Boca Raton.

Pollack A., Zagars, G. K., and Kavadi, V. S. (1994). Prostate specific antigen doubling time and disease relapse after radiotherapy for prostate cancer. *Cancer* **74**, 670–678.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine* **8**, 431–440.

Proust-Lima, C., and Taylor, J. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* **10**, 535–549.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika* **95**, 63–74.

Rizopoulos, D., Verbeke, G., and Lesaffre, E. (2009). Fully exponential Laplace approximation for the joint modelling of survival and

longitudinal data. *Journal of the Royal Statistical Society. Series B* **71**, 637–654.

Rizopoulos, D. (2010a). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of statistical software* **35**, 1–33.

Rizopoulos, D. (2010b). Reference manual of the package 'JM'. Available at: `http://cran.r-project.org/web/packages/JM/JM.pdf`. Last access: 22/03/2012.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **63**, 819–829.

Schumacher, M., Graf, E., and Gerds, T. (2003). How to assess prognostic models for survival data: a case study in oncology. *Methods of information in medicine* **42**, 564–71.

Shapiro, J. M., Smith, H., and Schaffner, F. (1979). Serum bilirubin: a prognostic factor in primary biliary cirrhosis. *Gut* **20**, 137–140.

Sweeting, M. J., and Thompson, S. G. (2011). Joint modeling of longitudinal and time-to-event data with aplication to predict abdominal aortic aneurysm growth and rupture. *Biomedical Journal* **53**, 750–763.

Song, X., Davidian, M., Tsiatis, A. (2002). A semiparameteric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742–753.

Taylor, J. M. G., and Wang, Y. (2002). Surrogate markers and joint models for longitudinal and survival data. *Controlled Clinical Trials* **23**, 626–634.

The Federal Medicare Agency Medicare hospice benefits. Available at:

http://www.medicare.gov/publications/pubs/pdf/hosplg.pdf. Last access: 05/03/2012

Thernau, T., and Grambsch, P. (2000). *Modeling survival data: extending the Cox model*. Springer-Verlag, New York.

Tsiatis, A., DeGruttola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90**, 27–37.

Tsiatis, A., and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447–458.

Tsiatis, A., and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* **14**, 809–834.

Vasan, R. S. (2006). Biomarkers of cardiovascular disease : molecular basis and practical considerations. *Circulation* **113**, 2335–2362.

Verbeke, G., Molenberghs, G., and Rizopoulos, D. (2010). *Longitudinal research with latent variables*, Springer-Verlag, Berlin Heidelberg.

WebMD Prostate-Specific Antigen (PSA). Available at: http://men.webmd.com/prostate-specific-antigen-psa. Last access: 06/03/2012

Weeks, J. C., Cook, E. F., O'Day, S. J., Peterson, L. M., Wenger, N., Reding, D., Harrell, F. E., Kussin, P., Dawson, N. V., Connors, A. F., Lynn, J., and Phillips, R. S. (1998). Relationship between cancer patients predictions of prognosis and their treatment preferences. *The journal of the American association* **279**, 1709–1714.

Wikipedia contributors Huntington's disease . Available at: http://en.wikipedia.org/w/index.php?title=Huntington%27s_disease&oldid=480115835. Last access: 06/03/2012

Wu, L. (2010). *Mixed effect model for complex data*. Chapman & Hall/CRC, London.

Wulfsohn, M., and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.

Yu, M., Law, N., Taylor, J., and Sandler, H. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica* **14**, 835–832.

Yurkovetsky, Z. R., Linkov, F. Y., Malehorn, D., and Lokshin, A. E. (2006). Multiple biomarker panels for early detection of ovarian cancer. *Future Oncology* **2**, 733–741.

Zagars, G. K., and Pollack, A. (1993). The fall and rise of prostate-specific antigen. Kinetics of serum prostate-specific antigen levels after radiation therapy for prostate cancer. *Cancer* **72**, 832–842.

Zethelius, B., Berglund, L., Sundstrm, J., Ingelsson, E., Basu, S., Larsson, A., Venge, P., and rnlv, J. (2008). Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. *The New England Journal of Medicine* **358**, 2107–2116.