KU LEUVEN

FACULTY OF SCIENCE

Joint Latent Class Modelling to Dynamically Predict the Risk for a Miscarriage

Helene VERMEULEN

Supervisor: Prof. dr. Bart De Moor
Affiliation: *KU Leuven*

Co-supervisor: Dr. Thibaut Vaulet
Affiliation: *KU Leuven*

Reader 1: Prof. Emmanuel Lesaffre
Affiliation: *KU Leuven*

Reader 2: Prof. Ingrid Van Keilegom
Affiliation: *KU Leuven*

Thesis presented in
fulfillment of the requirements
for the degree of Master of Science
in Statistics

Academic year 2018-2019

Acknowledgements

You are about to read the final project with which I complete the master program in statistics at the KULeuven. However, this work could never have been established without the help of many others. I would therefore like to thank the following persons.

First of all, I would like to thank professor dr. Bart De Moor for giving me the opportunity to fulfill this project and to provide me with the extremely interesting and challenging data used for the analysis. I also want to express my special gratitude to dr. Thibaut Vaulet for proofreading the assay and guiding me through the whole process. At last, I want to thank all friends and family who supported me during my studies, with special thanks to Milan Pelgrims who has always been my most understanding, encouraging and loving friend.

Helene Vermeulen

Contents

List of Figures

List of Tables

List of Abbreviations

AIC: Akaike's Information Criterion
BIC: Bayesian Information Criterion
BMI: Body Mass Index
CRL: Crown Rump Length
CVPOL: Cross-Validated Prognosis Observed Loss
CVPOL$_a$: Approximated Cross-Validated Prognosis Observed Loss
EPOCE: Expected Prognostic Observed Cross-Entropy
EPOS: Early Pregnancy Events and the Impact on Short-term and Long-term Pregnancy Outcomes
GA: Gestational Age
GOF: Goodness-of-fit
LMP: Last Menstrual Period
MSD: Mean Sac Diameter
MYS: Mean Yolk Sac size
PBAC: Pectoral Blood Assessment Chart
PC: Principal Complaint
PH: Proportional Hazards
PUV: Pregnancy of Unknown Viability
PUQE: Pregnancy-Unique Quantification of Emesis and Nausea
SCH: Subchriotonic Hematoma
TOP: Termination Of Pregnancy
UPT: Urinary Pregnancy Test

Abstract

Spontaneous abortion or miscarriage is defined as the loss of pregnancy before 24 weeks of gestational age and is the most common complication during early pregnancies. During the first trimester, miscarriages are estimated to occur in 15-30% of the pregnancies. A variety of risk factors have been proposed in the past and several studies have attempted to implement these potential risk factors in a predictive risk model. However, many of these previous studies used a univariate or multivariate logistic regression model in order to predict the risk for a miscarriage. These methods do not take into account the longitudinal aspect of a pregnancy. We applied the joint latent class modelling approach to the EPOS data-set, containing information on 753 pregnancies. A joint latent class model can simultaneously model a longitudinal process and a survival process, taking into account heterogeneity within the population. The joint latent class model consists of 3 sub-models, i.e. a multinomial logistic regression model to define class membership, a proportional hazards model to model the time-to-event and a linear mixed model to describe the evolution of a longitudinal variable. Parameter estimates for the joint latent class model are obtained using maximum likelihood theory. In this assay, we report on several different joint latent class models that aim to identify the risk for a miscarriage. These models include ultra-sound features as longitudinal outcome variable in the linear mixed model and have two to five latent classes. The best model identified in our study models the crown rump length of the fetus over time and has 5 latent classes. The model shows good discrimination, reasonable longitudinal and survival profiles and a good predictive accuracy. This model might be very useful in practice, as it allows for individual dynamic miscarriage predictions.

Chapter 1

Introduction

Spontaneous abortion or miscarriage is defined as the loss of pregnancy before 24 weeks of *gestational age* (GA) and is the most common complication during early pregnancies [1]. During the first trimester, miscarriages are estimated to occur in 15-30% of the pregnancies [2]. When fetal viability has been demonstrated by ultrasound, the incidence of a miscarriage in the first trimester reduces to 2-16% [3]. Among women attending an early pregnancy unit, the miscarriage rate is 17-46% [4]. The much rarer event of a mid-trimester miscarriage is estimated to occur in 2-3% of the pregnancies [5]. Although a miscarriage is often not associated with serious physical morbidity or mortality to the mother, it has a major social and psychological impact on the parents to be [4]. It appears that 10-50% of the patients experience a major depressive disorder after a pregnancy loss [6]. In the interest of the parent's mental health and the correct counseling and follow-up of the expecting mother, it is of great importance to accurately inform the future parents about the likelihood of an ongoing pregnancy.

A variety of risk factors for miscarriage have been proposed in the past. These factors can be be either modifiable or non-modifiable and can be acting before and/or during the pregnancy. Chromosomal aneuploidies have been proven to be the main cause for miscarriages accounting for 50-70% [1]. Modifiable risk factors mainly include lifestyle factors such as alcohol consumption, smoking status, caffeine intake, high *body mass index* (BMI), exercise, increased maternal age, stress, underweight, lifting more than 20kg daily and night work [7,8]. Nilsson et al. estimated that 25.2% of the miscarriages can be prevented when modifiable risk factors would be reduced to low levels [8]. Non-modifiable risk factors include low serum progesterone levels, vaginal bleedings, abdominal pain, gestational age at onset of bleeding, uterine size and fetal cardiac activity [7]. Nausea and vomiting lower the risk for a miscarriage, where woman with nausea and vomiting have a lower risk for a miscarriage than woman experiencing only nausea [9]. Also, daily vitamin intake is associated with a lower risk for miscarriages [10]. Further, also ultrasound features can be used to estimate the risk for a miscarriage. Falco et al. showed that fetal bradycardia, a discrepancy between the diameter of the gestational sac and the *crown-rump length* (CRL) and a discrepancy between the menstrual and sonographic gestational age of more than 1 week, increase the risk for a miscarriage [11].

Several studies have attempted to implement these potential risk factors in a predictive risk model for miscarriage, attaining variable performances [2–4, 11–15]. However, these studies often only contain a subset of the potential risk factors and are often focusing on a sub-population of pregnant women such as infertility patients or expectant mothers without a detectable embryo. Besides, many analyses use a univariate or multivariate logistic regression model in order to predict the risk for a miscarriage. Although logistic regression might be a suitable tool to predict the chance for a miscarriage, it does not take into account the longitudinal aspect of the pregnancy.

Joint modelling, a statistical method that is used to jointly model a longitudinal process as well as a survival process, might therefore be more suitable as compared to logistic regression to identify miscarriage risk factors and to build a predictive risk model. Due to the fact that a proportional hazards survival model is implemented in the joint modelling approach, the probability to have a miscarriage can be estimated over time. Besides, because the survival process is jointly modelled with the evolution of longitudinally measured variables by the implementation of a linear mixed model in the joint modelling framework, the correlation between both processes is taken into account. Because the majority of pregnant women will never experience a miscarriage, an extension of the joint modelling approach, i.e. a joint latent class model, is used to take into account the heterogeneity within the population of pregnant women. Although many researchers are unfamiliar with the joint latent class modelling approach, it has been used before in studies with similar data. Proust-Lima et al. used a joint latent class model to simultaneously model the longitudinal profile of prostate specific antigen and the risk of prostate cancer recurrence and to investigate the cognitive evolution of elderly and the associated risk of dementia [16,17].

The aim of this study was to build predictive miscarriage risk models based on the joint latent class modelling framework. Due to the complex structure of the data-set, including time-to-event data as well as longitudinally measured variables within a heterogeneous population, it is assumed that a joint latent class model is more appropriate for the prediction of a miscarriage as compared to previously described models. In chapter 2, an extensive description of the EPOS data-set is given as well as a throughout explanation of the joint modelling framework. In chapter 3, the obtained results are given. Chapter 4 gives a critical discussion on the study and chapter 5 concludes the assay.

Chapter 2

Research Methods

2.1 The EPOS Data-Set

The data set analyzed in this study originates from the EPOS study, that is a study on *early pregnancy events and the impact on short-term and long-term pregnancy outcomes*, conducted by the Imperial College in Londen. The data were collected at the Queen Charlottes and Chelsea Hospital at the Imperial College Healthcare NHS Trusts between 2013 and 2016 in the context of an observational prospective cohort study. The included study subjects are patients attending the early pregnancy unit in the first trimester of their pregnancy, with ages ranging from 17 to 48 years. Informed consent was obtained from all included subjects. Women suffering from an acute medical condition, women diagnosed with a miscarriage at the first examination and women who were unable to give fully informed consent were excluded from the study. Besides, study subjects were allowed to withdraw from the study at any time.

The EPOS data set contains information on baseline covariates measured at the beginning of the pregnancy, as well as on longitudinal variables measured repeatedly during the first trimester of the pregnancy. Information on the baseline covariates was obtained through the completion of a standard questionnaire concerning demographic and obstetric information. Important baseline covariates that might have an impact on the risk for a miscarriage are the maternal age, the alcohol consumption and smoking status of the mother, etc. A detailed list containing all baseline covariates can be found in table 21 in the appendix.

Information on pregnancy related symptoms such as bleeding, pain and vomiting was repeatedly measured throughout the first trimester. The amount of bleeding during the pregnancy was assessed using a modified *pectoral blood assessment chart* (PBAC) scoring system where a score is given on 4. The amount of pain experienced during the pregnancy was recorded using a visual pain score where patients were asked to document the grade and the duration of their pain with a score on 10. The amount of nausea and vomiting was recorded using the *pregnancy-unique quantification of emesis and nausea* (PUQE) scoring system with a maximum score of 15 [18]. Furthermore, the data set contains longitudinal information resulting from routine trans-vaginal ultrasound scans performed every one to two weeks during the first trimester. The outcome variable, with the three possible outcome values 'viable', '*pregnancy of unknown viability*' (PUV) and 'miscarriage', was

examined at every ultrasound scan. For many subjects, a miscarriage was not observed at the last scan, thus resulting in right censored observations. Note that although ultrasound scans were taken every one to two weeks, it can be assumed that the observation of a miscarriage is not interval censored due to the fact that women attend to the early pregnancy unit only when experiencing symptoms such as a bleeding, abdominal pain, etc. A detailed list containing all longitudinal variables can be found in table 22 in the appendix.

Due to the fact that we are dealing with censored data, baseline covariates and longitudinally measured variables, a specific data structure emerges from the EPOS study. The longitudinal variables were repeatedly measured at the moment ultrasound scans were taken. However, the occasions at which scans were taken, as well as the amount of scans that were taken, vary considerably between the different subjects. At least 1 and at maximum 6 scans were taken during the first trimester. At last, the data set also contains a considerable amount of missing data for all variables at all scan times, especially at scan 6. We therefore only used the data up to the fifth scan.

2.2 Outline of the Analysis

Before getting into detail, we want to provide the reader with a brief outline of the whole analysis process. The flowchart given in figure 1 shows the main steps in our analysis. As can be seen from this figure, the initial data-set contained information on 946 subjects. After application of the exclusion criteria, 753 subjects were retained. We then applied an exploratory analysis in order to identify possible covariates for each of the three sub-models that compose the joint latent class model, i.e. a linear mixed model to model fetus growth, a multinomial logistic regression model to model latent class membership and a proportional hazards model to model the time until a miscarriage. Subjects with missing values for any of the covariates included in any of the sub-models were further excluded from the analysis, thereby making the number of included subjects model-dependent. Several different models were fitted, including different combinations of the covariates in the three different sub-models. The maximum likelihood estimates for the parameters of the resulting joint latent class models were then obtained using a modified iterative Marquardt algorithm. After the different models were fitted, the most appropriate model was selected according to the *Bayesian information criterion* (BIC), the *goodness-of-fit* (GOF) and the predictive ability of the model. Finally, the selected model can be used to give individual dynamic predictions for future subjects.

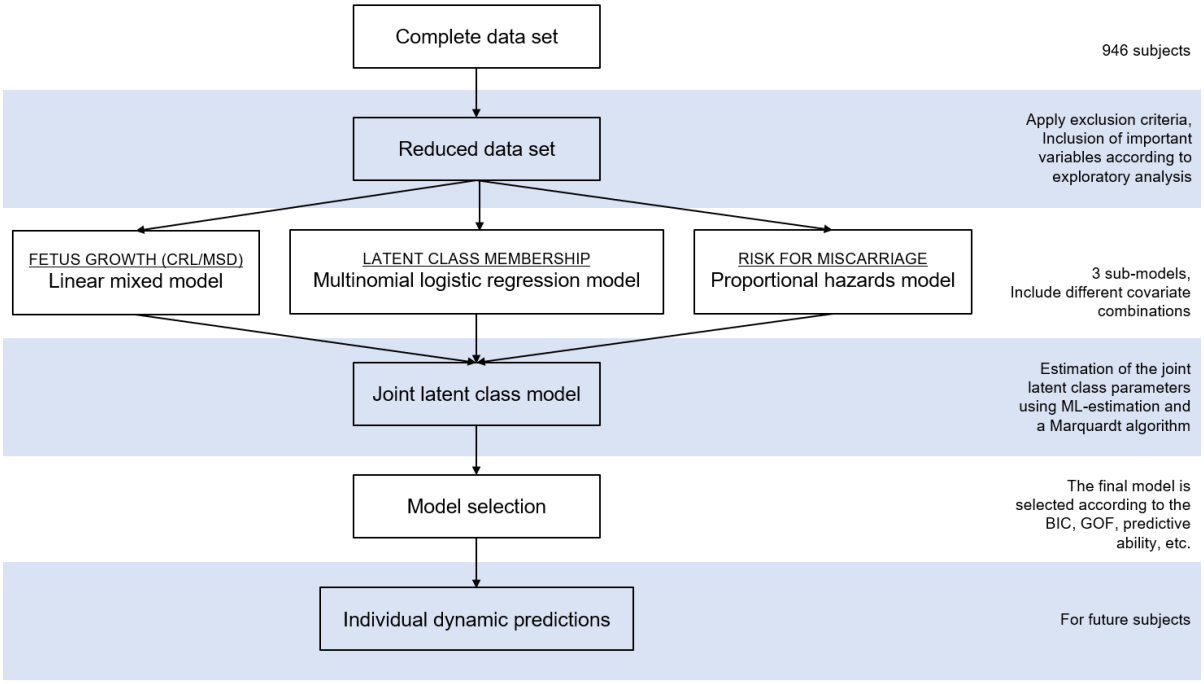

Figure 1: Flowchart showing the whole analysis process: *After application of the exclusion criteria, an exploratory analysis was performed to identify possible covariates for each of the three sub-models. Then subjects with missing values for any of these covariates were further excluded from the analysis. Different models were fitted, including different covariates in different sub-models. The maximum likelihood estimates were obtained using a modified iterative Marquardt algorithm. The most appropriate model was then selected according to the BIC, the goodness of fit in terms of posterior class membership and according to the predictive ability. Finally, the selected model can be used to give individual dynamic predictions.*

2.3 Data Cleaning

The initial data-set contained information on 946 subjects. After application of some exclusion criteria, 753 subjects were retained for analysis. The following subjects were excluded:

- subjects with a stillbirth
- subjects with a second trimester miscarriage
- subjects with a high uncertainty on their *last menstrual period* (LMP) (certainty-sore < 8)

Further, one subject (E1609) was further removed from the data-set due to some outlying observations with very low values for the *crown rump length* (CRL) and the *mean sac diameter* (MSD) at very high gestational ages. The initial data-set contained information on 63 variables. Variables giving information on the baby itself, such as the gender of the baby or the weight of the baby at birth were not included in our analysis. Further, to avoid correlation between different variables, some additional variables were excluded from our analysis. For example, it was decided to only include the BMI of the mother

and not to include the weight and height of the mother. An exploratory analysis was then performed on the resulting variables to see which variables might be of importance to predict the risk for a miscarriage. The variables that seemed to differ between the two groups, i.e. the group of miscarriages and the group of viable pregnancies, were eventually included in the initial models and are shown in table 1.

Table 1: Variables included in the initial models.

Variable	Included as
Maternal age	covariate in all three sub-models
Paternal age	covariate in all three sub-models
Smoking status	covariate in all three sub-models
Fetal heartbeat at scan 1	covariate in all three sub-models
Amnion at scan 1	covariate in all three sub-models
PUQE-score at scan 1	covariate in all three sub-models
Crown rump length (CRL)	outcome variable in the linear mixed model
Mean sac diameter (MSD)	outcome variable in the linear mixed model
Gestational age (GA)	time variable in the linear mixed model
Pregnancy outcome	event of interest in the proportional hazards model

For the estimation of the models, subjects with missing values for any of the covariates included in the model, except for the time variable, were further excluded from the analysis, thereby making the number of subjects used for estimation model-dependent. In the CRL- and MSD-models, 623 and 620 subjects were used for the estimation of the model respectively.

2.4 Joint Latent Class Modelling Framework

A joint latent class model is used to model the time until a miscarriage, taking into account the heterogeneity within the population as well as the correlation between the longitudinal process of the repeated measures and the survival process itself. A joint latent class model consists of three sub-models, i.e. a multinomial logistic regression model for latent class-membership, a linear mixed model for the longitudinal variable repeatedly measured over time and a proportional hazards model for the time until the event [17, 19]. In this study, the event of interest modeled by the proportional hazards model is a miscarriage and the longitudinal process modelled by the linear mixed model is the growth of the fetus, expressed in terms of CRL or MSD. Theory for all three sub-models as well as for the estimation of the resulting joint model parameters, is discussed below. All joint latent class models were fitted using the function 'jointlcmm' of the R-package 'lcmm' that was recently developed by Proust-Lima et al. [17].

2.4.1 Multinomial Logistic Regression Model

In general, latent class models assume that the population under study is heterogeneous. In case of the EPOS data-set, it was assumed that there are (at least) two latent classes, i.e. one group of women where the pregnancy will result in a first trimester miscarriage

and another group of women giving birth to a healthy child after a full term pregnancy. The probability to belong to one of these latent classes can be described by a binomial (or multinomial) logistic regression model within the joint latent class model. For G latent classes, the class-membership for subject i is given by a discrete random and latent variable c_i that equals g if subject i belongs to latent class g ($g = 1, \ldots, G$). The probability for the latent variable c_i is given by a multinomial logistic regression model according to the covariates x_{ci}:

$$\pi_{ig} = P(c_i = g|X_{ci}) = \frac{e^{\xi_{0g} + X_{ci}^\top \xi_{1g}}}{\sum_{l=1}^{G} e^{\xi_{0l} + X_{ci}^\top \xi_{1l}}} \tag{2.1}$$

with ξ_{0g} the intercept for class g and ξ_{1g} the vector of class-specific parameters associated with the time-independent covariates X_{ci}. The intercept ξ_{0G} and the vector of parameters ξ_{1G} associated with the reference class G should be equal to zero for identifiability [17,19].

2.4.2 Linear Mixed Model

Within each of the G latent classes, the longitudinal mean profile of the repeatedly measured variable is modelled by a class-specific linear mixed model, where the fixed effects as well as the distribution of the random effects can be class-specific. Within latent class g, the repeatedly measured Gaussian outcome variable is modelled as proposed by Laird and Ware [20]:

$$Y_{ij}|_{c_i=g} = X_{L1i}(t_{ij})^\top \beta + X_{L2i}(t_{ij})^\top v_g + Z_i(t_{ij})^\top u_{ig} + \epsilon_{ij} \tag{2.2}$$

where the vector X_{L1i} is associated with the fixed effects β that are common over the different latent classes and the vector X_{L2i} is associated with the class-specific fixed effects v_g. The vector Z_i is associated with the subject-specific random effects u_{ig} that follow a class-specific, zero-mean multivariate normal distribution with variance-covariance matrix $\omega_g^2 B$. Here, B is an unspecified variance-covariance matrix and ω_g is a proportionality coefficient allowing for class-specific variability of the subjects. For identifiability, the proportionality coefficient ω_G of the reference class should be equal to 1. The vector of parameters for modelling the symmetric positive definite variance-covariance matrix B is denoted by $vec(B)$. The measurement errors are given by ϵ_{ij} and are assumed to follow a normal distribution with variance σ_ϵ^2 and mean zero [17, 19]. Within this study, two measures of the growth of the fetus, namely the *crown rump length* (CRL) and the *mean sac diameter* (MSD) were included as outcome variables in the linear mixed model.

2.4.3 Proportional Hazards Model

In survival analysis, the true event time for subject i is denoted by $T^*{}_i$, the censoring time is denoted by $\tilde{T}_i$ and $T_i = min(T^*{}_i, \tilde{T}_i)$. The indicator variable E_i is equal to one if $T^*{}_i < \tilde{T}_i$ and is zero otherwise, indicating non-censored and censored observations respectively. The probability for an event to take place in a small interval after time t, provided the event has not occurred before t, is then given by the instantaneous failure rate or the hazard function which is defined as:

$$\lambda(t) = \lim_{h \to 0} \frac{P(t \leq T^* < t + h | T^* \geq t)}{h}. \tag{2.3}$$

The cumulative hazard function, defining the sum of risks up to time t is given by the integral of the hazard function over time:

$$\Lambda(t) = \int_0^t \lambda(x)dx \tag{2.4}$$

For subject i within latent class g, the hazard function at time t can be modelled by a class-specific Cox's regression model or proportional hazards model as:

$$\lambda_i(t)|_{c_i=g} = \lambda_{0g}(t)e^{X_{Si1}^\top \nu + X_{Si2}^\top \delta_g} \tag{2.5}$$

where the vector X_{Si1} is associated with the parameters v that are common over classes and the vector X_{Si2} is associated with the class-specific parameters δ_g. The baseline hazard λ_{0g}, parametrized by vector ζ_g, is class-specific and is either stratified on the latent class structure ($\lambda_{0g}(t) = \lambda_0(t;\zeta_g)$) or proportional in each latent class ($\lambda_{0g}(t) = \lambda_0(t;\zeta^*)e^{\zeta_g}$ with ζ^* the parameter vector defining $\lambda_0(t)$, with e^{ζ_g} a proportionality factor and $\zeta_G = 0$) [17, 19]. In this study, the baseline hazard function was either modelled with cubic M-splines as $\lambda_0(t;\zeta) = \sum_{l=1}^{n_z+2} \zeta_l M_l(t)$ where n_z denotes the number of knots and $(M_l(t))_{l=1,\ldots,n_z+2}$ denotes the basis of the M-splines [16] or with a Weibull function as $\lambda_0(t;\zeta) = \zeta_1\zeta_2(\zeta_1 t)^{(\zeta_2-1)}$, where parameters were restricted to be positive in both functions [17, 19].

2.4.4 Assumption of Conditional Independence

Given the latent classes, the longitudinal process and the survival process are assumed to be conditionally independent. In other words, it is thus assumed that the latent class structure captures the entire dependency between the longitudinally measured variable and the event-time and that there is no dependency through the random effects. The assumption of conditional independence can be tested using a score test as described by JacqminGadda et al. [21], with the alternative hypothesis that there is dependency between the event-time and the longitudinal variable through the random effects. Under the alternative hypothesis, the joint latent class model would thus contain shared random effects and the proportional hazards model described in equation (2.5) would then be:

$$\lambda_i(t)|_{c_i=g} = \lambda_{0g}(t)e^{X_{Si1}^\top \nu + X_{Si2}^\top \delta_g + u_{ig}\eta} \tag{2.6}$$

where η is a vector of length p associating the p random effects u_{ig} of the linear mixed model shown in equation (2.2) to the proportional hazards model. If the conditional independence assumption is satisfied, we would have that $\eta = 0$. Thus the null-hypothesis is that $\eta = 0$ while the alternative hypothesis is that $\eta \neq 0$.
This can be tested with a score test where the score U is given by:

$$U(\eta,\theta) = \frac{\partial l_a}{\partial \eta} \tag{2.7}$$

with l_a the log-likelihood under the alternative hypothesis. The score U is calculated for $\eta = 0$.

This derivative has been worked out by Jacqmin Gadda et al. [21] resulting in the following expression for U

$$U(0,\theta) = \sum_{i=1}^{N} \sum_{g=1}^{G} \hat{\pi}_{ig} \left(E_i - \Lambda_{ig} \left(T_i \right) \right) \hat{u}_{ig} \tag{2.8}$$

where $\hat{\pi}_{ig}$ is the probability of belonging to latent class g as described in equation (2.1), E_i is equal to one if the time-to-event T_i^* is smaller than the censoring time $\tilde{T}_i$ and is zero otherwise and where $\Lambda_{ig}(T_i)$ is the class-specific cumulative hazard. The score U can be interpreted as an estimate for the covariance between the empirical Bayes estimates of the random effects and the residuals from the survival model, weighted by $\hat{\pi}_{ig}$. The score test statistic is now given by $S = U^T Var(U)^{-1} U$, which follows a χ^2-distribution with p degrees of freedom under the null-hypothesis of conditional independence. A high score test statistic favors the alternative hypothesis, while a score test statistic close to zero favors the null-hypothesis of conditional independence [19, 21].

2.5 Exploratory Data Analysis

An exploratory data analysis, including descriptive and visual summary statistics, was performed in order to decide which covariates to include in which sub-models of the joint latent class models. The time variable used in all joint latent class models was the gestational age of the fetus in days, as defined by the last menstrual period (LMP) of the mother. Because the exact date of the LMP might sometimes be hard to recall, all subjects were asked to give a score from 0 to 10 on their certainty about their LMP. In order to have reliable data on the gestational age of the fetus, only subjects with a certainty score above 7 were included for analysis. The results from the exploratory data analysis are given in section 3.1.

2.5.1 Multinomial Logistic Regression Model

The averages of all variables included in the EPOS data-set were calculated for the group of women experiencing a miscarriage and for the group of women with a viable pregnancy in order to detect important covariates that might discriminate between the two groups. Further, the survival functions based on the Kaplan-Meier estimates were plotted for different values of the baseline and longitudinal covariates. (Quasi)continuous covariates were first discretized according to their median values in order to construct these plots.

2.5.2 Linear Mixed Model

The longitudinally measured variables CRL and MSD are the outcome variables modelled in the linear mixed model parts of the joint latent class models. A log transformation ($log(CRL+1)$) was applied to CRL to make its distribution more normal. A transformation of MSD was not needed. The time variable included in the linear mixed models is the gestational age of the fetus in days, as defined by the last menstrual period of the mother. To get an impression of the evolutions of CRL and MSD over time, their mean values were calculated at all scan times within the two groups, together with the mean gestational ages of the fetuses. However, since the gestational ages of different subjects at one scan

time were not the same, comparison between the two groups based on the average CRL and MSD values is difficult. We therefore also plotted the subject-specific evolutions for CRL, logCRL and MSD. At last, in order to see which covariates might influence the evolution of CRL and MSD over time, the longitudinal individual evolutions were plotted for different values of the covariates. For example, the individual evolutions of CRL for mothers younger and older than 32 years were plotted to investigate the impact of the maternal age on the growth of the fetus.

2.5.3 Proportional Hazards Model

In order to decide which baseline covariates and which longitudinal variables measured at the first scan should be included into the proportional hazards model, the survival functions for different values of the covariates were plotted. (Quasi)continuous covariates were first discretized according to their median values in order to investigate their influence on the survival process.

2.6 Estimation of the Joint Latent Class Model Parameters

Maximum likelihood estimation was used to estimate the parameters in the resulting joint latent class model. The vector of parameters to be estimated is denoted by θ_G and includes the parameters involved in the multinomial logistic regression model, in the class-specific linear mixed model and in the class-specific proportional hazards model. An overview of all parameters that have to be estimated in a joint latent class model is given in table 2.

Table 2: Parameters that have to be estimated in the joint latent class model.

Parameter	**Explanation**
Multinomial logistic regression model	
$(\xi_{0g})_{g=1,G-1}$	class-specific intercept for class g
$(\xi_{1g}{}^{\top})_{g=1,G-1}$	parameters for class g
Linear mixed model	
$\beta^{\top}$	fixed effects common over the latent classes
$(v_g)^{\top}_{g=1,G}$	class-specific fixed effects for class g
$vec(B)^{\top}$	parameters involved in the variance-covariance matrix B
$(\omega_g)_{g=1,G-1}$	proportionality coefficient for class g allowing class-specific variability
σ^2_{ϵ}	variance for the measurement error ϵ
Proportional hazards model	
$\nu^{\top}$	parameters common over the latent classes
$(\delta_g{}^{\top})_{g=1,G}$	class-specific parameters for class g
$(\zeta_g{}^{\top})_{g=1,G}$	parameters involved in the class-specific baseline hazard λ_{0g}

For the resulting joint latent class model, the individual contribution to the likelihood is given by:

$$L_i(\theta_G) = \sum_{g=1}^{G} \pi_{ig} \phi_{ig}(Y_i|c_i = g; \theta_G) e^{-\Lambda(T_i|c_i=g;\theta_G)} \lambda_i(T_i|c_i = g; \theta_G)^{E_i} \tag{2.9}$$

with π_{ig} the class-membership probability as defined in equation (2.1), with ϕ_{ig} a multivariate normal density function with mean $\mu_{ig} = X_{L1i}\beta + X_{L2i}v_g$ and variance $V_{ig} = Z_i B_g Z_i^\top + R_i + \Sigma_i$ as given by equation (2.2) and with $\lambda_i(T_i|c_i = g; \theta_G)$ the instantaneous hazard as defined in equation (2.5) with corresponding cumulative hazard $\Lambda_i(T_i|c_i = g; \theta_G)$. The log-likelihood is now given by [17]:

$$l(\theta_G) = \sum_{i=1}^{N} \log(L_i(\theta_G)). \tag{2.10}$$

The resulting maximum likelihood estimates correspond to the parameter values that maximize the log-likelihood function given in equation 2.10.

To obtain maximization of the log-likelihood, Proust-Lima et al. [22] modified the Marquardt algorithm, i.e. an iterative algorithm belonging to the Newton-Raphson family initially proposed by Marquardt et al. [23]. The modified Marquardt algorithm updates the parameter vector θ_G using the following equation at iteration $l + 1$:

$$\theta_G^{(l+1)} = \theta_G^{(l)} - \delta \left(\tilde{\mathcal{H}}^{(l)}\right)^{-1} \nabla \left(L\left(\theta_G^{(l)}\right)\right) \tag{2.11}$$

where $\delta = 1$ by default and can be modified at each iteration to ensure that the log-likelihood is improved. Further, $\tilde{\mathcal{H}}$ is the Hessian matrix with potentially inflated diagonal terms $\tilde{\mathcal{H}}_{ii}$ to ensure positive definiteness. The inflated diagonal terms $\tilde{\mathcal{H}}_{ii}$ are defined as $\mathcal{H}_{ii} + \lambda[(1-\eta)|\mathcal{H}_{ii}| + \eta \operatorname{tr}(\mathcal{H})]$ where λ and η are initially fixed at 0.01. These values for λ and η are reduced if $\mathcal{H}$ is positive-definite and are increased if $\mathcal{H}$ is not positive-definite. Further, $\nabla(L(\theta_G^{(l)})$ is the gradient of the log-likelihood at iteration l. The Marquardt algorithm is then reiterated until convergence as defined by the following three convergence criteria based on:

- parameter stability: $\sum_{j=1}^{n_\theta} (\theta_G(j)^{(l)} - \theta_G(j)^{(l-1)})^2 \leq \epsilon_a$
- log-likelihood stability: $|L^{(l)} - L^{(l-1)}| \leq \epsilon_b$
- the size of the derivatives: $\frac{\nabla(L(\theta_G^{(l)}))^\top \mathcal{H}^{(l)-1} \nabla(L(\theta_G^{(l)}))^\top}{n_\theta} \leq \epsilon_c$

with n_θ the length of the parameter vector θ_G. The default values for ϵ_a, ϵ_b and ϵ_c are equal to 10^{-4}. Note that although these thresholds might seem relatively large, the convergence criterion based on the derivatives is very stringent. Besides, because in complex settings such as the joint latent class model the log-likelihood might be relatively flat in some areas of the parameter space, all three convergence criteria must be met simultaneously. In this way, good convergence to the actual maximum of the log-likelihood is ensured and convergence to a local maximum is avoided [17, 19]. In order to further reduce the risk of converging to a local maximum, it is advised to run the algorithm multiple times

with different starting values [24, 25]. Further, an estimate for the variance-covariance matrix $V(\hat{\theta}_G)$ of the maximum likelihood estimates is given by the inverse of the Hessian matrix $\tilde{\mathcal{H}}$ [17, 19]. At last, after estimating the parameters for multiple models, each with a different number of latent classes, the optimal number of latent classes can be chosen according to the BIC:

$$BIC = -2L\left(\theta_G\right) + n_\theta \log(N) \tag{2.12}$$

where N denotes the sample size. The BIC value more often gives the correct number of latent classes as compared to *Akaike's information criterion* (AIC) [26].

2.7 Post-Fit Computations

After conduction of the exploratory analysis and after the estimation of the parameter vector θ_G, several post-fit computations can be performed such as the calculation of the posterior probability for class-membership, the calculation of the predicted profiles for the longitudinal variable, the prediction of the event of interest and the assessment of the predictive accuracy of the fitted model.

2.7.1 Posterior Class-Membership Probabilities and Posterior Classification

Posterior class-membership probabilities can be used to determine the posterior classification of subjects, by assigning each subject to the latent class for which they have the highest posterior class-membership probability ($\hat{c}_i = argmax_g(\hat{\pi}_{ig}^{(Y,T)})$). The posterior classification of subjects can then be used to assess the goodness-of-fit of the joint latent class model and to evaluate how well the model can discriminate between the different latent classes. The posterior class-membership probabilities are computed using Bayes theorem, given the already collected information, by:

$$\begin{aligned}\hat{\pi}_{ig}^{(Y,T)} &= P(c_i = g | X_{ci}, X_{Li}, X_{Si}, Y_i, T_i, E_i, \hat{\theta}_G) \\ &= \frac{\hat{\pi}_{ig}\phi_{ig}(Y_i|c_i = g; \hat{\theta}_G)e^{-\Lambda(T_i|c_i=g;\hat{\theta}_G)}\lambda(T_i|c_i = g; \hat{\theta}_G)^{E_i}}{\sum_{l=1}^{G}\pi_{il}\phi_{il}(Y_i|c_i = l; \hat{\theta}_G)e^{-\Lambda(T_i|c_i=l;\hat{\theta}_G)}\lambda(T_i|c_i = l; \hat{\theta}_G)^{E_i}}.\end{aligned} \tag{2.13}$$

The posterior classification table as defined in table 3 can be used to assess the goodness-of-fit and the discriminating power of the model. It shows the mean of the posterior probabilities of belonging to a latent class among the subjects classified *a posteriori* in that latent class. In case of perfect classification, the posterior classification table displays ones on the diagonal and zeros elsewhere [17, 19].

Table 3: Definition for the posterior classification table.

Final class $\hat{c}_i$	**Mean of the probabilities of belonging to class:** **1**	$\cdots$	**g**	$\cdots$	**G**
1	$\frac{1}{N_1}\sum_{i=1}^{N_1}\hat{\pi}_{i1}^{(Y,T)}$	$\cdots$	$\frac{1}{N_1}\sum_{i=1}^{N_1}\hat{\pi}_{ig}^{(Y,T)}$	$\cdots$	$\frac{1}{N_1}\sum_{i=1}^{N_1}\hat{\pi}_{iG}^{(Y,T)}$
$\vdots$	$\vdots$	$\ddots$			$\vdots$
g	$\frac{1}{N_g}\sum_{i=1}^{N_g}\hat{\pi}_{i1}^{(Y,T)}$	$\cdots$	$\frac{1}{N_g}\sum_{i=1}^{N_g}\hat{\pi}_{ig}^{(Y,T)}$	$\cdots$	$\frac{1}{N_g}\sum_{i=1}^{N_g}\hat{\pi}_{iG}^{(Y,T)}$
$\vdots$	$\vdots$			$\ddots$	$\vdots$
G	$\frac{1}{N_G}\sum_{i=1}^{N_G}\hat{\pi}_{i1}^{(Y,T)}$	$\cdots$	$\frac{1}{N_G}\sum_{i=1}^{N_G}\hat{\pi}_{ig}^{(Y,T)}$	$\cdots$	$\frac{1}{N_G}\sum_{i=1}^{N_G}\hat{\pi}_{iG}^{(Y,T)}$

2.7.2 Longitudinal Predictions

Class-specific predictions of the evolutions of the longitudinal variable can be calculated from the fitted model, both on the marginal level, as well as on the subject-specific level. The marginal predictions can be used when interest lies in the average evolution of a population of subjects with the same covariate values while the subject-specific predictions are used when one is interested in the evolution of a specific individual from the EPOS data-set. The random effects then show how the individual is deviating from its population average. For subject i, occasion j and class g, the predicted marginal evolution of the longitudinally measured variable is given by:

$$\hat{Y}_{ijg}^{(M)} = X_{L1i}(t_{ij})^\top \hat{\beta} + X_{L2i}(t_{ij})^\top \hat{v}_g \tag{2.14}$$

where the vector X_{L1i} is associated with the fixed effects β that are common over the latent classes and the vector X_{L2i} is associated with the class-specific fixed effects v_g. For subject i, occasion j and class g, the prediction of a subject-specific longitudinal trajectory is given by:

$$\hat{Y}_{ijg}^{(SS)} = X_{L1i}(t_{ij})^\top \hat{\beta} + X_{L2i}(t_{ij})^\top \hat{v}_g + Z_i(t_{ij})^\top \hat{u}_{ig} \tag{2.15}$$

where the vector Z_i is associated with the random effects u_{ig}. The marginal predictions can also be used to calculate and plot the predicted evolution of the longitudinal outcome variable for a fictive future subject with a hypothetical profile of covariates.

From these class-specific individual predictions, weighted class-specific predictions averaged over the individuals can be computed as:

$$\hat{Y}_{jg}^{(M)} = \sum_{i=1}^{N(j)} \hat{\pi}_{ig} \hat{Y}_{ijg}^{(M)} \tag{2.16}$$

and

$$\hat{Y}_{jg}^{(SS)} = \sum_{i=1}^{N(j)} \hat{\pi}_{ig}^{(Y,T)} \hat{Y}_{ijg}^{(SS)} \tag{2.17}$$

with $N(j)$ the number of subjects with measurements at occasion j, with $\hat{\pi}_{ig}$ the class membership probability as defined in equation 2.1 and with $\hat{\pi}_{ig}^{(Y,T)}$ the conditional class membership probability as defined in equation 2.13. In practice, time intervals are constructed to calculate these expressions. [17, 19].

2.7.3 Residual Analysis

The class-specific marginal and subject-specific longitudinal predictions $\hat{Y}_{ijg}^{(M)}$ and $\hat{Y}_{ijg}^{(SS)}$ as described in section 2.7.2 can be used to calculate the respective residuals. The class specific predictions were first averaged over the latent classes as:

$$\hat{Y}_{ij}^{(M)} = \sum_{g=1}^{G} \hat{\pi}_{ig} \hat{Y}_{ijg}^{(M)}$$

and

$$\hat{Y}_{ij}^{(SS)} = \sum_{g=1}^{G} \hat{\pi}_{ig}^{(Y,T)} \hat{Y}_{ijg}^{(SS)} \tag{2.18}$$

again with with $\hat{\pi}_{ig}$ the class membership probability as defined in equation 2.1 and with $\hat{\pi}_{ig}^{(Y,T)}$ the conditional class membership probability as defined in equation 2.13. The marginal and subject-specific residuals are then respectively given by:

$$R_{ij}^{(M)} = Y_{ij} - \hat{Y}_{ij}^{(M)}$$

and

$$R_{ij}^{(SS)} = Y_{ij} - \hat{Y}_{ij}^{(SS)}. \tag{2.19}$$

These residuals can be used to asses the goodness-of-fit of the joint latent class model by plotting a normal QQ-plot or by plotting the residuals against the longitudinal predictions [17].

2.7.4 Prediction of the Event of Interest

According to the fitted joint latent class model, predicted class-specific survival functions and individual dynamic predictions can be calculated and plotted. An individual dynamic prediction is defined as the predicted probability for a miscarriage to occur within the time-frame $(s, s+t)$, according to the information collected for subject i up to time s. We call s and t the landmark time and the horizon of the prediction respectively. Further, if we define:

- $Y_i^{(s)} = Y_{ij}, j = 1, \ldots, n_i,$ such that $t_{ij} \leq s$
  (= values of the longitudinally measured variable for subject i up to time s)
- X_{ci}
  (=values for the covariates in the multinomial regression model)
- $X_i^{(s)} = X_{L1i}(t_{ij}), X_{L2i}(t_{ij}), Z_i(t_{ij}), j = 1, \ldots, n_i,$ such that $t_{ij} \leq s$
  (= values of the covariates included in the linear mixed model for subject i up to time s)
- $X_{Si} = X_{Si1}, X_{Si2}$
  (= values of the covariates included in the proportional hazards model for subject i)

then the individual dynamic prediction is given by:

$$P(T_i \leq s+t|T_i \geq s, Y_i^{(s)}, X_i^{(s)}, X_{Si}, X_{ci}; \theta_G)$$

$$= \frac{\sum_{g=1}^{G} \pi_{ig} \left(S_i\left(s|X_{Si}, c_i = g; \theta_G\right) - S_i\left(s+t|X_{Si}, c_i = g; \theta_G\right)\right) \phi_{ig}(Y_i^{(s)}|X_i^{(s)}, c_i = g; \theta_G)}{\sum_{g=1}^{G} \pi_{ig} S_i(s|X_{Si}, c_i = g; \theta_G) \phi_{ig}(Y_i^{(s)}|X_i{}^{(s)}, c_i = g; \theta_G)} \tag{2.20}$$

with $\phi_{ig}(Y_i^{(s)}|X_i^{(s)}, c_i = g; \theta_G)$ the density of the longitudinal outcomes in class g, with π_{ig} the class-specific membership probability and with $S_i(s|X_{Si}, c_i = g; \theta_G)$ the class-specific survival function [17].

2.7.5 Assessment of Predictive Accuracy

The predictive accuracy of the fitted joint latent class models can be assessed using the *expected prognostic observed cross-entropy* (EPOCE) which is defined as:

$$EPOCE = E\left(-\ln f_{T|Y^{(s)}, T^* \geq s}|T^* \geq s\right) \tag{2.21}$$

where $f_{T|Y^{(s)}, T^* \geq S}$ is the conditional density of the event time as derived from the joint latent class model. The EPOCE is estimated using cross-validation, i.e. estimating the log-likelihood N_s times, each time leaving out one of the N_s observations still at risk, then calculating the EPOCE N_s times and finally computing the average EPOCE. The resulting EPOCE estimate is called the *cross-validated prognosis observed loss* (CVPOL). However, because cross-validation is computationally demanding, the CVPOL is approximated by the CVPOL$_a$ at a fixed time s as:

$$CVPOL_{a(s)} = -\frac{1}{N_s} \sum_{i=1}^{N_s} F_i(\hat{\theta}, s) + N \text{ Trace}\left(\mathcal{H}^{-1} K_s\right) \tag{2.22}$$

where N_s is the number of subjects still at risk at time s. $\mathcal{H}$ is the Hessian matrix of the joint log-likelihood. $K_s = \frac{1}{N_s(N-1)} \sum_{i=1}^{N} I\left(T_i \geq s\right) \hat{v}_i(s) \hat{d}_i^T$ with $\hat{v}_i(s)$ the gradients of the individual contributions to the conditional log-likelihood at time s using only $Y_i^{(s)}$, and with $\hat{d}_i$ the gradients of the individual contributions to the joint log-likelihood computed in $\hat{\theta}$ using the total vector of repeated measures Y_i. F_i is the individual contribution to the conditional log-likelihood which is given by:

$$F_i\left(\theta_G, s\right) = \ln\left(\frac{\sum_{g=1}^{G} \pi_{ig} \phi_{ig}\left(Y_i^{(s)}|c_i = g; \theta_G\right) \lambda_i\left(T_i|c_i = g; \theta_G\right)^{E_i} S_i\left(T_i|c_i = g; \theta_G\right)}{\sum_{g=1}^{G} \pi_{ig} \phi_{ig}\left(Y_i^{(s)}|c_i = g; \theta_G\right) S_i\left(s|c_i = g; \theta_G\right)}\right) \tag{2.23}$$

For a vector of different landmark times s, the predictive accuracy measure CVPOL$_a$ is thus calculated based on the subjects still at risk and the collected information up to a given landmark time. A lower value for CVPOL$_a$ corresponds to a better predictive ability of the model [19, 27, 28].

Chapter 3

Results

3.1 Exploratory Data Analysis

As discussed in section 2.5 subjects with a LMP certainty score below 8 were excluded from the analysis. From the remaining 753 subjects, 82 women experienced a first trimester miscarriage, that is 10.89% of the women. The mean and median gestational ages at which miscarriages occurred were 65.46 and 64.00 days respectively, ranging from a minimum of 29 days to a maximum of 92 days.

3.1.1 Multinomial Logistic Regression Model

Table 4 shows the mean values for the baseline and longitudinal covariates included in the EPOS data-set. Because time-varying variables can not be included into to multinomial logistic regression sub-models, it was decided to only include the measurements taken at the first scan for longitudinally measured variables. It can already be seen from this table that the average maternal and paternal age within the group of miscarriages is higher as compared to the group of viable pregnancies. Also, the average BMI and the percentage of smokers is larger within the group of miscarriages. On average, the worst bleeding score reported at the first scan is similar between the two groups, while the worst pain score and the PUQE-score reported at the first scan are higher for the group of viable pregnancies. At last, there seems to be a large discrepancy in the detection of a fetal heartbeat and the detection of the amnion between both groups. At the first scan, a fetal heartbeat is detected in 75.90% of the viable pregnancies and only in 41.46% of the pregnancies that will eventually end in a miscarriage. The amnion is detected in 18.32% of the viable pregnancies at the first scan, as opposed to only 7.41% in the group of the miscarriages.

Table 4: Mean values and percentages for baseline and longitudinally measured covariates of interest in both groups.

Baseline variables	**Miscarriage (range) (n=82)**	**Viable (range) (n=671)**
Maternal age	33.68 (22-43)	32.71 (18-48)
Paternal age	35.40 (22-52)	34.95 (19-53)
BMI	26.13 (18.52-43.44)	24.67 (15.84-46.44)
1st trimester miscarriages	1.16 (0-10)	0.709 (0-8)
Gravida	3.259 (1-14)	2.642 (1-11)
Para	0.7531 (0-6)	0.6319 (0-6)
Current alcohol consumption	0.0520 (0-2)	0.0815 (0.14)
Smokers	13.92 %	8.87 %
Surgery uterus	40.74%	38.75%
Progesterone intake	6.25%	7.154%
Metformin intake	0.00%	0.7452
Aspirin intake	5.00%	4.77%
Folic acid intake	92.50%	96.26%
Termination of pregnancy (TOP)	20.99%	16.54%
Ectopic pregnancy	7.41%	5.37%
Longitudinal variables		
Worst bleeding score	0.7805 (0-4)	0.7515 (0-4)
Worst pain score	1.987 (0-9)	2.461 (0-10)
PUQE score	3.635 (3-10)	4.539 (3-15)
Number of bleedings (SCH)	0.1358 (0-1)	0.1712 (0-2)
Fetal heartbeat detected	41.46 %	75.90 %
Amnion detected	7.41 %	18.32 %

Note that the summary measures for the longitudinal variables are calculated from the first scan only.

Figure 2 shows the survival functions based on the Kaplan-Meier estimates for different values of the baseline and longitudinal covariates. (Quasi)continuous covariates were first discretized according tot their median values in order to construct these plots. The values used for discretizing these variables are indicated in the respective plots. It can be seen that subjects where a fetal heartbeat was not detected at the first scan and subjects where the amnion was not detected at the first scan have a higher probability to miscarry. Also, subjects who reported a PUQE-score below or equal to 4 have a higher probability to miscarry. The risk for a miscarriage is higher with a maternal age above 32 and with a paternal age above 35. At last, the risk for a miscarriage is higher for smoking mothers. According to the mean values shown in table 4 and the survival plots shown in figure 2, it was decided to include the maternal age, the paternal age and the smoking status of the mother as baseline covariates. Further, the longitudinal variables measured at the first scan that seem to have an important role in the discrimination between the two groups are the detection of a fetal heartbeat, the detection of an amnion and the PUQE-score. These variables were therefore included in the initial multinomial logistic regression model.

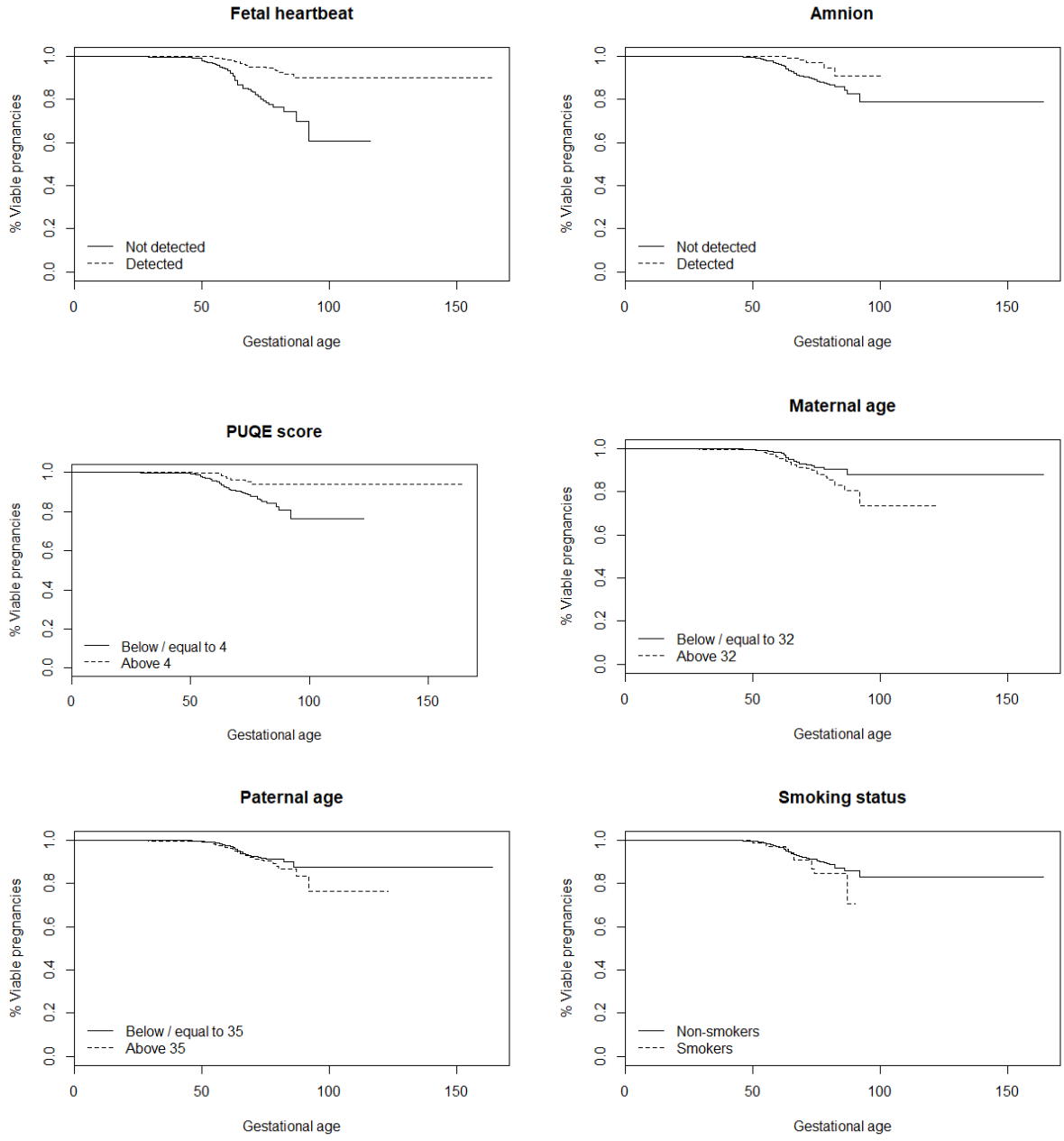

Figure 2: Survival functions for the important baseline and longitudinal covariates: *Survival functions for inclusion of baseline and longitudinal covariates in the multinomial logistic regression model and proportional hazards model. When a fetal heartbeat is not detected, when the amnion is not detected, when the PUQE-score is low, when the maternal or paternal age is high or when the mother smokes, the risk for a miscarriage is higher. Note that the survival functions for the longitudinal variables are based on data from the first scan only.*

3.1.2 Linear Mixed Model

Table 5 shows the mean values for CRL and MSD calculated at all scan times within the two groups, together with the mean gestational ages of the fetuses. Note again that the gestational ages are not the same for all subject at a specific scan time and that the

number of subjects is not the same at all scan times. From table 5, it can already be seen that the average values for CRL are lower in the miscarriage group than in the viable group at all scan times. The same is true for the average values of MSD, although this difference seems less prominent at the last scan. However, the mean gestational ages are also lower for the miscarriage group at all scan times. The observed differences between the CRL and MSD values of the two groups might thus be due only to the fact that the mean gestational ages are lower in the miscarriage group.

Table 5: The average CRL and MSD at all scan times for both groups.

	Outcome	**Scan 1**	**Scan 2**	**Scan 3**	**Scan 4**	**Scan 5**
Mean GA	Miscarriage	47.39	57.12	64.35	70.31	76.00
	Viable	52.37	62.87	71.18	76.86	80.33
CRL	Miscarriage	6.064	8.278	11.850	17.71	34.15
	Viable	13.39	21.25	30.88	37.8	40.50
MSD	Miscarriage	11.64	18.06	22.85	25.02	44.15
	Viable	21.02	33.06	40.95	44.38	45.42

We therefore also plotted the subject-specific evolutions for CRL, logCRL and MSD in figure 3, for 30 random subjects from both groups. Here, CRL and logCRL are showing increasing individual evolutions over time with mild curvature in both groups. The evolution of CRL is less steep in the group of miscarriages, although this effect is not that clear when looking at logCRL. The individual evolutions of MSD are also increasing but seem more linear. Again, the evolution seems less steep in the group of miscarriages. Based on these plots, we will therefore include a quadratic time effect as well as a main time effect in the linear mixed model for logCRL and only a main time effect will be included in the linear mixed model for MSD. In order to avoid computational difficulties due to the quadratic time effect in the linear mixed model for logCRL, the gestational age was divided by 10. Random effects for all effects of gestational age were included in the initial linear mixed models, together with random intercepts.

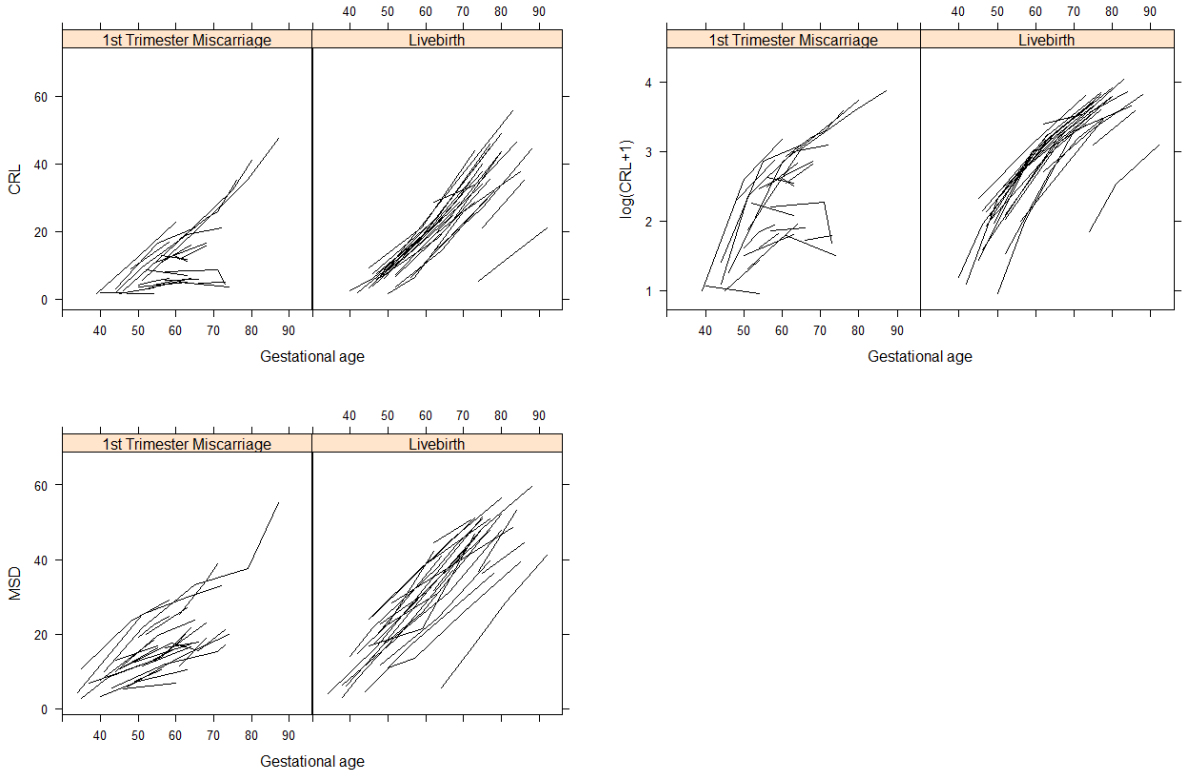

Figure 3: Individual longitudinal profiles for the evolution of CRL, logCRL and MSD within both groups: *The increasing evolutions of CRL and logCRL show mild curvature in both groups. The evolution of CRL/logCRL is less steep in the group of miscarriages. The increasing evolution of MSD seems more linear and is less steep in the group of miscarriages.*

In order to see which covariates might influence the evolution of CRL and MSD over time, the longitudinal individual evolutions were plotted for different values of the covariates. However, no obvious effects were found for any of the variables included in the EPOS data-set. It was therefore decided to only include the main effects of the variables identified in the previous section on the multinomial logistic regression model (section 3.1.1) as covariates in the linear mixed models.

3.1.3 Proportional Hazards Model

Figure 2 shows the survival functions for different values of some of the covariates included in the EPOS data-set. (Quasi)continuous covariates were first discretized according to their median values in order to construct these plots. The values used for discretizing the variables are indicated in these plots. According to this figure, it was decided to include the fetal heartbeat, amnion, PUQE-score, maternal age, paternal age and the smoking status in the initial proportional hazards model.

3.2 Fitted Models

Joint latent class models were fitted to the data, where in a first analysis the log transformed longitudinal outcome variable CRL was included in the linear mixed model part,

while in a second analysis the longitudinal outcome variable MSD was included. For both analyses, the initial model contains two latent classes and all covariates identified in the exploratory analysis (section 3.1). The covariates with a p-value above 0.50 were removed from the initial CRL- and MSD-models. Afterwards, the remaining covariates with a high non-significant p-value were removed in a sequential manner. Additional latent classes were then added in order to see whether the models could be improved. Further, except for the effect of gestational age in the linear mixed model, we assumed the effects of the covariates included in the linear mixed model and in the proportional hazards model to be the same over the latent classes. All models were fitted using M-splines with 3 equidistant knots as baseline hazard function as well as a with a Weibull function as baseline hazard function. For all models, lower BIC values were obtained when using a Weibull function as baseline hazard. Within this assay, we therefore only report on the models with a Weibull hazard function. At last, to avoid convergence to local maxima, the models were fitted using the *'gridsearch'* function from the 'lcmm' package. For all models, an automatic grid search was performed with 3 random sets of initial values and a maximum of 15 iterations. The resulting parameters that corresponded to the best log-likelihood were then used as initial values for the final estimation [24].

3.2.1 Models Including CRL

As stated above, a log transformation ($log(CRL + 1)$) was applied to CRL to make its distribution more normal. Because of the curvature observed in the longitudinal profiles of logCRL, all models include a quadratic time effect as well as a main time effect in the linear mixed model part. Further, since the gestational ages in the EPOS data-set can go up to 123 days, it was decided to use the gestational age divided by 10 (GA/10) as time variable to avoid computational difficulties caused by the quadratic term in the linear mixed model. Further, all models include a random intercept, a random slope for the time effect and a random slope for the quadratic time effect. The included covariates and the number of latent classes differ from model to model and a detailed overview of all fitted models is shown in table 6, indicating the number of latent classes and the covariates that were included in each of the sub-models.

Table 6: Overview of the fitted joint latent class models including CRL.

Covariates	**Sub-model**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Maternal age	Multinomial model	x						
	Mixed model	x	x	x	x	x	x	x
	Survival model	x	x	x	x	x	x	x
Fetal heartbeat	Multinomial model	x	x	x	x	x	x	x
	Mixed model	x	x	x	x	x	x	x
Smoking	Multinomial model	x	x					
	Survival model	x	x					
Paternal age	Survival model	x	x	x		x	x	x
Amnion	Survival model	x						
PUQE score	Multinomial model	x	x	x		x	x	x
	Latent classes	2	2	2	2	3	4	5
	BIC	582.5	573.7	561.9	804.6	471.16	448.62	411.4

The BIC value for the initial model including all covariates discussed in the exploratory analysis was 632.5.

Two latent classes

The initial model, containing the longitudinally measured outcome variable CRL, two latent classes and all covariates identified in the exploratory analysis, gave rise to a BIC value of 632.5. The covariates with the highest p-values were then sequentially removed from the model, resulting in model 1, model 2, model 3 and model 4 with respective BIC values of 582.5, 573.7, 561.9 and 804.6. Convergence was reached for all four models and the assumption of conditional independence could not be rejected in any of the models ($p_1 = 0.9026, p_2 = 0.8834, p_3 = 0.8614, p_4 = 0.0685$). According to these BIC values, model 3 appears to be the best two-class model and provides posterior classes 1 and 2 with proportions 92.46% and 7.54% respectively. Table 9 shows the proportion and number of subjects classified in class 1 and 2. 623 subjects, of which 51 (8.19%) experienced a miscarriage, were used for the estimation of CRL-model 3 and 24 parameters had to be estimated. The fetal heartbeat had a significant effect in the multinomial logistic regression model, the maternal age showed a significant effect in the proportional hazards model and both covariates had a significant impact in the linear mixed model. The parameter estimates and corresponding p-values for the model are given in table 23 in the appendix.

The posterior classification of subjects was used to assess the goodness-of-fit of the model and the discrimination between the two latent classes. Table 7 shows the proportion of subjects classified in each of the latent classes with a posterior probability above 0.7, 0.8 and 0.9. It can be seen that almost all subjects assigned to the first class were unambiguously classified since 99.65% of the subjects assigned to this class had a posterior probability above 0.9. Although the proportion of subjects unambiguously assigned to the second class is a bit lower as compared to the first class, high proportions are still obtained and 80.85% of the subjects assigned to this class had a posterior probability above 0.9. Table 8 shows the posterior classification table as defined in section 2.7.1. Again, it can be seen that the average probability of belonging to the first class, when classified in the first class, is higher as compared to the average probability of belonging to the second class, when classified in the second class. The respective probabilities are 0.9982 and 0.9471 for class 1 and class 2, indicating good discrimination between the two latent classes.

Table 7: Proportion of subjects classified in each of the latent classes with a posterior probability above 0.7, 0.8 and 0.9 for CRL-model 3.

	Class 1	**Class 2**
Probability > 0.7	100%	93.62%
Probability > 0.8	100%	89.36%
Probability > 0.9	99.65%	80.85%

Table 8: Posterior classification table for CRL-model 3.

	Probability 1	**Probability 2**
Class 1	0.9982	0.0018
Class 2	0.0529	0.9471

The goodness-of-fit of the model can be further investigated by looking at the obtained residuals for CRL-model 3. Although the marginal and subject-specific residuals seem to cluster around zero when plotted against the longitudinal predictions, a pattern can be observed for both the marginal and subject-specific residuals. Besides, the QQ-plots for the marginal and subject-specific residuals show deviation from the intersecting line. These results both indicate that the model does not fit the data very well and that an extension of the model might be needed.

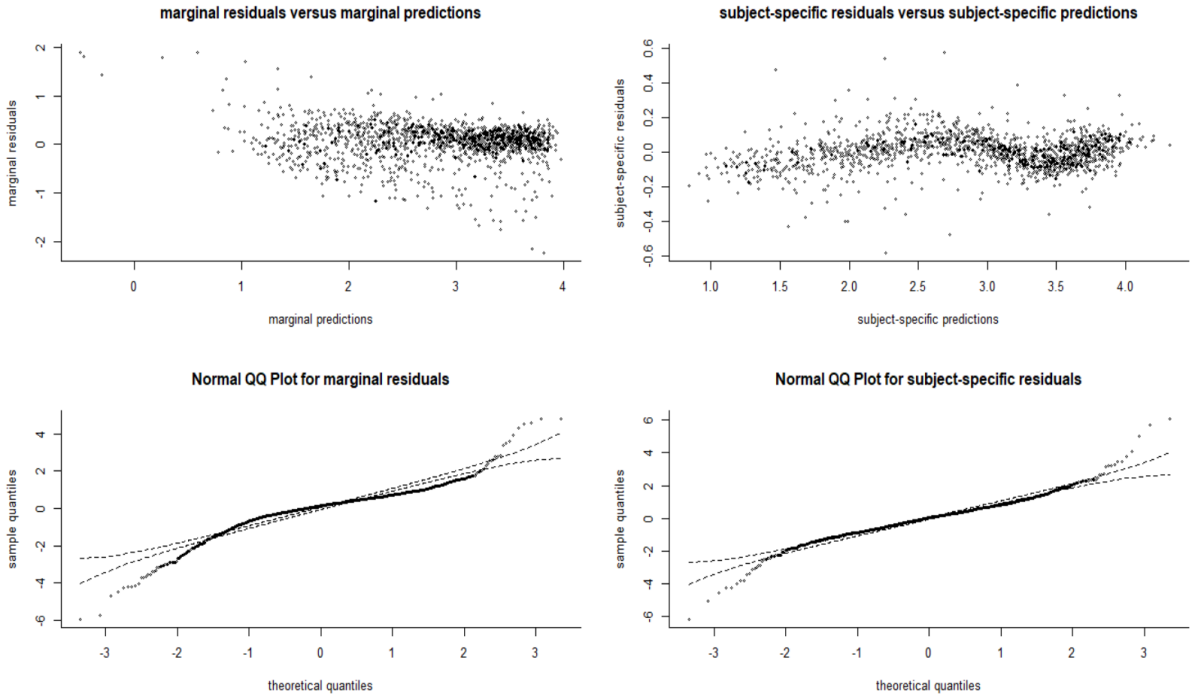

Figure 4: Residual plots for the assessment of the goodness-of-fit of CRL-model 3: *The marginal and subject-specific residuals are clustering around zero but patterns can be observed. The QQ-plots show deviation from the intersecting line, indicating that the model does not fit the data well.*

Figure 5 shows the observed and predicted marginal and subject-specific longitudinal profiles of logCRL as calculated from model 3. It can be seen that the predicted profiles fit the observed profiles very well. Besides, as was observed in the exploratory analysis in section 3.1.2, the evolution of logCRL seems to be less steep in class 2 as compared to class 1. According to these results, class 2 might correspond to the group of women who experienced a miscarriage while class 1 might correspond to the group of women who had a full term pregnancy.

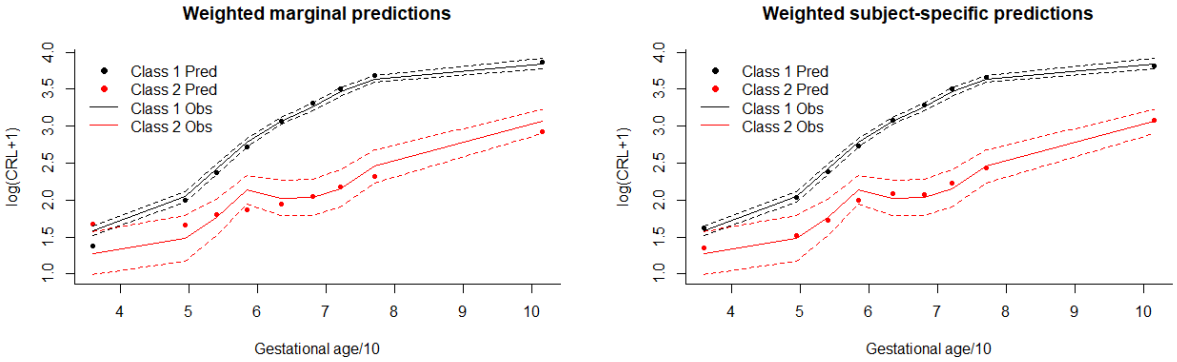

Figure 5: Comparison between observed and predicted marginal and subject-specific longitudinal profiles for log(CRL+1) calculated from model 3: *The predicted profiles fit the observed profiles well. The evolution of logCRL is less steep in class 2 than in class 1. Thus, class 2 might correspond to the group of women who experienced a miscarriage.*

Figure 6 shows the class-specific event-free probabilities over time for CRL-model 3. It can be seen that the probability for a miscarriage is higher in class 2 than in class 1 where the number of viable pregnancies remains stable during the whole first trimester. According to this figure, it can again be hypothesized that class 2 corresponds to the group of women who experienced a miscarriage while class 1 corresponds to the group of women with a viable pregnancy. However, if this truly would be the case, the survival function for class 2 should eventually decrease to zero, a characteristic that can not be observed from this figure. We can therefore conclude that a proportion of the subjects who still have a viable fetus after the first trimester is included in the group of subjects who experienced a miscarriage. Besides, because only a small fraction of the subjects was classified in class 2, this model is underestimating the true amount of miscarriages observed in our data-set. Therefore, more than 2 latent classes might be needed in order to find a good model.

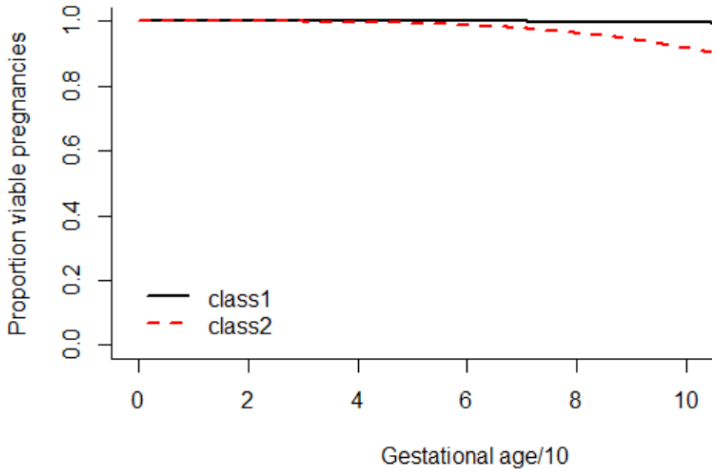

Figure 6: Class-specific event-free probability for CRL-model 3: *The probability for a miscarriage is higher in class 2 than in class 1. Thus, class 2 might correspond to the group of women who experienced a miscarriage. However, the survival function for class 2 is not decreasing to zero. We can therefore conclude that a proportion of the subjects who still have a viable fetus after the first trimester is included in the group of subjects who experience a miscarriage and that the true amount of miscarriages is underestimated.*

In order to see whether class 1 and class 2 indeed correspond to the group of viable pregnancies and the group of miscarriages respectively, the number of observed miscarriages were evaluated in both classes. Also, the average values for the variables included in CRL-model 3 were calculated for both classes to investigate how the latent classes differ from each other. Table 9 shows the proportion and number of subjects classified in each of the two classes, together with the number of subjects for which a miscarriage has been observed at the end of the first trimester and the average values for the variables included in CRL-model 3. The model contains 2 latent classes with probabilities of 92.46% for class 1 and 7.54% for class 2, corresponding to 576 and 47 subjects respectively. According to the longitudinal CRL-evolutions and the survival plots obtained from CRL-model 3, it was assumed that class 1 corresponds to the group of viable pregnancies while class 2 corresponds to the group of miscarriages. However, from the 47 subjects *a posteriori* classified into class 2, only 35 subjects have experienced a miscarriage, while 20 subjects classified in class 1 also experienced a miscarriage. It can thus be concluded that the proportion of subjects that experience a miscarriage is indeed higher in class 2 than in class 1. However, CRL-model 3 can not completely discriminate between women who do miscarry and women who do not miscarry. It can be seen that at the first scan, a fetal heartbeat was detected in 76.04% of the cases for class 1 and only in 48.94% of the cases for class 2. Also, the PUQE-score that was reported at the first scan is higher in class 1 than in class 2.

Table 9: Proportion and number of subjects assigned to the 2 latent classes for CRL-model 3.

	Class 1	**Class 2**
Proportion	92.46%	7.54%
Number of subjects	576	47
Number of final miscarriages	20	35
Maternal age	32.89	32.17
Paternal age	35.05	34.66
Fetal heartbeat at first scan	76.04%	48.94%
PUQE-score at first scan	4.531	3.681

More than 2 latent classes

Because of the improper results obtained for model 3, we then investigated whether this model could be further improved by including more than 2 latent classes, resulting in model 5, model 6 and model 7, with 3, 4 and 5 latent classes respectively. Again, convergence was reached for all of these models and the assumption of conditional independence could not be rejected for any of the models ($p_5 = 0.2014, p_6 = 0.1544, p_7 = 0.9137$). When fitting a model with more than 5 latent classes, convergence could no longer be reached. The model with 5 latent classes (model 7) resulted in the lowest BIC with a value of 411.4, as compared to BIC values of 471.16 and 448.62 for model 5 and model 6 with 3 and 4 latent classes respectively. We therefore selected model 7 as the final CRL-model, resulting in 5 latent classes with proportions of 4.82% for class 1, 87.80% for class 2, 3.05% for class 3, 1.28% for class 4 and 3.05% for class 5. These proportions and the corresponding number of subjects assigned to each of the 5 latent classes is given in table 12. Since the covariates included in model 7 are the same as those included in model 3, the same number of subjects was used for estimation of the model, i.e. 623 subject of which 51 (8.19%) experienced a miscarriage. For this model, 48 parameters had to be estimated. The maternal and paternal age showed a significant effect in the proportional hazards model and the fetal heartbeat and the maternal age had a significant impact in the linear mixed model. The parameter estimates and corresponding p-values for CRL-model 7 are given in table 24 in the appendix.

Again, the posterior classification of subjects was used to asses the goodness-of-fit of CRL-model 7 and to assess the discrimination between the corresponding 5 latent classes. Table 10 shows the proportion of subjects classified in each of the latent classes with a posterior probability above 0.7, 0.8 and 0.9 for model 7. It can be seen from this table that the model shows good discrimination between the 5 latent classes. The most confident classification of subjects can be done to class 2, since 98.72% of the subjects assigned to class 2 have a posterior probability above 0.90. The fact that such a high percentage of subjects can be unambiguously assigned to class 2 is very important since the majority of the subjects (87.80%) belong to this class. The percentage of subjects assigned to class 1 and class 4 with a probability above 0.90 is 83.33% and 87.50% respectively and subjects are thus assigned to these classes with good confidence. At last, the percentage of subjects assigned to class 3 and class 5 with a probability above 0.90 is only 68.42% and 63.16% respectively. However, when looking at the percentage of subjects assigned to these classes with a probability above 0.70, we can still state that assignment of subjects

to classes 3 and 5 is reliable. Table 11 shows the posterior classification table for model 7. Again, it can be seen from this table that model 7 can discriminate very well between the 5 latent classes. Indeed, the values on the diagonal all lie close to 1 with a minimum value of 0.8839 for class 5 and a maximum value of 0.9921 for class 2.

Table 10: Proportion of subjects classified in each of the latent classes with a posterior probability above 0.7, 0.8 and 0.9 for CRL-model 7.

	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Probability > 0.7	96.67%	99.45%	84.21%	100.0%	78.95%
Probability > 0.8	93.33%	99.27%	73.68%	87.5%	68.42%
Probability > 0.9	83.33%	98.72%	68.42%	87.5%	63.16%

Table 11: Posterior classification table for CRL-model 7.

	Probability 1	**Probability 2**	**Probability 3**	**Probability 4**	**Probability 5**
Class 1	0.9500	0.0025	0.0361	0.0114	0.0000
Class 2	0.0004	0.9921	0.0035	0.0001	0.0039
Class 3	0.0953	0.0265	0.8774	0.0008	0.0000
Class 4	0.0000	0.0000	0.0000	0.9715	0.0285
Class 5	0.0000	0.1155	0.0000	0.0005	0.8839

The goodness-of-fit for CRL-model 7 was further investigated using the residual plots shown in figure 7. Again, the residuals are clustered around zero and the same pattern as before can be recognized. However, the QQ-plots for CRL-model 7 show less deviation from the intersecting line as compared to the QQ-plots for CRL-model 3, thus indicating a model improvement.

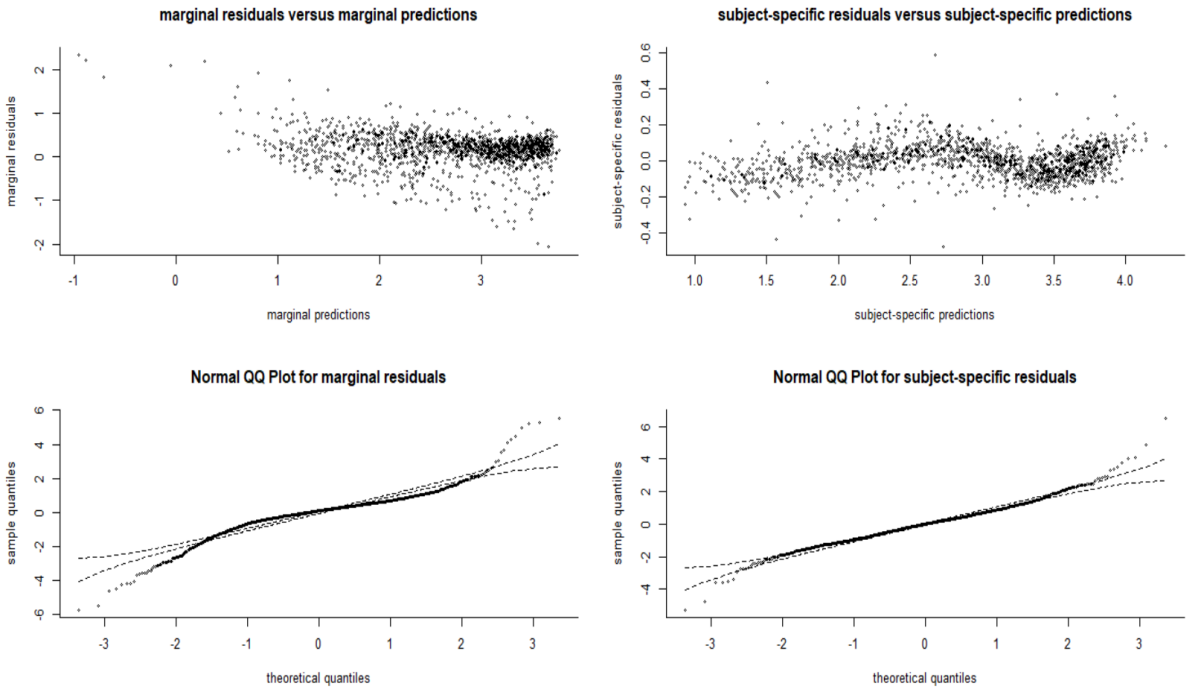

Figure 7: Residual plots for the assessment of the goodness-of-fit of CRL-model 7: *The residuals are clustered around zero and the same pattern as before can be recognized. The QQ-plots show less deviation from the intersecting line as compared to the QQ-plots for CRL-model 3.*

Figure 8 shows the observed and predicted marginal and subject-specific longitudinal profiles of logCRL as calculated from model 7. For class 2, the marginal predicted profile fits the observed profile very well. This is a good characteristic of the model since most of the subjects belong to this class. Indeed, it can be seen from this figure that the 95% confidence bands for the longitudinal profile in class 2 are very narrow, indicating good precision. Also, the marginal predicted profile for class 3 fits the observed profile very well. The marginal predicted longitudinal profiles in class 1 and 4 seem to fit the observed profiles reasonably well while the marginal predicted profile within class 5 clearly deviates from the observed profile at small gestational ages. When looking at the subject-specific evolutions, the observed and predicted profiles are corresponding well within all 5 classes. We can thus state that CRL-model 7 fits our data well. According to figure 8, class 2 shows the highest logCRL values at all times. It can therefore be hypothesized that class 2 corresponds to the group of women with a viable pregnancy.

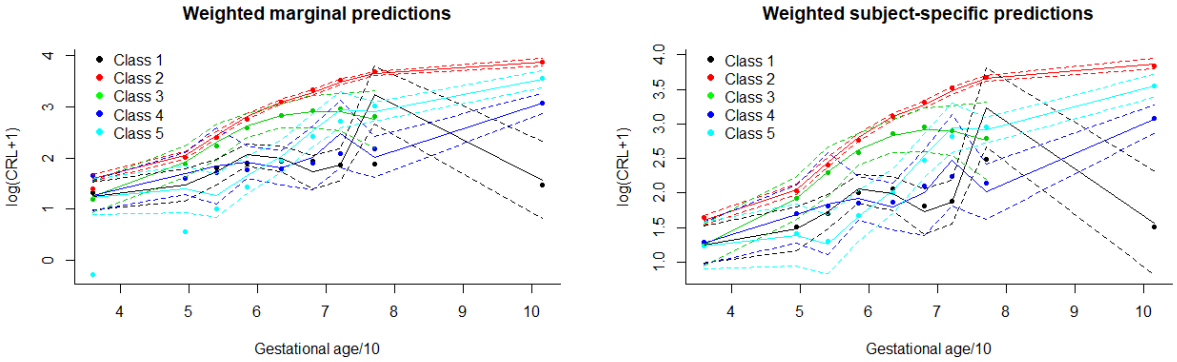

Figure 8: Comparison between observed and predicted marginal and subject-specific longitudinal profiles for log(CRL+1) calculated from model 7: *For class 2 and 3, the marginal predicted profile fits the observed profile very well. The marginal predicted longitudinal profiles in class 1 and 4 fit the observed profiles reasonably well. The marginal predicted profile within class 5 deviates from the observed profile at small gestational ages. The observed and predicted subject-specific profiles are corresponding well within all classes. class 2 shows the highest logCRL values at all times. Thus it is hypothesized that class 2 corresponds to the group of women with a viable pregnancy.*

Figure 9 shows the class-specific event-free probabilities for CRL-model 7 over time. Class 5 is clearly corresponding to women who do not miscarry while class 1 and 3 are clearly corresponding to women who do experience a miscarriage. Within class 2 and 4, the majority of the subjects will not experience a miscarriage. As compared to the survival curves plotted for CRL-model 3 in figure 6, the survival functions plotted for CRL-model 7 look much better.

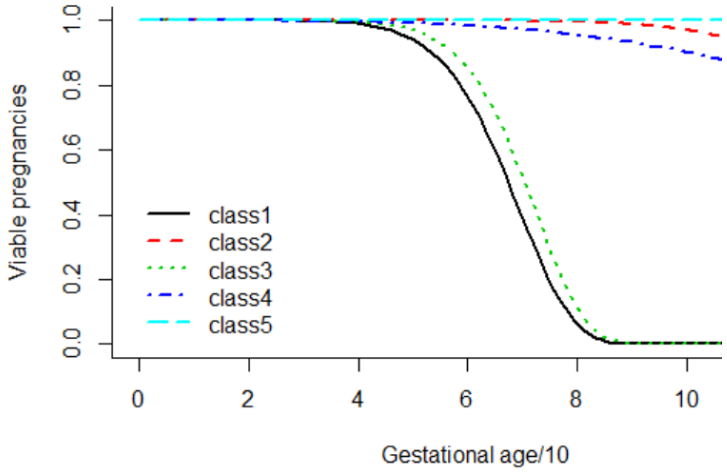

Figure 9: Class-specific event-free probability for CRL-model 7: *Class 5 is corresponding to women who do not miscarry. Class 1 and 3 are corresponding to women who do experience a miscarriage. Within class 2 and 4, the majority of the subjects will not experience a miscarriage.*

In order to see whether class 1 and class 3 indeed correspond to the group of miscarriages and whether classes 2, 4 and 5 correspond to the viable pregnancies, the number of observed miscarriages were evaluated in all latent classes. Also, the average values for

the variables included in CRL-model 7 were calculated for all latent classes to investigate how the latent classes differ from each other. Table 12 shows the proportion and number of subjects classified in each of the five classes, together with the number of subjects for which a miscarriage has been observed at the end of the first trimester and the average values for the variables included in CRL-model 7. CRL-model 7 contains 5 latent classes with probabilities of 4.82% for class 1, 87.8% for class 2, 3.05% for class 3, 1.28% for class 4 and 3.05% for class 5. The corresponding number of subjects classified in each of these classes is 30, 547, 19, 8 and 19 respectively. According to the longitudinal CRL-evolutions and the survival plots obtained from CRL-model 7, it was assumed that class 5 corresponds to the group of viable pregnancies while classes 1 and 3 correspond to women who experienced a miscarriage. For classes 2 and 4 it was assumed that the majority of the subjects would not experience a miscarriage. Indeed, from the 19 subjects *a posteriori* classified into class 5, only 1 subject had experienced a miscarriage at the end of the first trimester. From the 547 and 8 subjects classified in classes 2 and 4, miscarriages were observed for only 5 subjects and 1 subject respectively. Finally, for the 30 and 19 subjects classified in classes 1 and 3, 29 and 19 miscarriages were observed respectively. Classes 1 and 3 thus indeed correspond to the women who did experience a miscarriage, while the occurrence of a miscarriage in the other groups is only rare. The difference between the two classes that contain the group of women who do miscarry, i.e. class 1 and 3, appears to be the fact that a fetal heartbeat is more often detected in class 3 than in class 1 with 73.68% and 40.00% respectively. Besides, the average maternal age in class 3 is higher than the average maternal age in class 1. When looking at the classes that correspond to the women who do not miscarry, class 4 is standing out because a fetal heartbeat was only detected in 37.5% of the cases as compared to 76.05% and 84.21% for classes 2 and 5 at the first scan.

Table 12: Proportion and number of subjects assigned to the 5 latent classes for CRL-model 7.

	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Proportion	4.82%	87.8%	3.05%	1.28%	3.05%
Number of subjects	30	547	19	8	19
Number of final miscarriages	29	5	19	1	1
Maternal age	32.97	32.91	34.37	29.62	30.05
Paternal age	34.83	35.06	35.95	31.5	34.74
Fetal heartbeat at first scan	40.00%	76.05%	73.68%	37.50%	84.21%
PUQE-score at first scan	4.00	4.57	3.47	3.38	3.68

3.2.2 Models Including MSD

Because the longitudinal profiles of MSD seem to be linear, it was decided to only include a main time effect in the linear mixed model part. Since no quadratic term was included in the linear mixed model, gestational age was not divided by 10, as was done for the CRL-models in the previous section. Further, all MSD-models included a random intercept and a random slope. The included covariates and the number of latent classes differ from model to model. A detailed overview of all fitted MSD-models, indicating the number of latent classes and the covariates that were included, is shown in table 13.

Table 13: Overview of the fitted joint latent class models including MSD.

Covariates	Sub-model	Model 1	Model 2	Model 3
Maternal age	Mixed model	x	x	x
	Survival model	x	x	x
Fetal heartbeat	Multinomial model	x	x	x
	Mixed model	x	x	x
	Survival model	x	x	x
Paternal age	Survival model	x		x
Amnion	Multinomial model	x	x	x
PUQE score	Multinomial model	x		x
	Mixed model	x	x	x
	Latent classes	2	2	3
	BIC	11044.54	11989.52	11008.6

The BIC for the initial model including all covariates discussed in the exploratory analysis was 11047.19.

Two latent classes

The initial model containing the longitudinal variable MSD, two latent classes and all covariates identified in the exploratory analysis gave rise to a BIC value of 11047.19. The covariates with the highest p-values were then sequentially removed from the model, resulting in models 1 and 2 with respective BIC values of 11044.54 and 11989.52. Convergence was reached for both models and the assumption of conditional independence could not be rejected in any of the two models ($p_1 = 0.3237, p_2 = 0.4125$). According to these BIC values, model 1 was selected as the best MSD-model with two latent classes and resulted in posterior proportions of 89.52% and 10.48% for class 1 and 2 respectively. Table 16 shows the proportion and number of subjects classified in class 1 and class 2. 620 subjects, of which 58 (9.35%) experienced a miscarriage, were used for the estimation of MSD-model 1 and 22 parameters had to be estimated. The fetal heartbeat had a significant effect in the multinomial logistic regression model. Further, the fetal heartbeat and the maternal age had a significant impact on the proportional hazards model and the linear mixed model. The parameter estimates and corresponding p-values for this model are given in table 25 in the appendix.

The goodness-of-fit of MSD-model 1 was then assessed using the posterior classification of subjects. Table 14 shows the the proportions of subjects classified with a posterior probability above 0.7, 0.8 and 0.9. According to this table, many subjects are unambiguously assigned to any of the two classes, especially for latent class 1. Table 15 shows the posterior classification table for MSD-model 1, again indicating good discrimination.

Table 14: Proportion of subjects classified in each of the latent classes with a posterior probability above 0.7, 0.8 and 0.9 for MSD-model 1.

	Class 1	**Class 2**
Probability > 0.7	98.56%	93.85%
Probability > 0.8	98.02%	89.23%
Probability > 0.9	95.68%	80.00%

Table 15: Posterior classification table for MSD-model 1.

	Probability 1	**Probability 2**
Class 1	0.9843	0.0157
Class 2	0.0608	0.9392

The goodness-of-fit of MSD-model 1 was then further investigated using residual plots as is shown in figure 10. The residuals are clustering around zero, but the QQ-plots show some deviation from the intersecting line at the tails. Therefore, an extension of the model to a model with more than two latent classes seems necessary.

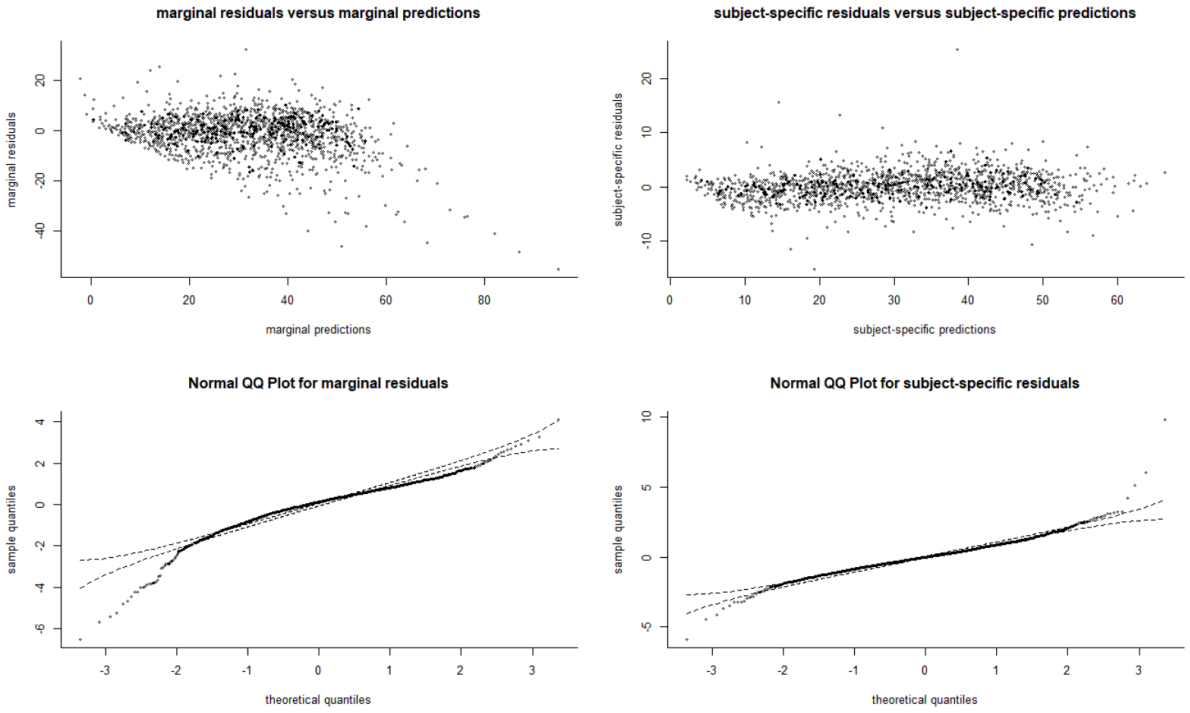

Figure 10: Residual plots for the assessment of the goodness-of-fit of MSD-model 1: *The residuals are clustering around zero, but the QQ-plots show deviation from the intersecting line at the tails.*

Figure 11 shows the predicted and observed marginal and subject-specific longitudinal profiles of MSD as given by model 1. The predicted profiles fit the observed profiles very well. The MSD values of class 1 are higher than the MSD values of class 2 at all time points. It is therefore hypothesized that class 2 corresponds to the group of women who experienced a miscarriage and that class 1 corresponds to the group of women with a viable fetus.

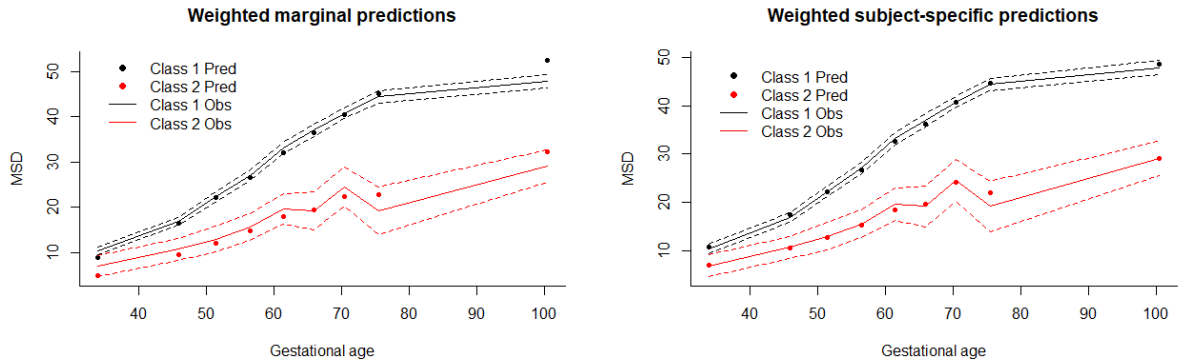

Figure 11: Comparison between observed and predicted marginal and subject-specific longitudinal profiles for MSD calculated from model 1: *The predicted profiles fit the observed profiles well. The MSD values of class 1 are higher than the MSD values of class 2 at all time points. Thus, it is hypothesized that class 2 corresponds to the group of women who experienced a miscarriage.*

Figure 12 shows the class-specific event-free probabilities for MSD-model 1. This figure indicates that the probability for a miscarriage is higher in class 2 than in class 1 where only a very small number of subjects is expected to have a miscarriage. This figure again indicates that class 2 corresponds to the women with a miscarriage and that class 1 corresponds to the women who did not experience a miscarriage. However, as was also seen for the two-class model including CRL (CRL-model 3), the survival function for class 2 does not go down all the way to zero. It thus appears that a large proportion of the subjects assigned to class 2 are still having a healthy fetus after completion of the first trimester. We therefore might have to extend this model to a model with more than two latent classes.

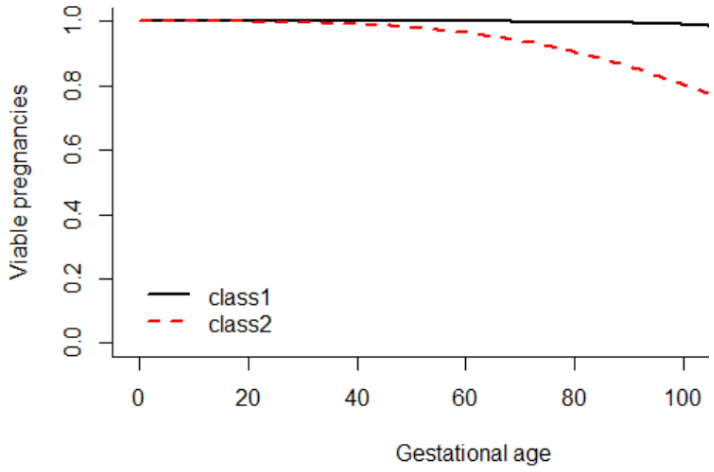

Figure 12: Class-specific event-free probability for MSD-model 1: *The probability for a miscarriage is higher in class 2 than in class 1. Class 2 corresponds to the women with a miscarriage and class 1 corresponds to the women who did not experience a miscarriage. The survival function for class 2 does not go down all the way to zero, indicating that a proportion of the subjects assigned to class 2 are still having a healthy fetus after completion of the first trimester.*

In order to see whether class 1 and class 2 indeed correspond to the group of viable pregnancies and the group of miscarriages respectively, the number of observed miscarriages were evaluated in both latent classes. Also, the average values for the variables included in MSD-model 1 were calculated for both latent classes to investigate how they differ from each other. Table 16 shows the proportion and number of subjects classified in each of the two latent classes, together with the number of subjects for which a miscarriage has been observed at the end of the first trimester and the average values for the variables included in MSD-model 1. MSD-model 1 contains 2 latent classes with probabilities of 89.52% for class 1 and 10.48% for class 2, corresponding to 555 and 65 subjects respectively. According to the longitudinal MSD-evolutions and the survival plots obtained from MSD-model 1, it was assumed that class 1 corresponds to the group of viable pregnancies while class 2 corresponds to the group of miscarriages. Indeed, from the 65 subjects classified into class 2, 54 subjects experienced a miscarriage by the end of the first trimester and only 9 subjects classified in class 1 experienced a miscarriage. Thus, the proportion of subjects that experienced a miscarriage is indeed much higher in class 2 than in class 1. Although the model does quite well, it can not completely discriminate between the women who do miscarry and the women who do not miscarry since 9 miscarriage cases were still classified in class 1. It can be seen that at the first scan, a fetal heartbeat was detected in 75.14% of the cases for class 1 and only in 49.23% of the cases for class 2. Also, the amnion was detected at the first scan for 18.92% of the subjects in class 1 and only for 4.62% of the subjects in class 2. At last, the PUQE-score that was reported at the first scan is higher in class 1 than in class 2 with values of 4.50 and 3.65 respectively.

Table 16: Proportion and number of subjects assigned to the 2 latent classes for MSD-model 1.

	Class 1	**Class 2**
Proportion	89.52%	10.48%
Number of subjects	555	65
Number of final miscarriages	9	54
Maternal age	32.81	33.43
Paternal age	35.03	35.28
Fetal heartbeat at first scan	75.14%	49.23%
Amnion at first scan	18.92%	4.62%
PUQE-score at first scan	4.50	3.65

More than 2 latent classes

To further improve MSD-model 1, three latent classes were fitted instead of two, resulting in model 3 for which a BIC value of 11008.6 was obtained. Covergence was reached for model 3 and the assumption of conditional independence could not be rejected ($p_3 = 0.1862$). When fitting models with more than 3 latent classes, there was no longer convergence. MSD-model 3 classified 9.19% of the subjects into class 1, 88.87% in class 2 and 1.94% in class 3. These proportions and the corresponding number of subjects assigned to each of the three latent classes are given in table 19. Since the covariates included in MSD-model 3 are the same as those included in MSD-model 1, the same number of subjects was used for the estimation of the model, i.e. 620 subject of which 58 (9.35%) experienced a miscarriage. For this model, 30 parameters had to be estimated.

The maternal and paternal age had a significant effect in the proportional hazards model and the fetal heartbeat and the maternal age had a significant effect in the linear mixed model part. The parameter estimates and corresponding p-values for the covariates included in MSD-model 3 are given in table 26 in the appendix.

As before, the goodness-of-fit of the model was assessed using the posterior classification of subjects to the latent classes. Table 17 gives the proportions of subjects with a posterior probability above 0.7, 0.8 and 0.9 and table 18 shows the posterior classification table. Subjects were unambiguously assigned to classes 1 and 2 where 96.49% of the subjects in class 1 and 98.37% of the subjects in class 2 had a posterior probability above 0.9. Although only 66.67% of the subjects in class 3 had a posterior probability above 0.9, 83.33% had a posterior probability above 0.7, which can still be considered relatively good.

Table 17: Proportion of subjects classified in each of the latent classes with a posterior probability above 0.7, 0.8 and 0.9 for the MSD-model 3.

	Class 1	**Class 2**	**Class 3**
Probability > 0.7	98.25%	99.64%	83.33%
Probability > 0.8	98.25%	99.27%	66.67%
Probability > 0.9	96.49%	98.37%	66.67%

Table 18: Posterior classification table for MSD-model 3.

	Probability 1	**Probability 2**	**Probability 3**
Class 1	0.9833	0.0167	0.0000
Class 2	0.0055	0.9933	0.0012
Class 3	0.0374	0.0835	0.8791

The goodness-of-fit of MSD-model 3 can be further investigated by looking at the residual plots shown in figure 13. However no great differences can be seen with the residual plots from the previous model (see fig. 10). Although the residuals are clustering around zero, the QQ-plots again show some deviation from the intersecting line at the tails.

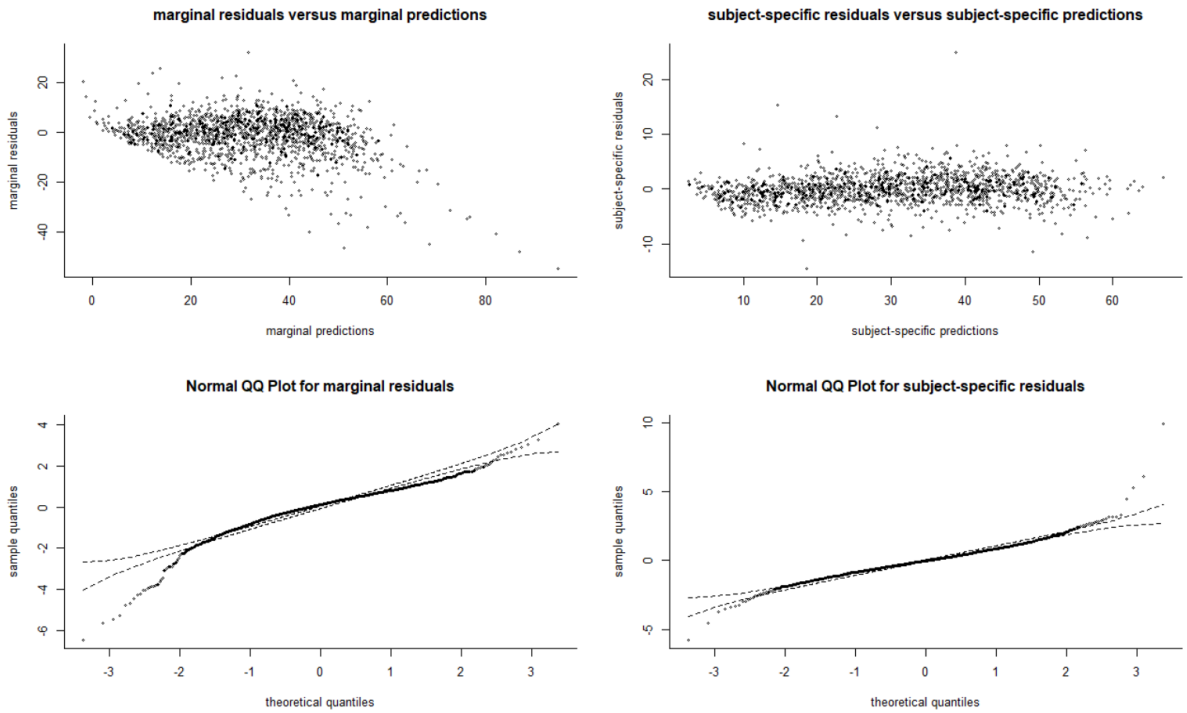

Figure 13: Residual plots for the assessment of the goodness-of-fit of MSD-model 3: *The residuals are clustering around zero, but the QQ-plots show some deviation from the intersecting line at the tails.*

Figure 14 shows the class-specific predicted and observed marginal and subject-specific longitudinal profiles for the evolution of MSD over time. Within class 1 and 2, the predicted MSD profiles fit the observed MSD profiles very well. However, for class 3, the predicted profile shows a bad fit to the observed profile. Class 2 has the highest MSD profiles at all times and we therefore hypothesize that this class corresponds to group of viable pregnancies. Since the majority of the subjects were assigned to class 2, the precision is much higher and the confidence bands are much more narrow than for class 1 and class 3.

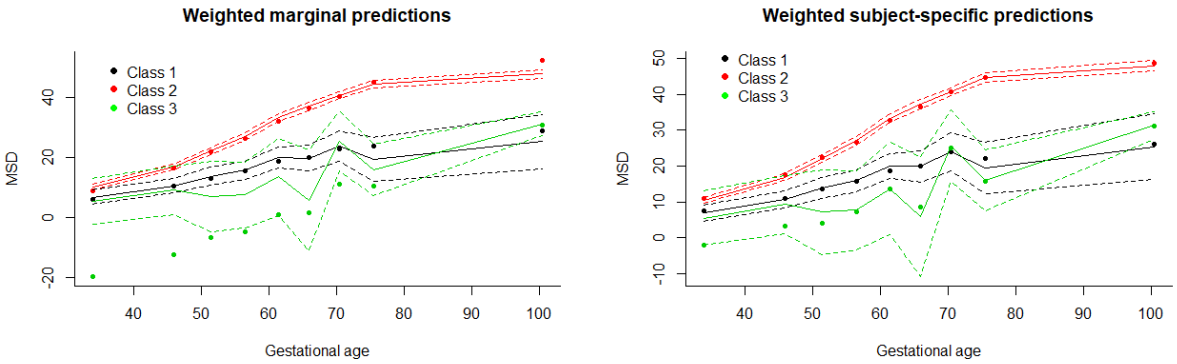

Figure 14: Comparison between observed and predicted marginal and subject-specific longitudinal profiles for MSD calculated from model 3: *Within class 1 and 2, the predicted MSD profiles fit the observed MSD profiles very well. For class 3, the predicted profile shows a bad fit to the observed profile. Class 2 has the highest MSD profiles at all times. It is thus hypothesized that class 2 corresponds to group of viable pregnancies.*

Figure 15 shows the class-specific survival functions for MSD-model 3. According to this figure, class 3 corresponds to the group of women who do not experience a miscarriage while class 1 corresponds to the group of women who do experience a miscarriage. Class 2 seems to represent a mixture of women who do and do not experience a miscarriage, although the majority of the subjects classified in class 2 will not have had a miscarriage after completion of the first trimester.

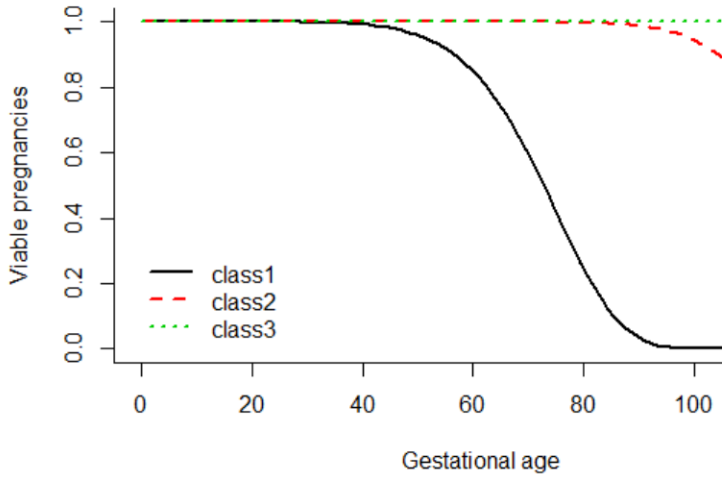

Figure 15: Class-specific event-free probability for MSD-model 3: *Class 3 corresponds to the group of women who do not experience a miscarriage. Class 1 corresponds to the group of women who do miscarry. Class 2 represents a mixture of women who do and do not experience a miscarriage, although the majority of the subjects in class 2 will not miscarry.*

In order to see whether class 1 indeed corresponds to the group of miscarriages while class 2 and class 3 correspond to the group of viable pregnancies, the number of observed miscarriages were evaluated in all three latent classes. Also, the average values for the variables included in MSD-model 3 were calculated for all latent classes to investigate how

they differ from each other. Table 19 shows the proportion and number of subjects classified in each of the three latent classes, together with the number of subjects for which a miscarriage has been observed at the end of the first trimester and the average values for the variables included in MSD-model 3. MSD-model 3 contains 3 latent classes with probabilities of 9.19% for class 1, 88.87% for class 2 and 1.94% for class 3, corresponding to 57, 551 and 12 subjects respectively. According to the longitudinal MSD-evolutions and the survival plots obtained from MSD-model 3, it was assumed that class 1 corresponds to the group of miscarriages, class 2 corresponds to a group with mainly viable pregnancies and class 3 corresponds to a group of women with viable pregnancies. Indeed, from the 57 subjects classified into class 1, 56 subjects had experienced a miscarriage by the end of the first trimester. Only 6 subjects classified in class 2 and only 1 subject classified in class 3 had a miscarriage after the first trimester. Class 1 thus corresponds to the group of women who miscarry while classes 2 and 3 mainly correspond to women who do not miscarry. Further, it can be seen that at the first scan, a fetal heartbeat was detected in 45.61% of the cases in class 1, in 75.68% of the cases in class 2 and in 50.00% of the cases in class 3. The amnion was detected at the first scan for 5.26% of the subjects in class 1, for 19.06% of the subjects in class 2 and for 0.0% of the subjects in class 3. Also, the maternal age is higher in class 1 than in class 2 and 3. At last, the average PUQE-score reported at the first scan was higher for class 2 than for class 1 and 3.

Table 19: Proportion and number of subjects assigned to the 3 latent classes for MSD-model 3.

	Class 1	**Class 2**	**Class 3**
Proportion	9.19%	88.87%	1.94%
Number of subjects	57	551	12
Number of final miscarriages	56	6	1
Maternal age	33.86	32.81	31.08
Paternal age	35.6	35.05	32.75
Fetal heartbeat at first scan	45.61%	75.68%	50.00%
Amnion at first scan	5.26%	19.06%	0.00%
PUQE-score at first scan	3.737	4.506	3.25

3.2.3 Predictive Accuracy of the Models

The predictive accuracy of the models was assessed using the EPOCE criterion, which is approximately estimated with the CVPOL$_a$ estimate. Figure 16 shows the CVPOL$_a$ values for the CRL- and MSD-models. It can be seen that the models with only two latent classes have higher CVPOL$_a$ values as compared to the models with more latent classes. Besides, the models including the longitudinally measured variable CRL have considerably lower CVPOL$_a$ values than the models including the longitudinal variable MSD. CRL-model 7 thus seems to have the highest accuracy to predict the risk for a miscarriage.

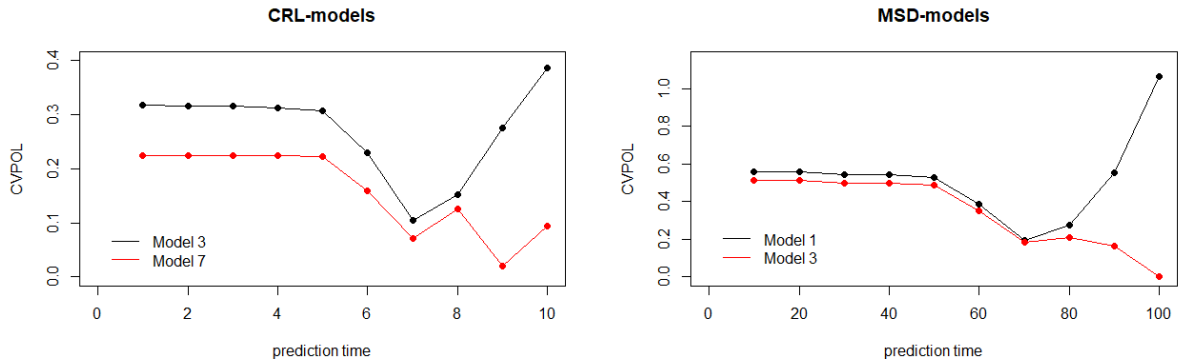

Figure 16: Comparison of the predictive accuracy for CRL-model 3 and 7 and MSD-model 1 and 3: *Models with only two latent classes have higher* $CVPOL_a$ *values as compared to the models with more latent classes. Models including the longitudinally measured variable CRL have considerably lower* $CVPOL_a$ *values than the models including the variable MSD.*

3.3 Individual Dynamic Predictions

To illustrate the use of a joint latent class model in practice, we applied dynamic predictions to two random subjects from the EPOS-data set according to CRL-model 7. The first subject (E1200) was 31 years old and delivered a healthy child at the end of the pregnancy. Also the father was 31 years old. At the first ultrasound scan which took place at 52 days of gestational age, the subject reported a PUQE-score of 3 and a fetal heartbeat was detected. The second subject (E1006) was 28 years old and had a first trimester miscarriage that was diagnosed at a gestational age of 73 days. At the first ultrasound scan which took place at 45 days of gestational age, no fetal heartbeat was detected and the subject reported a PUQE-score of 9. The father was 25 years old. The model with the best predictive accuracy, i.e. CRL-model 7, was then used to plot their predicted longitudinal logCRL profiles and to dynamically predict the risk for a miscarriage. Both subjects were *a priori* classified in class 2, where subject E1200 had a probability of 91.38% to belong to class 2 and subject E1006 had a probability of 86.63% to belong to class 2. Table 20 gives the probabilities to belong to each of the five latent classes for both subjects.

Table 20: Posterior probabilities to belong to each of the latent classes for subject E1200 and subject E1006 as computed from the final model.

	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Viable (E1200)	7.4052%	91.3755%	0.6581%	0.2437%	0.3175%
Miscarriage (E1006)	3.0947%	86.6263%	4.3703%	0.9749%	4.9338%

Figure 17 shows the predicted class-specific longitudinal trajectories of the logCRL values for both subjects. When looking at the trajectories predicted within class 2, it can be seen that the subject who experienced a miscarriage is starting with a lower logCRL value but is expected to have a steeper increase as compared to the subject that had a viable pregnancy.

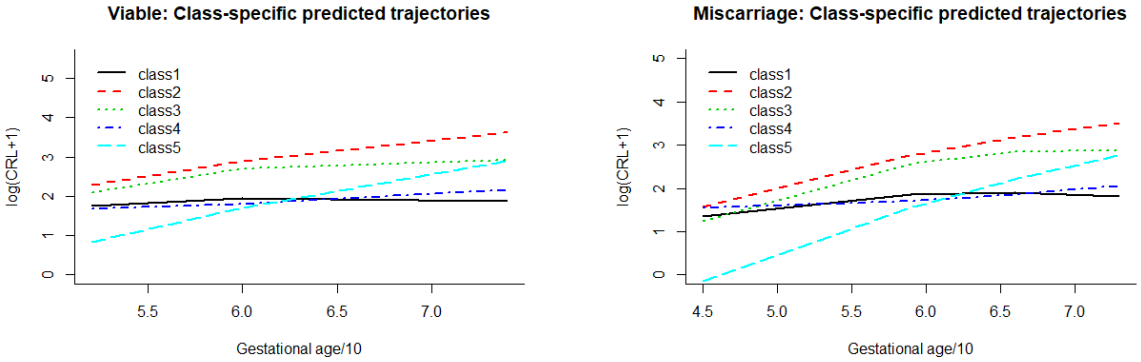

Figure 17: Predicted class-specific longitudinal profiles of log(CRL+1) for a random subject with a viable pregnancy and a random subject with a miscarriage: *For the trajectories predicted within class 2, the subject who experienced a miscarriage is starting with a lower logCRL value but is expected to have a steeper increase as compared to the subject that had a viable pregnancy.*

Figure 18 shows the dynamic predictions for both subjects at landmark times 7, 8 and 9 and horizon times 0.2, 0.5, 1, 1.5, 2, 2.5 and 3. Clearly, the predicted risk for a miscarriage is much lower at all landmark times for the subject who did not miscarry. Further, the risk for a miscarriage drops considerably when the landmark time increases. However, subject E1006 miscarried at a gestational age of 73 days. The dynamic predictions at landmark time 8 and 9 are therefore not relevant for this subject but show that if the subject would not have miscarried at the gestational age of 73 days, the risk for a miscarriage would have dropped gradually. This effect, although less prominent can also be seen for the subject that delivered a healthy baby at the end of the pregnancy. Note that all predictions have broad confidence bands, indicating high uncertainty for the individual dynamic prediction of a miscarriage.

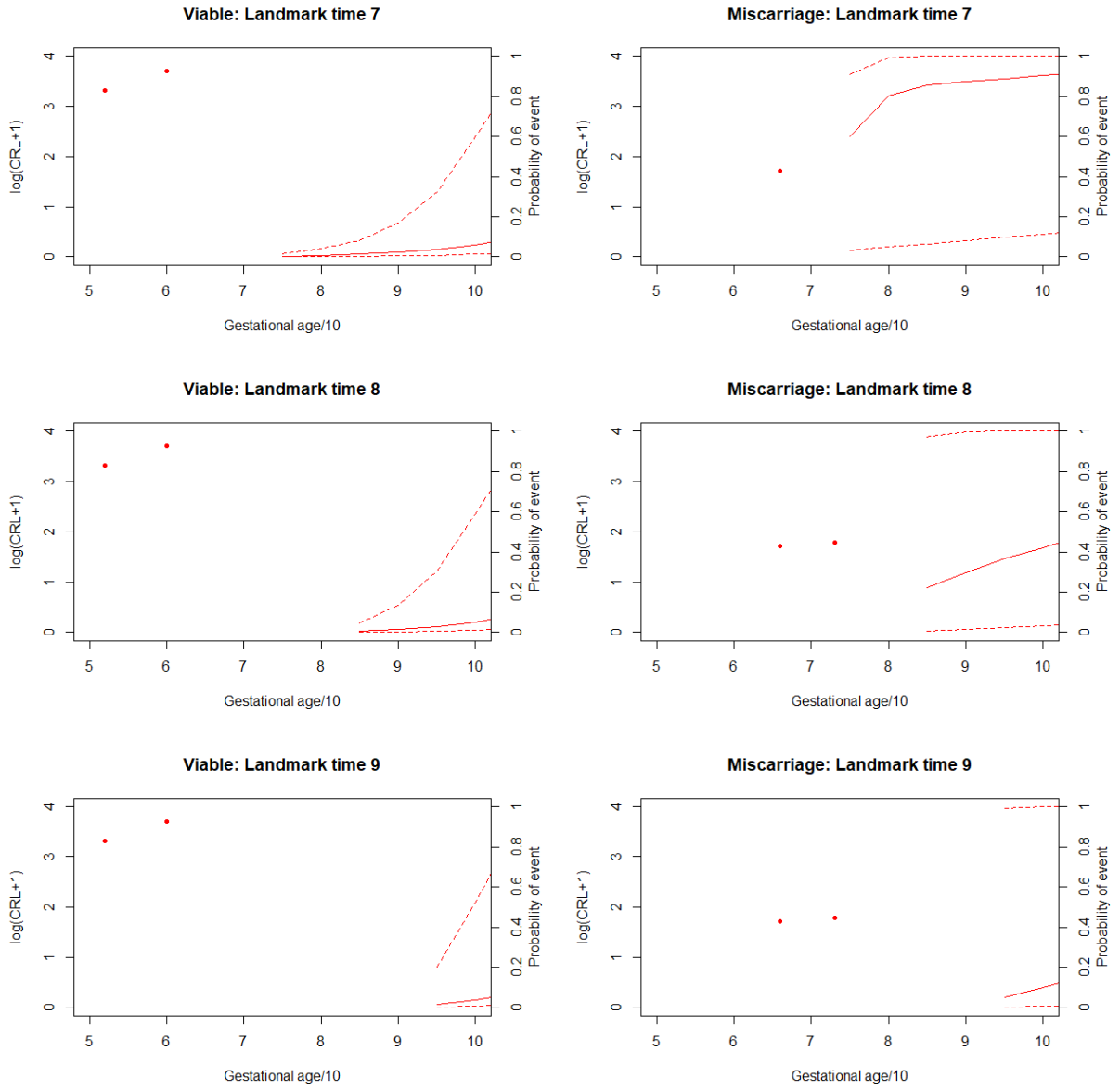

Figure 18: Dynamic predictions for a random subject with a viable pregnancy and a random subject with a miscarriage: *The predicted risk for a miscarriage is lower at all landmark time points for the subjects who did not miscarry and drops considerably when the landmark time increases. The dynamic predictions at landmark time 8 and 9 for the subject who miscarried at a gestational age of 73 days are not relevant for the subject itself but show that if the subject would not have miscarried, the risk for a miscarriage would have dropped gradually. All predictions have broad confidence bands.*

Chapter 4

Discussion

The aim of this study was to build a predictive model to identify the risk for a miscarriage using the joint latent class modelling approach. To our knowledge, this is the first study that applies the joint latent class methodology to pregnancy data. Several models were fitted, including different covariates and different numbers of latent classes. In this assay, we first selected the most appropriate CRL- and MSD-model with 2 latent classes and extended these models to 5 and 3 latent classes respectively.

CRL-model 3 contains 2 latent classes where the proportion of subjects that experienced a miscarriage is higher in class 2 than in class 1. However, CRL-model 3 can not completely discriminate between women who do miscarry and women who do not miscarry. When comparing the average values for the variables included in CRL-model 3 between the two classes, a fetal heartbeat was more often detected class 1 than in class 2. Also, the PUQE-score that was reported at the first scan is higher in class 1 than in class 2. It thus appears that the detection of a fetal heartbeat and a high PUQE-score at the first scan are related to a low risk for miscarriage.

CRL-model 7 contains 5 latent classes where classes 1 and 3 correspond to the group of women who experienced a miscarriage while the majority of women classified in class 2, class 4 and class 5 had a viable pregnancy. The difference between the two classes that contain the group of women who do miscarry, i.e. class 1 and 3, appears to be the fact that a fetal heartbeat is more often detected in class 3 than in class 1. Besides, the average maternal age in class 3 is higher than the average maternal age in class 1. Class 1 might thus correspond to the women who do miscarry due to the absence of a fetal heartbeat at the first scan, while class 3 might correspond to the group of women who do miscarry due to a high maternal age. When looking at the classes that correspond to the women who do not miscarry, class 4 is standing out because the probability to detect a fetal heartbeat in class 4 was much lower as compared to the probability to detect a fetal heartbeat in classes 2 and 5. Class 4 might thus correspond to the group of women who do not miscarry even if a fetal heartbeat was not detected at the first scan.

MSD-model 1 contains 2 latent classes where class 1 corresponds to the group of women with a viable pregnancy while class 2 corresponds to the group of women who experienced a miscarriage. However, the model can not completely discriminate between the women who do miscarry and the women who do not miscarry since 9 miscarriage cases were still

classified in class 1. The proportion of subjects for which a fetal heartbeat was detected at the first scan is much higher in class 1 than in class 2. Also, the proportion of subjects where the amnion was detected at the first scan is much higher in class 1 than in class 2. At last, the PUQE-score that was reported at the first scan is higher in class 1 than in class 2. The detection of a fetal heartbeat, the detection of the amnion and a high PUQE-score at the first scan thus appear to lower the risk for a miscarriage.

MSD-model 3 contains 3 latent classes where class 1 corresponds to the women who do miscarry while classes 2 and 3 correspond to the women who do not miscarry. The proportion of subjects where a fetal heartbeat was detected at the first scan is lower in classes 1 and 3. Also, the proportion of subjects where the amnion was detected at the first scan is lower in classes 1 and 3. The maternal age is higher in class 1 than in class 2 and 3. At last, the average PUQE-score reported at the first scan was higher for class 2 than for class 1 and 3. The difference between class 2 and class 3 might thus be the fact that although the probability to detect a fetal heartbeat and the probability to detect the amnion in class 3 is rather low, these subjects are still not expected to miscarry.

When comparing all four models, the CRL-models showed a better predictive ability as compared to the MSD-models, where the CRL-model with 5 latent classes (model 7) showed the highest predictive ability. Besides, the BIC values obtained by the CRL-models were considerably lower than the BIC values obtained by the MSD models. Again, CRL-model 7 with 5 latent classes showed the lowest BIC value of all models. At last, when comparing the longitudinal profiles and the event-free probabilities for all four models, model 7 is giving the most plausible and informative results. We therefore selected model 7 as the best model for the prediction of the risk for a miscarriage. This model resulted in 5 latent classes, where the majority of the subjects was classified into class 2, a class where only very few subjects are expected to miscarry. Also within class 4 and 5, only very few subjects are expected to miscarry. Further, subjects classified in class 1 and 3 are expected to miscarry during the first trimester. The number of subjects classified in one of these two classes *a posteriori* was equal to 49. From these 49 subjects, 48 indeed had a miscarriage at the end of the first trimester. Further, the goodness-of-fit of this model as assessed by the posterior classification of subjects and the residual plots was good.

In general, when looking at the results obtained for all four models, a high maternal and a high paternal age, the absence of a fetal heartbeat at the first scan, the absence of the amnion at the first scan and a low PUQE-score seem to be associated with a higher risk for a miscarriage. When looking at the final model, i.e. CRL-model 7, the maternal and paternal age showed a significant effect in the proportional hazards model and the fetal heartbeat and maternal age showed a significant effect on the linear mixed model (see table 24 in the appendix). The PUQE-score that was also included in the final model did not show a significant effect. Indeed, it has been shown in previous studies that the detection of a fetal heartbeat and the age of the mother have a large impact on the risk for a miscarriage. Previous studies also identified a high paternal age as a possible risk factor for miscarriage [7].

One of the greatest advantages of the joint latent class modelling approach used in this study is the fact that the correlation between the longitudinal process and the survival process is taken into account by the use of a relatively simple model that can be estimated using general maximum likelihood theory. Besides, due to the inclusion of latent classes, the heterogeneity within the population of pregnant women can be taken into account. As compared to other joint modelling methods, the joint latent class modelling approach has several advantages. Another popular approach to simultaneously model a longitudinal process as well as a survival process is the shared random effects model or selection model in which a function of random effects defining the longitudinal process is included as a covariate in the survival model [29]. The general shared random effects model has been further extended by Yu et al. to include a cured fraction that is modelled as a logistic function of baseline covariates [30]. However, the shared random effects model accounting for a cured fraction is not implemented in any software program. Because the joint latent class modelling approach uses two latent components, i.e. the random effects and latent classes, it allows a more flexible association structures as compared to the shared random effects model that assumes the random effects to capture the association between the repeatedly measured variable as well as the association between the longitudinal process and the survival process. Because of the independence between the longitudinal process and the survival process given on the latent classes, the likelihood remains tractable to compute as opposed to the shared random effects model that often requires numerical integration [31]. Besides, the flexibility of the joint latent class model allows for better predictions as compared to the shared random effects model [19].

Another big advantage of this approach is the fact that the joint latent class model allows for individual dynamic risk predictions. This might have great implications for patient follow-up and counseling during the pregnacy. Indeed, when the risk for a miscarriage appears to be high, medical advice can be given to the expecting mother in an attempt to avoid a miscarriage. The risk for a miscarriage can then be reevaluated over time for each subject individually and the medical advice can be adapted accordingly. In this study, we gave the dynamic predictions for two random subjects from the EPOS-data set as an example of how the final model can be used in practice.

The joint latent class modelling approach has some disadvantages as well. First of all, when one starts with a specific assumption on the dependency between the longitudinal process and the survival process, this assumption can be build in into the shared random effects model but not into the joint latent class model that is assuming independence between the two processes. Thus, the joint latent class model can not be used to evaluate specific assumptions regarding the characteristics of the longitudinal process that are the most influential on the risk for the event [19]. Also, the interpretation of the latent classes and the parameters within each class might come with a difficult interpretation. Another issue is that the log-likelihood of the joint latent class model might have multiple maxima. As a consequence, the model must be refitted multiple times with different starting values in order to avoid convergence to a local maximum. Besides, because the number of latent classes is not known *a priori*, multiple models must be fitted to find the best number of latent classes. Due to the fact that the models must be refitted multiple times, computation time can increase considerably [31].

Eventually, we also have to note that our study comes with some limitations. First of all, the number of subjects included in our study is not that large, especially when compared to the number of parameters that had to be estimated. Although 753 subject were eligible for the study, only 623 and 620 subjects could be used for the estimation of the CRL- and MSD models respectively. Besides, the number of subjects included in the latent classes is often considerably lower than 623 or 620, thus further compromising the precision of our estimates in each of the latent classes. Indeed, very broad confidence bands were seen in the individual dynamic predictions in section 3.3. We therefore recommend to estimate the model parameters on a larger data-set to obtain more precise estimates and better individual predictions.

A second limitation to our study is the fact that the predictive accuracy of the final model is expressed using the approximate estimate for the EPOCE criterion, i.e. CVPOL_a. Since most studies on miscarriage risk factors report the predictive ability of their models in terms of sensitivity and specificity or in terms of AUC, it is difficult to compare the predictive ability obtained in this study to the predictive ability obtained by other studies. At last, due to the fact that 48 parameters had to be estimated in CRL-model 7, the final model might be prone to over-fitting. To some extend, this feature was taken into account by the use of a cross-validation based method to assess the predictive ability of the model. However, we still recommend to validate this model on external pregnancy data.

At last, we want to note that the 'lcmm' package is only able to include one outcome variable into the linear mixed model part of the joint latent class model. However an extension of the package to include a multivariate linear mixed model into the joint latent class modelling framework might be of interest. Indeed, in our study, simultaneously modelling the evolutions of CRL and MSD might give better predictions as compared to models based on CRL or MSD alone.

Chapter 5

General Conclusion

In this study, we have build several predictive models that can be used to dynamically assess the risk for a miscarriage in pregnant women. To our knowledge, this is the first study that applies the joint latent class methodology to pregnancy data. We can summarize the results of our study with the following conclusions:

- In general, when looking at the results obtained for all fitted models, a high maternal and paternal age, the absence of a fetal heartbeat at the first scan, the absence of the amnion at the first scan and a low PUQE-score are associated with a higher risk for a miscarriage.
- The best model (CRL-model 7) identified in this study includes 5 latent classes, shows good discrimination, acceptable profiles and has the best predictive accuracy of all fitted models.
- The subjects classified in classes 2, 4 and 5 are not expected to miscarry even though the probability to detect a fetal heartbeat is class 4 is rather low.
- Classes 1 and 3 from the final model correspond to the group of women who miscarry and differ from each other in terms of the average maternal age and the proportion of subject where a fetal heartbeat was detected at the first scan.
- The predictive model identified in our study might be very useful in practice due to the fact that individual dynamic predictions can be made. However, due to the limited size of our study, broad confidence bands for the dynamic predictions were often obtained and re-estimating the model on a larger dataset is therefore recommended.
- Finally, it is recommended to validate the final model on external data.

Although our model suffers from several caveats as was discussed in section 4, the joint latent class modelling approach seems promising for the construction of predictive risk models and we believe this approach might be of use in practice due to the fact that it can be used dynamically over time. However, since this methodology is rather recent and remains unfamiliar to many researchers, its application should be further investigated. Future research could further establish the joint latent class modelling framework, for example by developing models that can include multiple longitudinal outcome variables in the linear mixed model part.

Chapter 6

Appendix

Table 21: Detailed list of the baseline covariates.

Variable	Type	Explanation
EPOS	Identifier	Patient identifier
Pregnancy outcome	Ordinal	Viable/PUV/ Miscarriage
Maternal age	Continuous	Age of the mother in years
Paternal age	Continuous	Age of the father in years
Maternal ethnicity	Nominal	Asian/ Caucasian/Afro-caribbean/Mixed/Other
Paternal ethnicity	Nominal	Asian/ Caucasian/Afro-caribbean/Mixed/Other
BMI	Continuous	BMI of the mother in kg/m^2
Height	Continuous	Height of the mother in m
Weight	Continuous	Weight of the mother in kg
Gravida	Continuous	Number of pregnancies in the past
Para	Continuous	Number of viable pregnancies in the past
TOP	Continuous	Number of terminated pregnancies in the past
TOP binary	Binary	Indicator variable for TOP in the past
Ectopic	Continuous	Number of ectopic pregnancies in the past
Ectopic binary	Binary	Indicator variable for ectopic pregnancies in the past
Cesarean Section	Continuous	Number of cesarean sections in the past
PSH Uterus	Binary	Indicator variable for surgery of the uterus in the past
PSH Cervix	Binary	Indicator variable for surgery of the cervix in the past
Progesterone	Binary	Indicator variable for the intake of progesterone
Metformin	Binary	Indicator variable for the intake of metformin
Aspirin	Binary	Indicator variable for the intake of aspirin
FolicAcid	Binary	Indicator variable for the intake of folic acid
PreconceptualFolicAcid	Binary	Indicator variable for the preconceptual intake of folic acid
Pre-pregnancy alcohol	Continuous	Alcohol use before the pregnancy in units per week
Current alcohol	Continuous	Current alcohol use in units per week
SmokingStatus	Binary	Indicator variable for smoking status
1stTMiscarriage	Continuous	Number of past miscarriages in the first trimester
2ndTMiscarriage	Continuous	Number of past miscarriages in the second trimester
CertaintyUPT	Ratio	Certainty of the urine pregnancy test
CertaintyLMP	Ratio	Certainty of the LMP
EDD LMP	Date	Expected delivery date by LMP
EDD by scan	Date	Expected delivery date by scan
Date of delivery	Date	Date of delivery
FinalGALMP	Continuous	Final gestational age as calculated by the LMP
FinalGAdays	Continuous	Final gestational age in days
FinalGADiff	Continuous	Difference between FinalGALMP and FinalGAdays in days
Baby Weight	Continuous	Weight of the baby in g
Baby gender	Binary	Male/Female

Table 22: Detailed list of the longitudinal variables measured at each scan.

Variable	Type	Explanation
GAbyLMP	Continuous	Gestational age as calculated by the LMP
GAbyCRL	Continuous	Gestational age as calculated by the CRL
GAbyMSD	Continuous	Gestational age as calculated by the MSD
PC	Nominal	Principal complaint: Pain/Bleeding/Pain & bleeding/reassurance
Site	Nominal	Anterior/Posterior/Fundal
BleedingDays	Continuous	Number of days with a bleeding
BleedingScoreAtPresent	Ratio	Bleeding score at the day of the ultrasound scan
WorstBleedingScore	Ratio	Worst bleeding score during each period between scans
PainDays	Continuous	Number of days with pain
PainScoreAtPresent	Ratio	Pain score at the day of the ultrasound scan
WorstPainScore	Ratio	Worst pain score during each period between scans
PUQEscore	Ratio	PUQE score indication the amount of nausea and vomiting
MSD	Continuous	Mean Sac Diameter
CRL	Continuous	Crown Rump Length
FH	Binary	Indicator variable for the presence of a fetal hearth beat
MYS	Continuous	Mean Yolk Sac Size
Amnion	Binary	Indicator variable for the detection of the amnion
AmnionSize	Continuous	Size of the amnion
SCHnumber	Continuous	Number of subchriotonic hematomas
SizeSCH	Continuous	Size of the largest SCH
SCHperc	Ratio	Percentage of the largest SCH
SCHloc	Nominal	Location of the largest SCH
SCHcont	Nominal	Content of the largest SCH: Homogene/Heterogene
outcomeAtScan	Ordinal	Pregnancy outcome at each scan: Viabe/PUV/Miscarriage
NumberOfScan	Continuous	Number of performed ultrasound scans during the first trimester

Table 23: Parameter estimates and p-values for CRL-model 3.

Variable	**Estimate**	**P-value**
Multinomial logistic regression model		
Intercept	1.06592	0.02301
Fetal heartbeat	1.09433	0.00098
PUQE-score	0.19467	0.10552
Proportional hazards model		
Weibull 1,1	0.17869	< 0.0001
Weibull 2,1	2.18863	< 0.0001
Weibull 1,2	0.23001	< 0.0001
Weibull 2,2	1.96382	< 0.0001
Maternal age	0.15191	0.00032
Paternal age	-0.05347	0.11291
Linear mixed model		
Intercept 1	-5.31828	< 0.0001
Intercept 2	1.16613	0.06315
Gestational age 1	1.93996	< 0.0001
Gestational age 2	-0.07994	0.64736
(Gestational age 1)^2	-0.10494	< 0.0001
(Gestational age 2)^2	0.02534	0.04102
Fetal heartbeat	0.06645	0.00042
Maternal age	0.00827	< 0.0001

Class 2 is the reference class for the multinomial logistic regression model. The indices 1,1-2,2 in the proportional hazards model denote the two parameters of the Weibull baseline function for the two latent classes. The numbers 1 and 2 in the linear mixed model denote the two latent classes.

Table 24: Parameter estimates and p-values for CRL-model 7.

Variable	Estimate	P-value
Multinomial logistic regression model		
Intercept 1	0.80598	0.46555
Intercept 2	2.43759	0.01438
Intercept 3	0.28218	0.84374
Intercept 4	0.90827	0.63974
Fetal heartbeat 1	-2.05355	0.00941
Fetal heartbeat 2	-0.64697	0.35248
Fetal heartbeat 3	-0.55232	0.54558
Fetal heartbeat 4	-2.13885	0.03937
PUQE-score 1	0.26038	0.31793
PUQE-score 2	0.35829	0.11898
PUQE-score 3	0.04962	0.89377
PUQE-score 4	-0.13031	0.81091
Proportional hazards model		
Weibull 1,1	0.37651	< 0.0001
Weibull 2,1	2.85026	< 0.0001
Weibull 1,2	0.26394	< 0.0001
Weibull 2,2	3.12891	< 0.0001
Weibull 1,3	0.36907	< 0.0001
Weibull 2,3	3.01454	< 0.0001
Weibull 1,4	0.23247	0.00070
Weibull 2,4	1.91919	0.02704
Weibull 1,5	0.05113	0.99794
Weibull 2,5	2.28086	0.99380
Maternal age	0.05101	0.00441
Paternal age	0.03533	0.00033
Linear mixed model		
Intercept 1	-3.98431	< 0.0001
Intercept 2	-5.26524	< 0.0001
Intercept 3	-9.97518	< 0.0001
Intercept 4	1.99128	0.15273
Intercept 5	-8.82872	< 0.0001
Gestational age 1	1.74867	< 0.0001
Gestational age 2	1.95855	< 0.0001
Gestational age 3	3.60791	< 0.0001
Gestational age 4	-0.34154	0.29740
Gestational age 5	2.41605	< 0.0001
(Gestational age 1)^2	-0.13408	< 0.0001
(Gestational age 2)^2	-0.10696	< 0.0001
(Gestational age 3)^2	-0.25675	< 0.0001
(Gestational age 4)^2	0.04459	0.02329
(Gestational age 5)^2	-0.11702	< 0.0001
Fetal heartbeat	0.05195	0.00172
Maternal age	0.00672	< 0.0001

Class 5 is the reference class for the multinomial logistic regression model. The indices 1,1-2,5 in the proportional hazards model denote the two parameters of the Weibull baseline function for the five latent classes. The numbers 1-5 in the linear mixed model denote the five latent classes.

Table 25: Parameter estimates and p-values for MSD-model 1.

Variable	Estimate	P-value
Multinomial logistic regression model		
Intercept	0.82802	0.06714
Fetal heartbeat	0.71686	0.01572
PUQE-score	0.16113	0.15982
Amnion	1.22094	0.08625
Proportional hazards model		
Weibull 1,1	0.07086	< 0.0001
Weibull 2,1	2.58852	< 0.0001
Weibull 1,2	0.08061	< 0.0001
Weibull 2,2	1.87717	< 0.0001
Fetal heartbeat	-0.74510	0.01319
Maternal age	0.12753	0.00574
Paternal age	-0.04840	0.15893
Linear mixed model		
Intercept 1	-40.87039	< 0.0001
Intercept 2	-26.33839	< 0.0001
Gestational age 1	0.97598	< 0.0001
Gestational age 2	0.51202	< 0.0001
Fetal heartbeat	2.72209	< 0.0001
Maternal age	0.31057	< 0.0001
PUQE-score	0.20528	0.07480

Class 2 is the reference class for the multinomial logistic regression model. The indices 1,1-2,2 in the proportional hazards model denote the two parameters of the Weibull baseline function for the two latent classes. The numbers 1 and 2 in the linear mixed model denote the two latent classes.

Table 26: Parameter estimates and p-values for MSD-model 3.

Variable	Estimate	P-value
Multinomial logistic regression model		
Intercept 1	-0.34178	0.86988
Intercept 2	0.55636	0.78694
Fetal heartbeat 1	-0.48927	0.50328
Fetal heartbeat 2	0.52683	0.43956
PUQE-score 1	0.64857	0.31363
PUQE-score 2	0.79260	0.21438
Amnion 1	7.25963	0.88480
Amnion 2	8.34027	0.86789
Proportional hazards model		
Weibull 1,1	0.11437	< 0.0001
Weibull 2,1	2.73037	< 0.0001
Weibull 1,2	0.09065	< 0.0001
Weibull 2,2	3.81552	< 0.0001
Weibull 1,3	0.01242	0.99242
Weibull 2,3	2.36687	0.99245
Fetal heartbeat	-0.02863	0.93783
Maternal age	0.14631	0.00235
Paternal age	-0.11401	0.00060
Linear mixed model		
Intercept 1	-23.73738	< 0.0001
Intercept 2	-39.89915	< 0.0001
Intercept 3	-69.44138	< 0.0001
Gestational age 1	0.49655	< 0.0001
Gestational age 2	0.97241	< 0.0001
Gestational age 3	0.95169	< 0.0001
Fetal heartbeat	2.50250	< 0.0001
Maternal age	0.29570	< 0.0001
PUQE-score	0.18936	0.08039

Class 3 is the reference class for the multinomial logistic regression model. The indices 1,1-2,3 in the proportional hazards model denote the two parameters of the Weibull baseline function for the three latent classes. The numbers 1-3 in the linear mixed model denote the three latent classes.

Bibliography

[1] Sinad M. O'Neill, Patricia M. Kearney, Louise C. Kenny, Ali S. Khashan, Tine B. Henriksen, Jennifer E. Lutomski, and Richard A. Greene. Caesarean Delivery and Subsequent Stillbirth or Miscarriage: Systematic Review and Meta-Analysis. *PLoS ONE*, 8(1), 2013.

[2] Yan Yi, Guangxiu Lu, Yan Ouyang, Ge Lin, Fei Gong, and Xihong Li. A logistic model to predict early pregnancy loss following in vitro fertilization based on 2601 infertility patients. *Reproductive Biology and Endocrinology*, 14(1):1–7, 2016.

[3] Nicole Stamatopoulos, Chuan Lu, Ishwari Casikar, Shannon Reid, Max Mongelli, Nigel Hardy, and George Condous. Prediction of subsequent miscarriage risk in women who present with a viable pregnancy at the first early pregnancy scan. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 55(5):464–472, 2015.

[4] Cecilia Bottomley, Vanya Van Belle, Emma Kirk, Sabine Van Huffel, Dirk Timmerman, and Tom Bourne. Accurate prediction of pregnancy viability by means of a simple scoring system. *Human Reproduction*, 28(1):68–76, 2013.

[5] Kelly M Mcnamee, Feroza Dawood, and Roy G Farquharson. Mid-Trimester Pregnancy Loss. *Obstet Gynecol Clin N Am*, 41:87–102, 2014.

[6] Ingrid H Lok and Richard Neugebauer. Psychological morbidity following miscarriage. *Best Practice & Research Clinical Obstetrics and Gynaecology*, 21(2), 2007.

[7] Leong Jin Kouk, Ghim Hoe Neo, Rahul Malhotra, John Carson Allen, Suan Tiong Beh, Thiam Chye Tan, and Truls Østbye. A prospective study of risk factors for first trimester miscarriage in Asian women with threatened miscarriage. *Singapore Medical Journal*, 54(8):425–431, 2013.

[8] S. Feodor Nilsson, PK. Andersen, K. Strandberg-Larsen, and A-M. Nybo Andersen. Risk factors for miscarriage from a prevention perspective: a nationwide follow-up study. *Royal College of Obstetricians and Gynaecologists*, 2014.

[9] Katherine J. Sapra, K. S. Joseph, Sandro Galea, Lisa M. Bates, Germaine M.Buck Louis, and Cande V. Ananth. Signs and symptoms of early pregnancy loss: A systematic review. *Reproductive Sciences*, 24(4):502–513, 2017.

[10] Courtney D Lynch and Jos M Maisog. Lifestyle and Pregnancy Loss in a Contemporary Cohort of Women Recruited Prior to Conception, LIFE Study. *HHS Public Access*, 106(1):180–188, 2017.

[11] P. Falco, V. Milano, G. Pilu, C. David, G. Grisolla, N. Rizzo, and L. Bovicelli. Sonography of pregnancies with first-trimester bleeding and a viable embryo: A study of prognostic indicators by logistic regression analysis, 1996.

[12] S. Choong, L. Rombauts, A. Ugoni, and S. Meagher. Ultrasound prediction of risk of spontaneous miscarriage in live embryos from assisted conceptions. *Ultrasound in Obstetrics and Gynecology*, 22(6):571–577, 2003.

[13] Bruno C. Casanova, Mary D. Sammel, Jesse Chittams, Kelly Timbers, Jennifer L. Kulp, and Kurt T. Barnhart. Prediction of Outcome in Women with Symptomatic First-Trimester Pregnancy: Focus on Intrauterine Rather Than Ectopic Gestation. *Journal of Women's Health*, 18(2):195–200, 2009.

[14] Megan Othus, Bart Barlogie, Michael L. LeBlanc, and John J. Crowley. Cure models as a useful statistical tool for analyzing survival. *Clinical Cancer Research*, 18(14):3731–3736, 2012.

[15] J. Elson, R. Salim, A. Tailor, S. Banerjee, N. Zosmer, and D. Jurkovic. Prediction of early pregnancy viability in the absence of an ultrasonically detectable embryo. *Ultrasound in Obstetrics and Gynecology*, 21(1):57–61, 2003.

[16] Cecile Proust-Lima and Jeremy M.G. Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of post-treatment PSA: A joint modeling approach. *Biostatistics*, 10(3):535–549, 2009.

[17] Cecile Proust-Lima, Viviane Philipps, and Benoit Liquet. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. *Journal of Statistical Software*, 78(2), 2015.

[18] Gideon Koren, Radinka Boskovic, Marjie Hard, Caroline Maltepe, Yvette Navioz, and Adrienne Einarson. Motherisk - PUQE (pregnancy-unique quantification of emesis and nausea) scoring system for nausea and vomiting of pregnancy. *American Journal of Obstetrics and Gynecology*, 186(5):228–231, 2002.

[19] Cecile Proust-Lima, Mbery Sene, Jeremy M G Taylor, and Helene Jacqmin-Gadda. Joint latent class models for longitudinal and time-to-event data: A review. *Statistical Methods in Medical Research*, 23(1):74–90, 2014.

[20] Nan M Laird and James H Ware. Random-Effects Models for Longitudinal Data Author (s): Nan M . Laird and James H . Ware Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2529876 REFERENCES Linked references are available on JSTOR for this article. *Biometrics*, 38(4):963–974, 1982.

[21] Helene Jacqmin-Gadda, Cecile Proust-Lima, Jeremy M.G. Taylor, and Daniel Commenges. Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. *Biometrics*, 66(1):11–19, 2010.

[22] Cecile Proust and Helene Jacqmin-Gadda. Estimation of linear mixed models with a mixture of distribution for the random effects. *Computer Methods and Programs in Biomedicine*, 78(2):165–173, 2005.

[23] Donald W. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[24] Christophe Biernacki, Gilles Celeux, and Grard Govaert. Choosing starting values for the EM algorithm for getting the highest likehood in multivariate Gaussian mixture models. *Computational Statistics and Data Analysis*, 41(3-4):561–575, 2003.

[25] John R. Hipp and Daniel J. Bauer. Local solutions in the estimation of growth mixture models. *Psychological Methods*, 11(1):36–53, 2006.

[26] Dollena S. Hawkins, David M. Allen, and Arnold J. Stromberg. Determining the number of components in mixtures of linear models. *Computational Statistics and Data Analysis*, 38(1):15–48, 2001.

[27] Daniel Commenges, Benoit Liquet, and Cecile Proust-Lima. Choice of Prognostic Estimators in Joint Models by Estimating Differences of Expected Conditional Kullback-Leibler Risks. *Biometrics*, 68(2):380–387, 2012.

[28] Daniel Commenges, Cecile Proust-Lima, Cecilia Samieri, and Benoit Liquet. A universal approximate cross-validation criterion for regular risk functions. *International Journal of Biostatistics*, 11(1):51–67, 2015.

[29] Michael S Wulfsohn, Gilead Sciences, Lakeside Drive, and Foster City. A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics*, 53(1):330–339, 1997.

[30] Menggang Yu, Jeremy M.G. Taylor, and Howard M. Sandler. Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association*, 103(481):178–187, 2008.

[31] Dimitris Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data: With Applictions in R.* Chapman & Hall/CRC Biostatistics Series, 2012.

Department of Electrical Engineering (ESAT)
Kasteelpark Arenberg 10 bus 2440
3001 LEUVEN, BELGIË
tel. + 32 16 32 11 30
fax + 32 16 32 19 86
www.esat.kuleuven.be