Modeling Longitudinal Data of Blood Donors

Kazem Nasserinejad

Printed by: Ipskamp drukkers, Enschede
Cover design: Johan de Rooi and Kazem Nasserinejad

ISBN: 978-94-028-0247-4

The work presented in this thesis was accomplished at department of Biostatistics at Erasmus Medical Center in Rotterdam and was partially financially supported by Sanquin Blood Supply Foundation in the Netherlands.

Modeling Longitudinal Data of Blood Donors

Modelleren van longitudinale data van bloeddonoren

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
woensdag 7 september 2016 om 13:30 uur

door

Kazem Nasserinejad

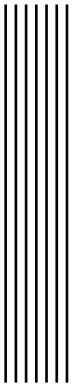
geboren te Mashhad, Iran

Erasmus University Rotterdam

Promotiecommissie

Promotoren:	Prof.dr. E. Lesaffre Prof.dr. W. de Kort
Overige leden:	Prof.dr E. W. Steyerberg Dr. D. Rizopoulos Dr. R.B. Geskus
Co-promotor:	Dr. J. van Rosmalen

To all blood donors



Contents

1 General introduction

1.1 Introduction

Blood transfusion is an essential part of modern healthcare, which helps save millions of lives each year. Since blood is a unique resource for which an artificial substitute has yet to be found, blood donations are in great need. There are several types of blood donation, but the most common type is whole blood donation. Whole blood is a term used in transfusion medicine for a standard (500 ml) blood donation as opposed to plasma and platelet donation.

Although blood donations and subsequent transfusions are meant to help patients, they might be harmful for both the recipients and the donors. For example, blood products are associated with the possibility of transmission of infectious agents (Murphy, 2002). Therefore, to secure blood safety and donor health, donors must fulfill a number of eligibility criteria prior to each donation.

The most important potential harm for donors is that whole blood donation causes a loss of iron and blood cells. After a donation, on average a male donor loses 242 mg and a female donor 217 mg of iron (Simon, 2002), which is 4-8% of total body iron. Since iron is the most important element of the Hb protein, the loss of iron can lead to depleted iron stores and lowering of Hb levels (Skikne et al., 1984; Amrein et al., 2012). Potential symptoms of iron deficiency include fatigue, decreased physical endurance and work capacity, and impairment in attention, concentration, and other cognitive functions (Brittenham, 2011; Popovsky, 2012; Newman et al., 2006; Dallman, 1986; Schiepers et al., 2010). Restless legs syndrome, a neurologic disorder with irresistible need to move the legs, and pica, a disorder in which a person is craving and consuming nonnutritive substances, have also been repeatedly linked to blood donation related iron deficiency (Ulfberg and Nyström, 2004; Birgegård et al., 2010; Bryant et al., 2013; Spencer et al., 2013). Repeated donations could lead to iron depletion and ultimately to anemia (Skikne et al., 1984; Brittenham, 2011).

To protect donors from developing iron deficiency anemia after blood donation, the Hb value of blood donors is assessed prior to each donation. Donors with a Hb value that is too low are deferred from donation to protect donor health. Furthermore, deferral of donors with too low Hb values ensures that the quality of blood units for transfusion meets the required standards for Hb content (Baart, 2013). The minimum Hb values for donation are 8.4 mmol/l (135 g/L) for men and 7.8 mmol/l (125 g/L) for women based on the European Commission Directive (Baart, 2013). Each year a considerable proportion of prospective blood donors are temporarily deferred from donation because of too low Hb values (Popovsky, 2012; Newman, 2004). For instance in the Netherlands in 2011, men and women were deferred in 2.2% and 5.5% of the visits to the blood bank due to a low Hb value. Although deferrals are meant to protect donors and the quality of blood units for transfusion, Hb deferrals decrease the cost-effectiveness of blood supply, because (i) testing and deferring a donor are expensive; (ii) for every deferred donor, another donor

needs to be invited to reach collection targets; and (iii) lapsing donors need to be replaced by new donors because deferred candidates rarely return for donation (Halperin et al., 1998).

Hb recovery process

The loss of iron due to blood donation causes the Hb concentration to decrease and to reach its nadir a few days after donation (Wadsworth, 1955; Boulton, 2004; Kiss et al., 2015). The Hb value will then gradually recover to its pre-donation value (Wadsworth, 1955; Boulton, 2004; Kiss et al., 2015). In healthy donors with sufficient iron stores this may not be problematic. Repeated donations could however deplete iron stores, leading to iron depletion and ultimately anemia (Skikne et al., 1984; Brittenham, 2011). The time needed for Hb value to return to its pre-donation value is the so-called recovery time. This recovery time varies between individuals and might depend on some biological factors.
Basically, when the iron intake is not sufficient to replenish the iron loss due to donation or the interval between two successive donations is too short (that Hb value does not recover completely), a negative iron balance occurs. This results in gradually declining Hb values over time, which are currently not detected until the donor meets any of the deferral criteria. Therefore, blood donors on average need several weeks to replenish the lost iron after a blood donation (Fowler and Barer, 1942). However, there are wide variations in the duration of the recovery period among individuals.
Already in the 1940s, it was estimated that the body needs around 50 days to recover to pre-donation Hb levels (Fowler and Barer, 1942; Alstead, 1943). Therefore, currently guidelines for blood donation in the Netherlands impose a minimum interval of 56 days between donations, with a yearly maximum of 5 donations for men and 3 for women (Baart, 2013). A declining trajectory has been observed in Hb values for some individuals, and therefore these guidelines may not be safe for all individuals. The Hb recovery process after blood donation is illustrated in Figure 1.1. In this figure δ indicates the time that Hb reaches its minimum value after donation, RT indicates the recovery time that is needed for Hb to return to its pre-donation value, and θ shows the amount of reduction in Hb after donation.

Factors associated with Hb value

Several factors are known to be associated with the Hb value and hence may be used as predictors for Hb, i.e., sex (Yip et al., 1984), season (Hoekstra et al., 2007), age (Yip et al., 1984), and body mass index (BMI) (Yip et al., 1984). Men on average have higher Hb values than women. The Hb value decreases with increasing age in men, whereas in women, the Hb value rises by the effect of menopause due to hormonal changes and

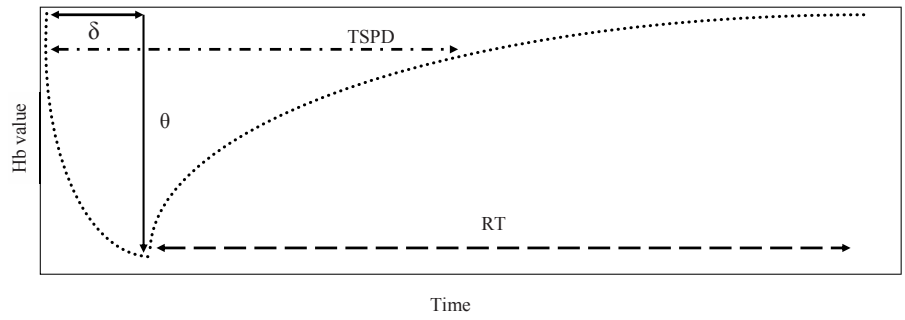


Figure 1.1: Hb recovery process after blood donation: δ indicates the time that Hb reaches its minimum value after donation, RT indicates the recovery time that is needed for Hb to return to its pre-donation value, and θ shows the amount of reduction in Hb after donation.

the cessation of iron loss through menstruation (Yip et al., 1984). Although women have a lower threshold for the Hb value to be eligible for donation, Hb deferral occurs more frequently in women than in men (Baart, 2013). BMI is also associated with Hb levels, i.e., a greater BMI is associated with higher Hb levels (Micozzi et al., 1989). Furthermore, donation history including number of donation in the past and the level of Hb in the previous visits/donations may affect the Hb value. The longer the interval between two donations, the more time for the donor to recover from the previous donation (Garry et al., 1995; Simon et al., 1981; Finch et al., 1977). A list of factors potentially associated with Hb value can be found in previous literature (Baart, 2013).

Sanquin

Sanquin is the national blood service in the Netherlands. In Sanquin donors must be between 18 and 70 years old, and prior to each donation donors must fill out an eligibility questionnaire to identify known medical conditions and high-risk behaviors. In a physical examination, body weight, pulse rate, blood pressure and the hemoglobin concentration[1] (Hb) value are measured. According to European guidelines, donors must have a body weight of at least 50 kg, a regular pulse, a systolic blood pressure between 90 and 180 mmHg, a diastolic blood pressure between 50 and 100 mmHg, and a Hb value of at least 8.4 mmol/L (135 g/L) for men and 7.8 mmol/L (125 g/L) for women. For blood collection, all measured data were entered into the blood-bank computer system (e)PROGESA (MAK-SYSTEM International Group, France). Prior to donation, Hb and other parameters undergo a check to determine whether the prospective donor is eligible. Sanquin dose not allow a newly registered donor donates blood at the first visit (i.e., the screening visit),

[1]Hereafter, whenever we say hemoglobin we mean hemoglobin concentration.

which consists of a health check only. At every subsequent visit, donors who pass all eligibility checks may donate 500 ml of whole blood. Finally, guidelines impose a minimum interval of 56 days between donations, with a yearly maximum of 5 donations for men and 3 for women (Baart, 2013). In 2014, 312,206 whole blood donors and 441,403 whole blood donations were registered at Sanquin.

Donor InSight data

The Donor InSight Study is a large prospective cohort study that was conducted by Sanquin among a random sample of 50,000 whole blood and plasma donors in 2007-2009. This data set is based on a self-administered questionnaire aimed at gaining insight into characteristics and motivation of the Dutch donor population (Atsma et al., 2011). In Figure 1.2, profiles of the Hb value after the screening visit for a subset of male and female donors in the Donor InSight data are displayed. The horizontal lines represent the eligibility thresholds for donation. The Donor InSight study was approved by the Medical Ethical Committee Arnhem-Nijmegen in the Netherlands and all participants gave their written informed consent. Further information about the Donor InSight study can be found on the Sanquin website at http://www.sanquin.nl/en/research/donor-insight/. The Donor InSight data set has been used in several chapters of this thesis.

1.2 Modeling longitudinal data of blood donors

Due to the fact that there are repeated measurements of Hb per donor, the observed Hb values can be seen as a longitudinal data set. To analyze the recovery, and to predict future Hb values and future rejections due to low Hb, statistical methods for longitudinal data are required. A large number of such techniques are available, such as linear mixed models (Molenberghs and Verbeke, 2001) and transition models (Diggle et al., 2002).
The longitudinal data of Hb values of blood donors have a number of unique aspects, namely a) heterogeneity of the initial Hb value, b) state dependence (i.e., within-subject correlations) of a donor's Hb values, c) varying time intervals between donations, d) the temporary reduction in Hb after blood donation, and e) the fact that the recovery process may change with the number of donations and may differ between donors. Due to these aspects, existing methods for analyzing longitudinal data are not directly applicable. Thus the new statistical methodology is required to properly analyze these blood donation data.
In analyzing longitudinal and repeated measures data, the central problem is the presence of association between repeated measures that were recorded for the same individual. For valid inferences, this association should be taken into account. The association structures can be defined as **true** and **apparent** contagion. In the former case, actual and future observation are directly influenced by previous observation(s). The latter case arise

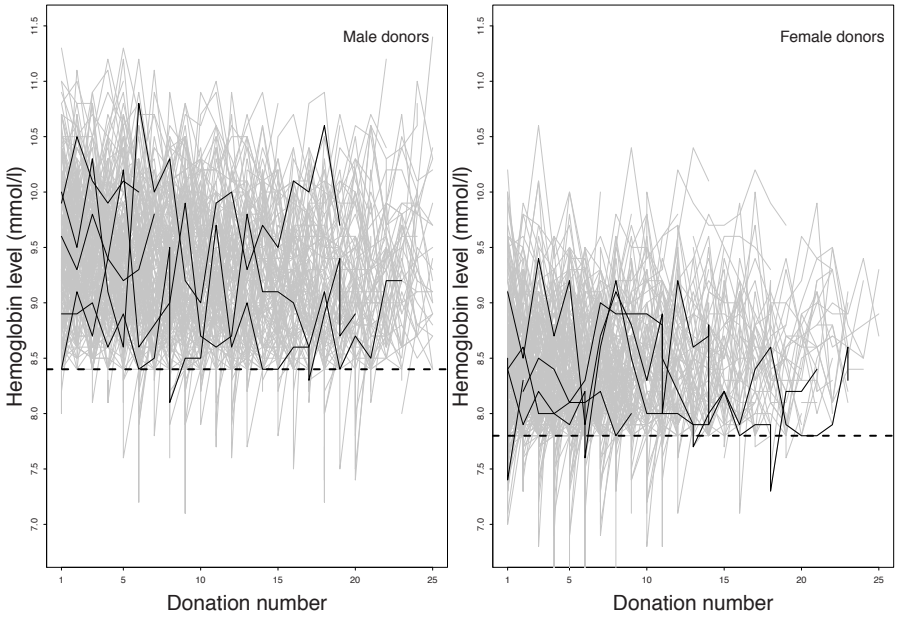


Figure 1.2: Hb profiles after the screening visit for a subset of male and female donors. Five random profiles are highlighted for both sexes. The bold dashed lines show the corresponding thresholds of eligibility for donation.

when individuals are drawn from **heterogeneous** populations, each population having a constant, but different, evolution (Aitkin and Alfò, 2003). Basically, considering a single individual (conditionally), previous observations do not influence the subsequent observations. However, the aggregate analysis of heterogeneous individuals could produce a misleading statistical finding, since this unobserved but persistent heterogeneity can be interpreted as being caused by a strong serial dependence. The distinction between true and apparent contagion is not always straightforward (Feller, 1943).

1.3 Overview of statistical models for longitudinal data

This section provides an overview of the statistical techniques that are applied in this thesis. This includes the statistical approaches that are usually applied in the analysis of longitudinal data. Namely, we will describe mixed effects models, transition models and mixture models, which cover several chapters of this thesis. The statistical techniques in this thesis were implemented in both Bayesian and frequentist frameworks.

Linear mixed-effects models

Linear mixed effects models (LMMs) are linear models with both fixed and random effects. A specific case of a LMM is a longitudinal growth study, where the baseline responses for the individuals differ but their linear growth is the same. This yields the random intercepts model, given by:

$$y_{it} = b_i + \beta^{'} x_{it} + \epsilon_{it}, \tag{1.1}$$

where y_{it} is the tth observation of the ith individual, x_{it} represents the q-dimensional covariate vector with fixed effects vector β having length q, and b_i represents the random intercept. It is assumed that b_i and ϵ_{it} are normally distributed and mutually independent with mean zero and different constant variances, i.e., $b_i \sim N(0, \sigma_b^2)$, and $\epsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ (Molenberghs and Verbeke, 2001). Furthermore, in the random intercept model the correlation between two observations of a subject is constant and is equal to the intra-class correlation given by $\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2}$ (Verbeke and Molenberghs, 2000; McCulloch and Neuhaus, 2001).

The linear mixed effects model in 1.1 can be extended by adding more random effects on top of the random intercept. For example in the previous longitudinal growth study, not only the baseline measurement, but also the true linear growth trend for each individual can be assumed to be different. This results in a linear random intercept and slope model. In general, the linear mixed-effects model can be written as follows:

$$y_{it} = b_i^{'} z_{it} + \beta^{'} x_{it} + \epsilon_{it}, \tag{1.2}$$

where b_i represents the q-dimensional random-effects with a multivariate normal distribution having mean zero and covariance matrix Σ_b, and z_{it} is the corresponding vector of q covariates. The other terms are the same as in 1.1. Mixed-effects models are the most famous models for capturing apparent contagion (heterogeneity) between different individuals in longitudinal studies.

Transition models

A transition model, also known as an autoregressive panel data model in the econometrics literature, is a dynamic regression model, in which the current response of a particular individual is regressed on previous responses of that subject as well as on other covariates (Diggle et al., 2002). A transition model of order q can be expressed as:

$$y_{it} = \beta^{'} x_{it} + \sum_{r=1}^{q} \gamma_r y_{it-r} + \epsilon_{it}, \tag{1.3}$$

where y_{it} is the tth observation of the ith individual, x_{it} represents the q-dimensional covariates vector with fixed effects vector β having length q, y_{it-r} is the rth lagged response and γ_r is the corresponding coefficient of the rth lag. Classically it is assumed that

the residuals ϵ_{it} are normally distributed and mutually independent with mean zero and constant variance, i.e., $\epsilon_{it} \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$. In a transition model with order q, the predicted values depend on q lagged previous observations; however, to calculate the predicted value using 1.3, there are not enough previous observations for the first few visits of a donor. Transition models are well-known models for capturing true contagion (state dependence) between observations in longitudinal studies.

Mixed-effects transition model

In some cases neither of the aforementioned approaches (i.e., mixed-effects model and transition model) are able to adequately explain the correlation structure alone, due to the presence of both unobserved heterogeneity and state dependence. For these reasons, we may combine the mixed-effects model and the transition model by including unobserved individual-specific effects and a lagged endogenous variable in a single regression model. Such a mixed-effects transition model is popular in econometrics for forecasting (Diggle et al., 2002), but less commonly used in medical applications (Funatogawa and Funatogawa, 2012). This model is given by:

$$y_{it} = b_{i1} + \beta^{'} x_{it} + \gamma y_{it-1} + \epsilon_{it}, \tag{1.4}$$

where b_{i1} controls the heterogeneity now partly explaining the intra-subject correlation, and γ is the lagged impact of the previous observation (y_{it-1}). For a stationary process, i.e., $|\gamma| < 1$, the correlation between two subsequent measurements, can be expressed as (Rikhtehgaran et al., 2012):

$$\rho_{y_{it}, y_{it-1}} = \gamma + \frac{1-\gamma}{1 + (1-\gamma)\sigma_\varepsilon^2 / [(1+\gamma)\sigma_b^2]}. \tag{1.5}$$

When the lag impact γ is negligible, this correlation reduces to the intra-class correlation (ICC) in the mixed-effects model. On the other hand, when there is no heterogeneity between individuals, i.e., $\sigma_b^2 \approx 0$, the correlation is equal to the lag impact only.

Mixed-effects model with serial correlation

Another possible solution for capturing true contagion is a model with serial correlation (Anderson and Hsiao, 1982). This model can be given as follows:

$$y_{it} = b_i + \beta^{'} x_{it} + \gamma(y_{it-1} - \beta^{'} x_{it-1}) + \epsilon_{it}. \tag{1.6}$$

In this model, in contrast with the mixed-effects transition model the current response is affected merely by current covariates and not by previous covariates i.e., x_{it-1}, x_{it-2}, ..., x_{it-q}. In other words, in a transition model a change in a time-dependent covariate (x) that affects the distribution of y in the current period will also affect this distribution in

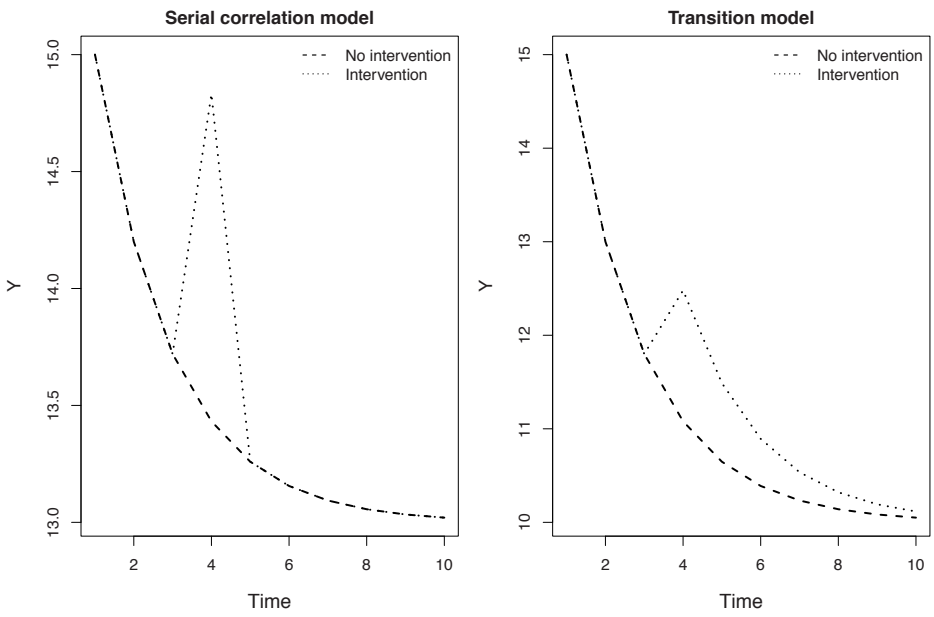


Figure 1.3: Intervention on time-dependent covariate and its impact on transition and serial correlation models.

forthcoming periods (Funatogawa et al., 2007). This issue is exemplified here. Assumed a transition model and a serial correlation model as 1.7 and 1.8, respectively. Where the inital observation is 15 (i.e., $y_1 = 15$) and $x_t = 10$ for all times ($t = 1, \ldots, 10$). We apply an intervention to x_t at time $t = 4$ and change $x_4 = 5$. This intervention at time $t = 4$ has an effect only on y_4 in 1.7. However, in 1.8 the intervention has an effect not only on y_4, but also on later observations, i.e., y_5, ..., y_{10}. Figure 1.3 illustrates the impact of this intervention in both transition and serial correlation models.
A serial correlation model:

$$y_t = 6 - 0.2x_t + 0.6(y_{t-1} + 0.2x_{t-1}) + \epsilon_{it}, \tag{1.7}$$

A transition model:

$$y_t = 6 - 0.2x_t + 0.6y_{t-1} + \epsilon_{it}. \tag{1.8}$$

Growth mixture models

In the mixed-effects models framework, the intercept (and slope) vary across individuals and the heterogeneity between individuals is captured by random effects (i.e., continuous

latent variables). This approach assumes that all individuals come from a single population and that a single growth trajectory can adequately approximate an entire population. Also, it is assumed that covariates that affect the growth trajectory influence each individual in the same way. Therefore, this approach may not be able to capture entire heterogeneity and might oversimplify the complex growth pattern in population.
To solve this problem, growth mixture modeling has been proposed, which is a statistical technique that allows for heterogeneity in the growth trajectories. This heterogeneity is presented by a latent categorical variable that defines k latent classes of individuals, each of which are potentially described by a unique set of growth model parameters. Each of these latent classes is then characterized by a mixed-effects model, so that heterogeneity within each group is captured by random effects. The objective of these approaches is to capture information about inter-individual differences in intra-individual change over time (Nesselroade, 1991). The density function of $\mathbf{y}$ in a growth mixture model can be expressed as:

$$f(\mathbf{y}|\boldsymbol{\lambda}, \boldsymbol{\theta}) = \sum_{k=1}^{K} \lambda_k f_k(\mathbf{y}|\theta_k),$$

where $f_k(y|\theta_k)$ $(k = 1, \ldots, K)$ are density functions that describe the trajectory for each class. The w_ks represent the proportion of subjects in each class, and must satisfy $0 \leq \lambda_k \leq 1$ and $\sum_{k=1}^{K} \lambda_k = 1$ (McLachlan and Peel, 2004). The vector θ_k represents the parameters that are associated with the trajectory of class k. Latent class growth analysis (LCGA) is a special type of GMM, where the variance and covariance estimates for the growth factors within each class are assumed to be fixed to zero. LCGA assumes that individuals heterogeneity is completely captured by the different growth trajectories of the k latent classes, which means all individual growth trajectories within a class are assumed to be homogeneous (Tofighi and Enders, 2008).

1.4 Bayesian approach

The central idea of the Bayesian approach is to combine the likelihood **(data)** with the prior knowledge **(prior probability)** to result in a revised probability **(posterior probability)**. The parameter θ is given a probability distribution, which is in contrast to the frequentist approach, where it is a fixed but unknown value. The parameters become stochastic due to our uncertainty of their values. We denote $p(\theta)$ as the prior distribution of θ that can be specified on the basis of historical studies and/or with **expert** knowledge, but without observing the current data y. $L(\theta|y)$ is the likelihood defined by the model specification. The probability distribution of θ, which is called the posterior distribution, is obtained by combining the information from the prior and the data according to Bayes' Theorem:

$$p(\theta|y) = \frac{L(\theta|y) \times p(\theta)}{p(y)}. \tag{1.9}$$

The denominator $p(y)$ can be written as the integral of the likelihood $L(\theta|y)$ with respect to the variable θ, and is therefore called the averaged likelihood. Bayes' theorem shows one of the advantages of the Bayesian approach, which can utilize a priori information to increase the power for estimation. Another merit of the Bayesian approach is that the parameters are stochastic, which means uncertainty of their values is taken into account, which makes the interpretation of parameter estimates straightforward.
In case there is no prior information about the parameters in the model, a non-informative prior is often implemented. However, there exists no purely non-informative prior. Some relatively uninformative priors were suggested as vague priors e.g., Jeffreys priors (Jeffreys, 1946) or reference priors (Bernardo, 1979). The aim of using vague priors is to ensure the impact of the prior on the posterior is minimal, so that the posterior represents information in the data. This kind of Bayesian analysis is called **objective Bayesian** in contrast to **subjective Bayesian**.
Note that in the denominator of 1.9, each integral may involve multiple integrations depending on their respective dimensions. Because of the high-dimensional integration, the Bayesian approach was for about two centuries impossible to use for real-life problems (Lesaffre and Lawson, 2012).

Bayesian computational techniques/softwares

Computing a posterior distribution analytically is almost impossible. There are several techniques to estimate the posterior. In the 1980s and 1990s a powerful class of numerical procedures, called Markov Chain Monte Carlo (MCMC) techniques (Gelfand and Smith, 1990), was launched that revolutionized the Bayesian approach. The MCMC technique is based on a sampling approach, i.e., the integral is approximated by Monte Carlo sampling (Ripley, 1988). There are two major classes of MCMC techniques: Gibbs sampling and Metropolis-Hastings (MH) sampling. These sampling techniques are available in standard open access Bayesian software e.g., WinBugs (Lunn et al., 2000), OpenBugs (Thomas and OHara, 2004), Jags (Plummer, 2011). Other sampling techniques have recently been developed e.g., Hamiltonian or Hybrid Monte-Carlo Sampling (HMC) and The No-U-Turn Sampler (NUTS) (Homan and Gelman, 2014), which can be implemented via Stan (Stan, 2015). Alternative methods of inference are Approximate Bayesian Computation (ABC) and Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009), which can be implemented in R-INLA (Martins et al., 2013) in R. Thanks to modern and fast computers, Bayesian statistics have dramatically increased in popularity in the last decades.

1.5 List of methodological issues

The statistical approaches discussed in the last two sections are relatively complex, and not always straightforward to implement in practice. There are several methodological issues

that need to be taken into account when applying these statistical techniques. Below, we give an overview of these issues, and some suggested solutions from the literature. These issues are addressed in the next chapters of this thesis.

The initial conditions problem

One of the assumptions in classical mixed-effects models is that the covariates in the model are exogenous, i.e., the covariance between the covariates and the random effects are zero. But this assumption is violated in mixed-effects transition models where one of the covariates is the lag variable, which is endogenous. This issue relates to the initial conditions problem (ICP), which is well-known in the econometrics literature. The ICP occurs due to the fact that the individual effects, b_{i1}, that capture the unobserved heterogeneity are correlated with the initial observations, i.e., $\text{cov}(y_{i1}, b_{i1}) \neq 0$ (Kazemi and Davies, 2002; Kazemi and Crouchley, 2006). Ignoring the ICP and thus the endogeneity of y_{i1} results in inconsistent estimates in the model (Kazemi and Davies, 2002; Kazemi and Crouchley, 2006), i.e., an upward bias of the estimated state dependence and a downward bias in the estimated coefficients of explanatory variables (Kazemi and Crouchley, 2006). Several solutions to the ICP have been proposed in the literature (Bhargava et al., 1983; Hsiao et al., 2002; Wooldridge, 2005). A possible solution is to incorporate the association of the initial observations and the random effects jointly with the model for the subsequent response observations. The model is assumed to be similar to the main model, but without the lagged response variables (Kazemi and Crouchley, 2006). Using this solution, the regression parameters as well as the residual variance are allowed to differ between the initial and the subsequent observations. The joint modeling approach enables one to capture the correlation between the individual effects, b_{i1}, and the initial observations and provides reliable estimates for the regression parameters (Kazemi and Crouchley, 2006).
A possible model for the initial values could be:

$$y_{i1} = \beta_0^{'} x_{i0} + b_{i0}, \tag{1.10}$$

where $b_{i0} \sim N(0, \sigma_{b0}^2)$, and the model allows for a correlation between b_{i0} and the random effects in the model for the subsequent values.

Choosing the number of clusters in latent class models

Identifying the number of correct latent classes (K) in finite mixture models, which can appropriately capture the entire heterogeneity in the population, is a challenging issue for researchers (Lee et al., 2008; McGrory and Titterington, 2007; Richardson and Green, 1997; Rousseau and Mengersen, 2011). Several criteria exist for choosing the number of latent classes in mixture models in both a frequentist and a Bayesian framework. Whereas information criteria such as the Akaike information criterion (AIC) (Akaike, 1973) and the

Bayesian information criterion (BIC) (Schwarz et al., 1978) seem to be the most popular criteria in a frequentist setting, no clear consensus on the optimal criterion in a Bayesian setting has yet emerged. Although the deviance information criterion (DIC) (Spiegelhalter et al., 2002) is a well-known criterion for comparing different Bayesian models, unfortunately this criterion is not suited to the case of mixture models. Several adaptations of this criterion for mixture models were proposed (Celeux et al., 2006). Alternatively, models with different numbers of latent classes can be compared by computing marginal likelihoods, Bayes factors, or by using reversible jump MCMC techniques (Green, 1995). Recently, Rousseau and Mengersen (2011) showed that in overfitted mixture models, the overfitted latent classes will asymptotically become empty if the Dirichlet hyperparameter is small enough. They suggested that this mathematical result can be used to choose the number of latent classes, by estimating the finite mixture model with a relatively number of classes and then omitting all classes that have a small proportion.

Change-point problems

The change-point problem occurs when data have a natural ordering. To tackle this issue a change-point model is needed. This model allows the sequence of data be broken down into segments with the observations following the same statistical model within each segment, but different models in different segments (Hawkins, 2001). The best-known application of change-point modeling in data analysis is that of regression trees. In the most widely used implementation (Breiman et al., 1984), the data set is ordered by a continuous or ordinal predictor and then split into two subsequences those cases whose predictor value falls below some change-point and those whose predictor value is above the change-point. The change-point is chosen to maximize the separation between the two subsequences (Hawkins, 2001). This change-point is normally unknown and must be estimated via data. The Bayesian change point model can be estimated using the Gibbs sampler (Western and Kleykamp, 2004).

Identifiability problem

The main identifiability issues in Bayesian statistics can be defined as follows.

Lack of information

Weak identifiability occurs when the technical conditions for identifiability are met but the data provide little information about the pertinent parameter (Carlin and Louis, 1997; Garrett and Zeger, 2000). Indeed, the only information that we have about the parameter, θ, is supplied by its prior ($p(\theta|y) \simeq p(\theta)$), and the posterior and the prior distribution are the same (Garrett and Zeger, 2000). A practical example of this issue is estimating the Hb

recovery process where the Hb values are only observed with at least a 56 days intervals between visits. The Hb recovery process was illustrated in Figure 1.1.

Label switching problem

An interesting, challenging issue that arises in the Bayesian analysis of mixture models is the non-identifiability of the latent classes. The problem is mainly caused by the invariance of the posterior distribution to permutations of the parameter labeling under symmetric priors and likelihood (Dellaportas and Papageorgiou, 2006). This leads to so called label switching in the MCMC output. This will mean that ergodic averages of component specific quantities will be identical and thus useless for inference (Jasra et al., 2005).
Various solutions have been proposed to tackle this problem, e.g., imposing artificial identifiability constraints (Richardson and Green, 1997), random permutation sampling (Frühwirth-Schnatter, 2001), relabeling algorithms (Celeux, 1998; Stephens, 2000), and label invariant loss functions methods (Celeux et al., 2000). The problem is further complicated in higher dimensions, since the number of identifiability constraints on the parameter space is very large and relabeling algorithms will require a lot of computing time (Dellaportas and Papageorgiou, 2006). Figure 1.4 shows the original (Panel A) and relabeled (Panel B) trace plots of Gibbs samples of a univariate gaussian mixture distribution with two latent classes.
Another challenging issue in the Bayesian analysis of mixture models is overparametrization. Overfitting induces a special type of non-identifiability in the posterior distribution of mixture models. Theoretically, any mixture distribution can be represented equally well by another mixture distribution with a larger number of latent classes, where some (extra) latent classes have either been merged together or have proportions equal to zero (Garrett and Zeger, 2000; van Havre et al., 2015).

1.6 List of research questions

A number of research questions will be in this thesis, with either a clinical or a methodological focus. These research questions are as follows.

- Modeling Hb values to have a better prediction for the future values
- Incorporating prior information to better estimate the Hb recovery process
- Modeling the heterogeneity in the trajectory of Hb values
- Identifying the appropriate number of latent classes in blood donors
- The ultimate aim is to determine the best time for the donor to return for next donation to avoid unnecessary deferral due to low Hb value.

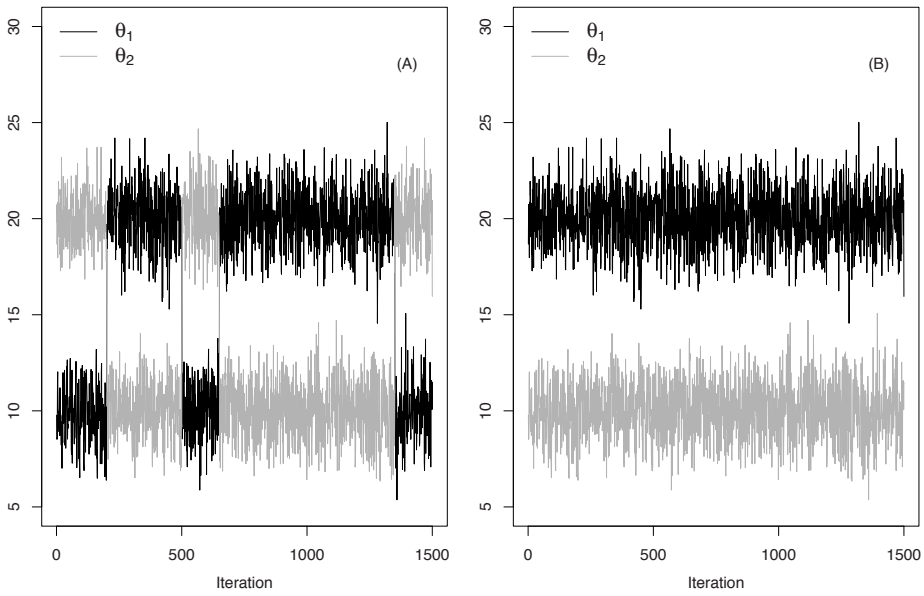


Figure 1.4: Label switching problem in fitting a mixture model on a univariate gaussian mixture distribution data with two latent classes, the original MCMC output (Panel A) and the relabeled MCMC output (Panel B).

1.7 Outline of this thesis

This section provides an overview of the separate chapters within this thesis. Per chapter the main problem and proposed solutions are shortly described. Often followed by some theoretical background concerning the used datasets.

Chapter 2. Predicting hemoglobin levels in whole blood donors using transition models and mixed effects models

To optimize the planning of blood donations but also to continue motivating the volunteers it is important to streamline the practical organization of the timing of donations. While donors are asked to return for donation after a suitable period of at least 56 days, still a relevant proportion of blood donors is deferred from donation each year due to a too low hemoglobin level. Rejection of donation may demotivate the candidate donor and implies an inefficient planning of the donation process. Hence, it is important to predict the future hemoglobin level to improve the planning of donors' visits to the blood bank. We developed a hemoglobin prediction rule based on longitudinal data from blood donations collected by Sanquin. We explored and contrasted two popular statistical models,

i.e., the transition model and the mixed effects model as plausible models to account for the dependence among subsequent hemoglobin levels within a donor. The predictors of the future hemoglobin level are age, season, hemoglobin levels at the previous visits, and a binary variable indicating whether a donation was made at the previous visit.

Chapter 3. Prevalence and determinants of declining versus stable hemoglobin levels in whole blood donors

A too short recovery time after blood donation results in a gradual depletion of iron stores and a subsequent decline in hemoglobin (Hb) levels over time. This decline in Hb levels may depend on individual, unobserved characteristics of the donor. We used a data set of 5388 Dutch blood donors from the Donor InSight study. The statistical analysis is based on a Bayesian growth mixture model, which assumes that each donor belongs to one of several groups. Each group implies a different Hb trajectory, and donors with similar longitudinal trajectories belong to the same group.

Chapter 4. Prediction of hemoglobin in blood donors using a latent class mixed-effects transition model

Blood donors experience a temporary reduction in their hemoglobin (Hb) value after donation. At each visit, the Hb value is measured, and a too low Hb value leads to a deferral for donation. Because of the recovery process after each donation as well as state dependence and unobserved heterogeneity, longitudinal data of Hb values of blood donors provide unique statistical challenges. To estimate the shape and duration of the recovery process, and to predict future Hb values, we employed three models for the Hb value: (i) a mixed-effects models; (ii) a latent-class mixed-effects model; and (iii) a latent-class mixed-effects transition model. In each model, a flexible function was used to model the recovery process after donation. The latent classes identify groups of donors with fast or slow recovery times and donors whose recovery time increases with the number of donations. The transition effect accounts for possible state dependence in the observed data. All models were estimated in a Bayesian way, using data of new entrant donors from the Donor InSight study. Informative priors were used for parameters of the recovery process that were not identified using the observed data, based on results from the clinical literature.

Chapter 5. Comparison of different criteria for choosing the number of latent classes

In this chapter we deal with identifying the number of latent classes in Bayesian finite mixture models. Identifying the number of latent classes in Bayesian finite mixture mod-

els is a challenging problem. Several criteria have been proposed, such as adaptations of the deviance information criterion, marginal likelihoods, Bayes factors, and reversible jump MCMC techniques. It was recently shown that in overfitted mixture models, the overfitted latent classes will asymptotically become empty under specific conditions for the prior of the class sizes. This result may be used to construct a criterion for finding the true number of latent classes, based on the removal of latent classes that have negligible proportions. Unlike some alternative criteria, this approach can easily be implemented in complex statistical models such as latent class mixed-effects models and multivariate mixture models using standard Bayesian software. We performed an extensive simulation study to develop practical guidelines to determine the appropriate number of latent classes based on the posterior distribution of the class proportions, and to compare this criterion with alternative criteria.

Chapter 6. Conclusions

In the last chapter of this thesis we summarize the major conclusions from the different chapters. In addition, a short more general discussion focuses on the interpretation of the statistical models. Proposals for future research are made where possible.

2 Predicting Hb levels using transition models and mixed-effects models

This chapter is published as: Nasserinejad K, de Kort W, Baart M, Komarek A, van Rosmalen J and Lesaffre E. Predicting hemoglobin levels in whole blood donors using transition models and mixed effects models. *BMC Medical Research Methodology*, 13:62, 2013. doi:10.1186/1471-2288-13-62

Abstract.

To optimize the planning of blood donations but also to continue motivating the volunteers it is important to streamline the practical organization of the timing of donations. While donors are asked to return for donation after a suitable period, still a relevant proportion of blood donors is deferred from donation each year due to a too low hemoglobin level.

Rejection of donation may demotivate the candidate donor and implies an inefficient planning of the donation process. Hence, it is important to predict the future hemoglobin level to improve the planning of donors' visits to the blood bank.

The development of the hemoglobin prediction rule is based on longitudinal (panel) data from blood donations collected by Sanquin (the only blood product collecting and supplying organization in the Netherlands). We explored and contrasted two popular statistical models, i.e., the transition (autoregressive) model and the mixed effects model as plausible models to account for the dependence among subsequent hemoglobin levels within a donor.

The predictors of the future hemoglobin level are age, season, hemoglobin levels at the previous visits, and a binary variable indicating whether a donation was made at the previous visit. Based on cross-validation, the areas under the receiver operating characteristic curve (AUCs) for male donors are 0.83 and 0.81 for the transition model and the mixed effects model, respectively; for female donors we obtained AUC values of 0.73 and 0.72 for the transition model and the mixed effects model, respectively.

We showed that the transition models and the mixed effects models provide a much better prediction compared to a multiple linear regression model. In general, the transition model provides a somewhat better prediction than the mixed effects model, especially at high visit numbers. In addition, the transition model offers a better trade-off between sensitivity and specificity when varying the cut-off values for eligibility in predicted values. Hence transition models make the prediction of hemoglobin level more precise and may lead to less deferral from donation in the future.

2.1 Introduction

Blood transfusion is an essential part of modern healthcare which helps save millions of lives each year. Since blood is a unique resource for which an artificial substitute has yet to be found, blood donations are in great need. However, occasionally donation cannot be accepted. There may be several reasons for the ineligibility of a blood donor for donation, a common reason being low hemoglobin level of the donor (Gómez-Simón et al., 2007; Tong et al., 2010). A hemoglobin (Hb) level of 8.4 mmol/l (135 g/l) and 7.8 mmol/l (125 g/l) for men and women, respectively, is widely accepted as the lower cut-off value of eligibility for donation (Tong et al., 2010; Radtke et al., 2005; Baart et al., 2011). While donors are asked to return for donation after a suitable period, a relevant proportion of blood donors are temporarily deferred from donation each year due to low Hb levels (Tong et al., 2010). Rejection of donation may demotivate the candidate donor and implies inefficient planning of the donation process (Halperin et al., 1998; Custer et al., 2007). Hence, it is important to predict the future Hb level to improve the planning of donors' visits to the blood bank. Prediction models for low Hb level deferral have been developed previously (Baart et al., 2011, 2012).
The main goal of this paper is to illustrate the use of two well-known longitudinal models in predicting the future Hb level after a visit to the blood bank. An adequate prediction will help the blood bank to apply appropriate interventions (e.g., postponing the next invitation) for blood donation when the Hb value falls below the cut-off value. Prediction is based on models developed using historical data of Hb levels obtained from Sanquin Blood Supply in the Netherlands. More specifically, in this paper we examine the predictive performance of the transition (autoregressive panel data) model and the mixed effects model.

2.2 Materials and methods

Data

The data have been obtained from Sanquin Blood Supply, which is the only blood product collecting and supplying organization in the Netherlands. In this paper, we analyze newly registered whole blood donors whose first visit to the collection centers occurred in the period between January 1, 2007 and December 31, 2009 and have donated at least twice during this period. Whole blood is a term used in transfusion medicine for a standard blood donation as opposed to plasma and platelet donation. The data were collected from 16,158 newly registered whole blood donors (54.6% women). The reason for selecting this set of blood donors is that they constitute a relatively homogeneous group that did not donate prior to establishing the Sanquin database. We excluded donors who had missing values for the Hb level, and the data of the remaining $15,625$ donors were used

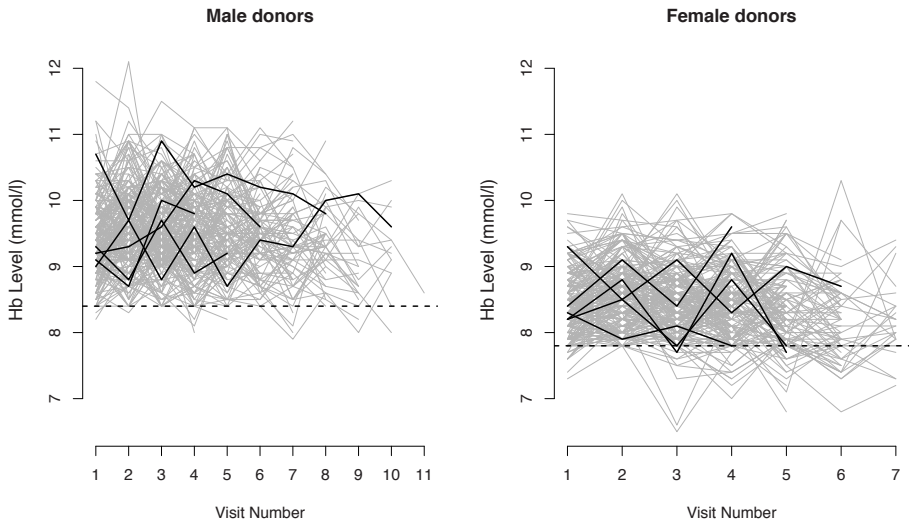


Figure 2.1: Profile of hemoglobin levels for successive visits to the blood bank of a random sample of male and female donors. The profiles of 5 randomly selected donors are highlighted. The dashed horizontal lines show the Hb cut-off values of eligibility for donation

in the analyses.

In Sanquin Blood Supply, a candidate has to register prior to donation; after registration he/she will receive an information package and an invitation to attend a blood donor health check. If the test results are satisfactory, the candidate will be invited to donate blood. Therefore, the first visit to the Sanquin Blood Supply is not a donation but a health check that includes a measurement of the Hb level. After a successful whole blood donation, a male (female) donor is allowed to return for the next donation after a period of at least 8 weeks with a maximum of 5 (3) donations per year.

In each visit, prior to donation, the candidates are screened for health risks that might make the donation unsafe for either the donor or the recipient. These tests include taking fingerstick capillary samples for measuring Hb level and filling out a health appraisal form. Based on the results of these tests, the candidate may not be eligible for donation due to a too low Hb level or other reasons that he/she mentioned in the health appraisal form. Finally, eligible candidates will donate 500 milliliters (ml) blood. We defined donation status in each visit as a binary variable in our data set (donation =1, no donation =0).

In Figure 2.1, profiles of the Hb level are displayed for male and female donors separately. The dashed horizontal lines show the corresponding Hb level cut-off points of eligibility for donation. Several factors are known to be associated with the Hb level and hence may be used as predictors for Hb level, i.e., sex (Yip et al., 1984), age (Yip et al., 1984), and body

mass index (BMI) (Micozzi et al., 1989; Skjelbakken et al., 2006).
In this study, we take into account the effect of sex and age in our models, but we decided to ignore the effect of BMI due to the fact that the BMI was not recorded for approximately 40% of donors. Also, based on a pilot study we found that the impact of BMI on Hb level is secondary. The season in which the visit takes place also affects the Hb level, namely in a warm season Hb level is lower on average (Hoekstra et al., 2007; Kristal-Boneh et al., 1993). Here season is used as a binary covariate, i.e., cold season (=0) includes fall and winter and warm season (=1) includes spring and summer. Male and female donors have different Hb profiles. Therefore, we analyzed the data for men and women separately.
Inter-visit intervals differ between donors, in our data set the median inter-visit interval for male donors is 72 (inter-quartile range: $29 - 92$) days and for female donors it is 93 (inter-quartile range: $25 - 131$) days. In principle, varying intervals between visits require continuous-time models, but these models are beyond the scope of this paper. Therefore, we decided to ignore this feature of the data, and we used the sequential number of the visit rather than the actual time of the visit. We also take into account the status of the previous visit (donation or deferral) as a binary covariate in the prediction model. Since no donations have been made prior to the first visit, the value of donation at previous visit (DPV) for the first visit is defined to be no donation.
This research has been performed with the approval of the ethical advisory council of the Sanquin Blood Supply Foundation. Moreover, all donors have given their consent by stating that part or all of their donations can be used for research aiming at improving the blood supply chain. Our ethical advisory council includes members of both Sanquin and non-Sanquin affiliations. This committee includes members with the background training and experience required for such ethical committees.

Statistical analysis

Since successive Hb levels on the same subject are correlated, we need to employ statistical models that can take this correlation into account. For this purpose, we applied two well-known models, namely the transition model and the mixed effects model. However, we commence with a multiple linear regression model as a benchmark to show the capability of transition and mixed effects models. These statistical analyses were performed in R (R Development Core Team, 2010) version 2.15.2 using the *stats package* for the multiple linear regression models, the *nlme package* for the mixed effects models, the *KalmanLike* and the *mle* functions in the *stats4 package* for the transition models, and the *mixAK* and *pROC packages* to draw profile and ROC curve plots. We used a significance level of $\alpha = 0.05$ and no correction for multiple testing was implemented.

Multiple linear regression model

A naive approach to analyze the successive Hb levels is a multiple linear regression model, in which the current response of a particular subject is regressed only on time-varying covariates, i.e., age, season, and DPV. A multiple linear regression model can be expressed as:

$$y_{it} = \alpha + \beta_1 \text{Age}_{it} + \beta_2 \text{Season}_{it} + \beta_3 \text{DPV}_{it} + \epsilon_{it}, \tag{2.1}$$

where y_{it} is the tth observation of the ith individual, α is an unknown constant (intercept), and the β's are unknown regression coefficients. It is assumed that the residuals ϵ_{it} are normally distributed and mutually independent with mean zero and constant variance, i.e., $\epsilon_{it} \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$. Due to the fact that this model cannot take into account the intra-subject correlations and the previous Hb levels, it is only presented as a benchmark model to show the capability of transition and mixed effects models.

Transition model

A transition model, also known as an autoregressive panel data model in the econometrics literature, is a dynamic regression model, in which the current response of a particular subject (donor) is regressed on previous responses of that subject as well as on other covariates (Diggle et al., 2002). A transition model of order q can be expressed as:

$$\begin{aligned} y_{it} = \alpha + \beta_1 \text{Age}_{it} + \beta_2 \text{Season}_{it} + \beta_3 \text{DPV}_{it} + \\ \sum_{r=1}^{q} \gamma_r (y_{it-r} - (\beta_1 \text{Age}_{it-r} + \beta_2 \text{Season}_{it-r} + \beta_3 \text{DPV}_{it-r})) + \epsilon_{it}, \end{aligned} \tag{2.2}$$

where y_{it} is the tth observation of the ith individual, α is an unknown constant, and the β's are unknown regression coefficients, y_{it-r} and $(Age_{it-r}, Season_{it-r}, DPV_{it-r})$ are rth lagged response and covariates, respectively and γ_r is the corresponding coefficient of the rth lag. Classically it is assumed that the residuals ϵ_{it} are normally distributed and mutually independent with mean zero and constant variance, i.e., $\epsilon_{it} \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$. In a transition model with order q, the predicted values depend on q lagged previous observations; however, to calculate the predicted value using 2.2, there are not enough previous observations for the first few visits of a donor. We employed the method of maximum likelihood via a linear quadratic estimation (Kalman filter) algorithm to estimate the parameters in the transition model. This algorithm enables us to calculate the exact likelihood function, which includes the distribution of the first few observations of each donor (Kalman, 1960; Harvey and Phillips, 1979; Tsay, 2010). As a result, the maximum likelihood estimation also includes the information of donors who have made fewer visits than the order of the transition model.

Linear mixed effects model

The linear mixed effects (LME) model which contains a mixture of fixed effects and random effects provides another way to deal with longitudinal responses within a subject. The correlation among responses pertaining to one subject is now induced by introducing random effects, which can be regarded as subject-specific terms (Verbeke and Molenberghs, 2000; McCulloch and Neuhaus, 2001). A special case of the mixed effects model is the random intercept model which can be expressed as:

$$y_{it} = \alpha + b_{0i} + \beta_1 \text{Age}_{it} + \beta_2 \text{Season}_{it} + \beta_3 \text{DPV}_{it} + \epsilon_{it}, \tag{2.3}$$

where α is an unknown constant, the β's are regression coefficients (fixed effects) and the b_{i0} is the random intercept. The random intercept b_{i0} can be viewed here as the deviation of the ith subject-specific mean of Hb levels from the population mean of Hb levels. It is assumed that b_{0i} and ϵ_{it} are normally distributed and mutually independent with mean zero and different constant variances, i.e., $b_{0i} \sim N(0, \sigma^2_{b0})$, and $\epsilon_{it} \sim N(0, \sigma^2_{\varepsilon})$ (Molenberghs and Verbeke, 2001). Furthermore, in the random intercept model the correlation between two observations of a subject is constant and is equal to the intra-class correlation given by $\rho = \frac{\sigma^2_{b0}}{\sigma^2_{b0}+\sigma^2_{\varepsilon}}$ (Verbeke and Molenberghs, 2000; McCulloch and Neuhaus, 2001).
Although the simplicity of the mixed model with only random intercept is appealing, it poses the restriction that the correlation between the repeated measurements remains constant over time. An extension that allows for a more flexible specification of the covariance structure is a mixed model with random intercept and slope; this model introduces an additional random effects term (e.g., age), and assumes that the rate of change in the covariates (age) differs between subjects. The mixed effects model with random intercept and slope can be expressed as:

$$y_{it} = \alpha + b_{0i} + (b_{1i} + \beta_1)\text{Age}_{it} + \beta_2 \text{Season}_{it} + \beta_3 \text{DPV}_{it} + \epsilon_{it}, \tag{2.4}$$

where α is an unknown constant, the β's contains population-specific parameters. $b_i = (b_{i0}, b_{i1})$ contains subject-specific parameters (intercept and the effects of age) describing how the evolution of the ith individual deviates from the average evolution in the population, and where the residual component ϵ_i=$(\epsilon_{i1}, \ldots, \epsilon_{in_i})'$ is a vector containing the common error components, with $\epsilon_i \sim N(0, \Sigma_i)$. In this paper, we assumed that $\Sigma_i = \sigma^2 I_{n_i}$, so that, conditional on the values of the random effects, a person's measurements of the Hb level are independent. However, additional correlation among the errors can be accommodated by allowing for a more general covariance structure (e.g., autoregressive) in the model. It is assumed b_i has a bivariate normal distribution with mean zero and a diagonal covariance matrix, so that ϵ_i and b_i are mutually independent.
To estimate the parameters in the mixed effects models we employed the method of restricted maximum likelihood (REML). We applied an empirical Bayes method (EB) to predict a person's random intercept and slope based on his/her all previous observations

(McCulloch and Neuhaus, 2001).
We used a likelihood ratio test to choose between the mixed model with random intercept and the mixed model with random intercept and slope. In this case, the likelihood ratio test statistic for testing a random slope in the model is a mixture of chi-squared distributions with 1 and 2 degrees of freedom (Verbeke and Molenberghs, 2000).
Note that the linear mixed effects model is based on quite different assumptions than the transition model. In principle, if one model is correct, the other model must be wrong. However, in practice we never know the truth and in fact it is possible that both models are wrong. Despite this, we can still check which of the two models performs better in predicting the Hb level.

Prediction performance

To avoid a too optimistic assessment of the model predictions by using the data twice, i.e., for model building and parameter estimation as well as model evaluation, we have randomly divided the data set ($n = 15,625$ donors) into two parts: a training data set consisting of all observations of 7,709 donors and a validation data set consisting of all observations of the remaining 7,916 donors (Kohavi, 1995). The models are estimated using the training data set, and the model predictions are evaluated using the validation data set. We used a dynamic prediction approach in the sense that to predict Hb level at a visit we used the observations of all previous visits, therefore for each visit we updated our prior information. Since no prior information is available for the first visit, the predicted values are based only on the sex and age of the donor and the season in which the visit takes place.
The ultimate purpose of our longitudinal model is to predict future Hb values, given previously measured Hb values of a blood donor. Two criteria for choosing a model are Akaike's information criterion (AIC) (Akaike, 1973) and the related Bayesian information criterion (BIC) (Schwarz et al., 1978). We report the values of AIC and BIC for the training data set. In addition, we have chosen to estimate the predictive accuracy using some simple and intuitively clear measures, i.e., *mean squared prediction error* (MSPE) as a function of the visit number. At the tth visit, the MSPE is computed as:

$$\text{MSPE}_t = \sum_{i=1}^{N_t} \frac{(\hat{y}_{it} - y_{it})^2}{N_t}, \tag{2.5}$$

where $\hat{y}_{it}$ and y_{it} are the predicted and observed values, respectively and N_t is the total number of subjects at occasion t. MSPE_t is a well-known measure to evaluate prediction. The MSPE values are calculated for the validation data set only.
We also computed the sensitivity and specificity of the predicted values for assessing the eligibility for donation in the validation data set. Specifically, we computed the proportion of individuals that are correctly predicted to be eligible for donation based on the clinical

cut-off value (i.e., an Hb level of at least 8.4 mmol/l and 7.8 mmol/l for men and women, respectively). However, one may also optimize the cut-off value for the predicted values to obtain a receiver-operating characteristic (ROC) curve. In this ROC curve, the state variable is a dichotomous variable indicating whether the Hb level is below the clinical cut-off value of 8.4 mmol/l for men or 7.8 mmol/l for women; the test variable is the predicted value $\hat{y}_{it}$. Varying the cut-off value for the predicted value will change the sensitivity to detect that a donor will be eligible; however the assessment of donors' eligibility is based on the clinical cut-off value, which is not changed in the ROC analysis. We calculated the area under the curve (AUC) to compare the models. The difference in the AUCs between the models was tested using a bootstrap technique (Robin et al., 2011; Hanley and McNeil, 1983) that takes into account the correlation between the areas that is induced by the paired nature of the data.

2.3 Results

Table 2.1 presents descriptive statistics of the training and validation data sets. Different models are applied on the Sanquin data. We start with a multiple linear regression model (*Model LR*) that includes age, season, and donation at previous visit (DPV) as covariates. This model ignores the correlation among the subsequent hemoglobin values and hence is not a candidate choice, however, it serves as a benchmark to evaluate the more realistic models. In addition to the multiple linear regression model, a mixed effects model (*Model LME*) and transition (autoregressive) models of different orders are fitted to the training data set. The transition models are denoted as *Models AR(1) to AR(5)*, where the number indicates the order of the transition model. The data for male donors supported only a mixed model with random intercept (p-value$= 0.19$), but the data for female donors supported a mixed model with random intercept and slope (p-value < 0.001). Tables 2.2

Table 2.1: Descriptive statistics of the training and validation data sets.

Data set	Sex	#Donor	#Deferral	#Cold Season	Age: Mean (SD)	Visit: Med (IQR)
Training data set	Male	3610	769 (4.58%)	10213 (50.05%)	34.57 (12.9)	5 (3)
	Female	4306	1596 (9.62%)	10387 (49.71%)	32.66 (12.8)	5 (1)
	Total	7916	2365 (7.08)%	20600 (49.88%)	33.53 (12.9)	5 (2)
Validation data set	Male	3449	688 (4.27%)	9781 (49.95%)	34.28 (12.6)	5 (3)
	Female	4260	1729 (10.41%)	10341 (49.54%)	32.77 (12.8)	5 (2)
	Total	7709	2417 (7.38%)	20122 (49.74%)	33.45 (12.7)	5 (2)

Note: SD= Standard deviation, IQR= Interquartile range

and 2.3 display the results of the fitted models on the training data set for male and female donors, respectively. These tables indicate that all transition effects (regression coefficients

of past Hb values) are significant, although the effect of previous Hb level decreases with the lag. The effect of age is negative for male donors and positive for female donors, these results are consistent with previous studies e.g., see Baart et al. (2011, 2012). During warm seasons Hb level is lower on average than during cold seasons; this result is also supported by previous studies e.g., see Hoekstra et al. (2007); Kristal-Boneh et al. (1993). Furthermore, our models show that having had a donation in the previous visit has a negative effect on the current Hb level.

Table 2.2: Parameter estimates (standard errors) of the models estimated using the training data set for male donors.

Parameter	Model LR	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	Model LME
intercept	9.6448	9.6309	9.6441	9.6560	9.6617	9.6633	9.6719
	(0.0142)	(0.0206)	(0.0231)	(0.0243)	(0.0246)	(0.0247)	(0.0243)
Age	-0.0045	-0.0043	-0.0044	-0.0045	-0.0047	-0.0047	-0.0049
	(0.0003)	(0.0005)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0006)
Season(Warm)	-0.0627	-0.0615	-0.0681	-0.0699	-0.0693	-0.0694	-0.0698
	(0.0089)	(0.0074)	(0.0066)	(0.0066)	(0.0067)	(0.0067)	(0.0067)
DPV	-0.0610	-0.0469	-0.0350	-0.0385	-0.0440	-0.0474	-0.0636
(Donation)	(0.0092)	(0.0089)	(0.0079)	(0.0074)	(0.0072)	(0.0072)	(0.0068)
γ_1	—	0.5158	0.3685	0.3053	0.2746	0.2630	—
	—	(0.0061)	(0.0068)	(0.0076)	(0.0082)	(0.0087)	—
γ_2	—	—	0.2888	0.2080	0.1766	0.1621	—
	—	—	(0.0078)	(0.0087)	(0.0084)	(0.0091)	—
γ_3	—	—	—	0.2207	0.1730	0.1581	—
	—	—	—	(0.0095)	(0.0104)	(0.0109)	—
γ_4	—	—	—	—	0.1488	0.1257	—
	—	—	—	—	(0.0123)	(0.0129)	—
γ_5	—	—	—	—	—	0.0829	—
	—	—	—	—	—	(0.0167)	—

The AIC and BIC values for different models based on the training data set and the MSPE values based on the validation data set are shown in Table 2.4 for men and women. The results in Table 2.4 show that, for both sexes, AIC and BIC prefer a 5th order transition model over transition models that use fewer lagged observations. However, if we include all models, the smallest AIC and BIC value for the data of female donors are obtained with the mixed model with random intercept and random slope. The assessment of predictive accuracy based on MSPE confirms that all transition models and the mixed effects (LME) model provide much better predictions than the multiple linear regression model. In addition, the results indicate that the transition model usually provides a better prediction than the mixed effects model, especially at high visit numbers, see Figure 2.2. Based on the fitted models, we calculated the predicted Hb levels for donors from the validation data set and predicted the eligibility (Hb>8.4 for men and Hb>7.8 for women) of a donor at a

Table 2.3: Parameter estimates (standard errors) of the models estimated using the training data set for female donors.

Parameter	Model LR	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	Model LME
intercept	8.2737	8.2394	8.2555	8.2678	8.2698	8.2702	8.2832
	(0.0123)	(0.0164)	(0.0180)	(0.0186)	(0.0187)	(0.0187)	(0.0181)
Age	0.0042	0.0044	0.0042	0.0040	0.0040	0.0040	0.0037
	(0.0003)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Season(Warm)	-0.0347	-0.0405	-0.0415	-0.0413	-0.0415	-0.0415	-0.0411
	(0.0078)	(0.0062)	(0.0060)	(0.0062)	(0.0061)	(0.0061)	(0.0062)
DPV	-0.1106	-0.1411	-0.1273	-0.1307	-0.1335	-0.1346	-0.1387
(Donation)	(0.0079)	(0.0075)	(0.0067)	(0.0064)	(0.0063)	(0.0063)	(0.0060)
γ_1	—	0.4669	0.3457	0.3012	0.2878	0.2830	—
	—	(0.0062)	(0.0067)	(0.0074)	(0.0080)	(0.0084)	—
γ_2	—	—	0.2573	0.1963	0.1793	0.1693	—
	—	—	(0.0080)	(0.0088)	(0.0089)	(0.0099)	—
γ_3	—	—	—	0.1742	0.1486	0.1360	—
	—	—	—	(0.0100)	(0.0112)	(0.0121)	—
γ_4	—	—	—	—	0.0831	0.0623	—
	—	—	—	—	(0.0157)	(0.0182)	—
γ_5	—	—	—	—	—	0.0681	—
	—	—	—	—	—	(0.0264)	—

Table 2.4: AIC and BIC values for different models for both sexes based on the training data set.

	Male donors			Female donors		
Model	AIC	BIC	MSPE	AIC	BIC	MSPE
Linear Regression	37087.8	37127.2	4.14	35968.9	36008.6	2.29
Mixed Effects	30524.3	30571.6	2.90	30058.0	30113.6	1.75
AR(1)	32051.0	32098.3	3.07	31559.1	31606.7	1.81
AR(2)	30936.4	30991.6	2.85	30664.7	30720.3	1.73
AR(3)	30471.9	30535.0	2.78	30375.1	30438.7	1.71
AR(4)	30342.5	30413.4	2.78	30341.7	30413.2	1.72
AR(5)	30321.4	30400.2	2.79	30325.1	30404.5	1.72

Note: Lower values of AIC, BIC, and MSEP indicate better model fit.

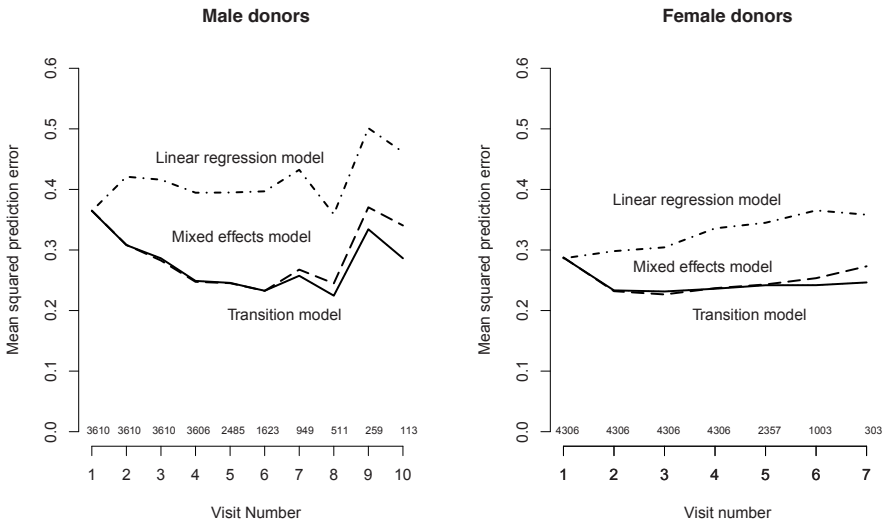


Figure 2.2: Mean squared prediction error of the linear regression model, the linear mixed effects model, and the 5th order transition model, as a function of the visit number. The included numbers of individuals are displayed above the horizontal axis.

particular visit. Figure 2.3 displays the ROC curves for the 5th order transition model and the mixed effects model for male donors; since the results for female donors are similar, the ROC curves for female donors are not shown. All observations in the validation data set (n = 7,916 donors) were used to compute these ROC curves. The AUCs for the transition model and mixed effects model are 0.83 and 0.81 for men, respectively; for women we obtained AUC values of 0.73 and 0.72, respectively. The difference in AUCs between the two models is statistically significant (p-value < 0.001), namely the transition model has a larger AUC than the mixed effects model and thus offers a better trade-off between sensitivity and specificity.

2.4 Discussion

In this article, we presented transition models with different numbers of autoregressive terms and mixed effects models (a mixed effects model with random intercept for male donors and a mixed effects model with random intercept and random slope based on age for female donors), as plausible models to account for the dependence among subsequent Hb levels within a donor and as models to predict the future hemoglobin level.
Based on the results for the validation data set, we showed that the transition model and the mixed effects model have almost the same predictive accuracy at the first few visits

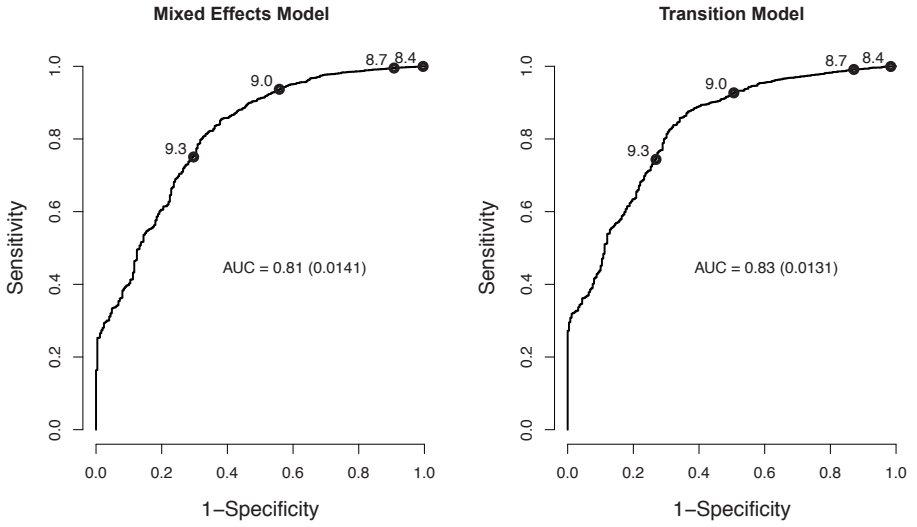


Figure 2.3: ROC curves of the prediction of eligibility for donation in male donors, for two different models. The standard errors of the AUCs are shown in parentheses. Different cut-off points for the predicted value are displayed on the curves.

of a donor; however, for longer time series the transition model offers somewhat better predictions.

To give an idea of the predictive performance, we have computed the ROC curve. Our results confirm that the transition model shows a small but significant improvement in the AUC compared to the mixed effects model.

Both the transition and the mixed effects models use the data of a persons previous observations for making predictions. In the transition model only the last q observations are used for prediction the current response. However, in the mixed effects model, the empirical Bayes method for estimating a persons random effects uses all previous observations. Therefore, the mixed effects model requires more historical information than the transition model. Since the transition model is convenient in practice and needs less historical information compared to the mixed effects model, blood banks may use this model to predict the future hemoglobin level of a candidate and to determine which candidates should not be invited for the next donation.

Our approach of using transition or autoregressive models is quite novel in biomedical research, however in other fields such as econometrics, autoregressive modeling is a very well-known technique for tackling correlated financial phenomena and time series problems (Earnest et al., 2005).

We do not claim that our final model is optimal; further research is needed to arrive at a

better prediction model. First, the data set used in this paper is unbalanced in the sense that the time intervals between visits vary considerably, though this was not taken into account here. Nevertheless, we believe that our paper shows the capabilities of longitudinal models in prediction and our findings may help reduce the number of deferred candidate in the blood banks. Second, there are more factors that are possibly associated with Hb level than those which we have investigated in this study, such as physical activity (Beard and Tobin, 2000), race (Johnson-Spear and Yip, 1994), nutrition (Brussaard et al., 1997) and smoking status (Skjelbakken et al., 2006; Kristal-Boneh et al., 1997).

Finally, the ultimate purpose of the prediction exercise is not the prediction of the future Hb value, but rather to determine the best time for the donor to return for donation. Hence, prediction models for Hb levels after blood donation should focus on the optimal timing of future donations, instead of on predicting future Hb levels. We are currently investigating such models.

2.5 Supplementary

R codes to implement the transition model with different order are presented in below. Note that the " > " symbol in the line shown is the prompt in R and not part of what the user types.

```
### Required packages
> require(dlm)
> require(stats4)
### Required packages

> Myfunction<-function(par1, par2, par3, par4, par5,
> par6, par7, par8, par9, Data){
### Par1 to par5 are the autoregressive parameters.
### Par6, Par7, and par8 are beta_0, beta_1, and beta_2.
### Par9 is the residuals variance estimate.
> pars<-c(par1,par2,par3,par4,par5)
> Likl<-0

> for (i in 1:length(unique(Data$Id)))
  > {
  > Data.P<-Data[Data$Id%in%i,]
  > DataPerson<-Data.P$Hb-par6-par7*Data.P$Season-par8*Data.P$Age
  > Likl<-Likl+dlmLL(DataPerson, mod=dlmModARMA(ar=pars,
  > sigma2=par9), debug=TRUE)
  > }

> if ((is.na(Likl)) || Likl>10^10) {
  > Res<-10^10
  > } else{ Res<-Likl }
> }

#### An example of AR(1)
> m1<- mle(Myfunction, start=list(par1=0.3, par6=8,
> par7=0, par8=0, par9=0.3),
> fixed=list(Data=Data, par2=0, par3=0, par4=0, par5=0),
> method="BFGS", control=list(trace=TRUE, REPORT=1,
> reltol=10^-10, maxit=10^3))
> summary(m1)
```

```
#### An example of AR(2)
> m2<- mle(Myfunction, start=list(par1=0.3, par2=0.1,
> par6=8, par7=0, par8=0, par9=0.3),
> fixed=list(Data=Data, par3=0, par4=0, par5=0), method="BFGS",
> control=list(trace=TRUE, REPORT=1, reltol=10^-10, maxit=10^3))
> summary(m2)
```

3 Prevalence and determinants of declining versus stable hemoglobin levels in whole blood donors

This chapter is published as: Nasserinejad K, van Rosmalen J, van den Hurk K, Baart M, Hoekstra T, Lesaffre E, and de Kort W. Prevalence and determinants of declining versus stable hemoglobin levels in whole blood donors. *Transfusion*, 2015 ;55(8):1955-63. doi: 10.1111/trf.13066

Abstract.

A too short recovery time after blood donation results in a gradual depletion of iron stores and a subsequent decline in hemoglobin (Hb) levels over time. This decline in Hb levels may depend on individual, unobserved characteristics of the donor. We used a data set of 5388 Dutch blood donors from the Donor InSight study. The statistical analysis is based on a Bayesian growth mixture model, which assumes that each donor belongs to one of several groups. Each group implies a different Hb trajectory, and donors with similar longitudinal trajectories belong to the same group.

Analyses were performed for male and female donors separately. For both sexes the model identified four groups of donors. Stable Hb trajectories were found among 14% of male donors and 15% of female donors; declining Hb trajectories were observed in the remaining groups of donors. The percentage of donor deferrals differed strongly between groups. The model can be used to predict to which group a donor belongs, and this prediction can be updated after each donation. This is of high practical importance because early identification of donors with declining Hb levels could help to tailor donation intervals and to prevent iron deficiency and donor deferrals.

3.1 Introduction

Whole blood donation poses a risk of iron deficiency to blood donors (Skikne et al., 1984). A whole blood donation implies a loss of erythrocytes and iron, resulting in a temporary decrease in hemoglobin (Hb) levels. In healthy donors with sufficient iron stores this may not be problematic. Iron balance is achieved by more efficient absorption of dietary iron in blood donors (Skikne et al., 1984). Repeated donations could however deplete iron stores, leading to iron depletion and ultimately anemia (Skikne et al., 1984; Brittenham, 2011).
Because healthy donors give blood voluntarily, iron depletion and subsequent anemia should be prevented as much as possible. In the Netherlands, several measures were taken to prevent donors from becoming anemic. Already in the 1940s, it was estimated that the body needs around 50 days to recover to pre-donation Hb levels (Fowler and Barer, 1942; Alstead, 1943). Therefore, guidelines impose a minimum interval of 56 days between donations, with a yearly maximum of 5 donations for men and 3 for women (Baart et al., 2012). Furthermore, the iron status of blood donors is assessed prior to donation. This is done by measuring whether Hb levels are $\geq$ 8.4 mmol/L (135 g/L) for men or $\geq$ 7.8 mmol/L (125 g/L) for women (Baart et al., 2012). Donors with Hb levels below these cut-offs, or 1.5 mmol/L (24 g/L) below the previous donation level, are temporarily deferred from donation. Deferrals can be demoralizing for donors and have a negative effect on donor return rates (Halperin et al., 1998; Custer et al., 2014; Zou et al., 2008). Hb deferrals thus decrease the cost-effectiveness of blood supply, because a) testing and deferring a donor is expensive, b) for every deferred donor another donor needs to be invited to reach collection targets, and c) lapsing donors need to be replaced (Hillgrove et al., 2011).
Both recent and historical data suggest that individual donors may differ in their recovery from blood donation, indicating that a donation interval of 56 days may not be desirable for each individual donor (Fowler and Barer, 1942; Cable et al., 2011). This may result in gradually declining Hb levels over time, which are currently not detected until the donor meets any of the deferral criteria. Distinguishing between donors with different Hb trajectories after repeated blood donations may help to select donors and tailor their donation intervals to prevent anemia. To our knowledge, no data on individual Hb trajectories exist in literature. Therefore, an objective of this study is to investigate whether different Hb trajectories can be distinguished in whole blood donors. Also, we determine whether the type of trajectory is associated with the probability of deferral due to low Hb. Finally, we aim to predict the type of Hb trajectory of a newly registered blood donor.

3.2 Material and methods

Study population

For blood collection, all measured data are entered into the blood bank computer system (e)PROGESA (MAK-SYSTEM International Group, France). Prior to every donation, Hb and other parameters are required to check whether the prospective donor is eligible to donate. In the Netherlands, a newly registered donor must not donate blood at the first visit, which consists of a health check only. At every subsequent visit, donors who pass all eligibility checks can donate 500 ml of whole blood.
For the present study, data are extracted from (e)PROGESA. The Donor InSight data set is used for data analysis. The Donor InSight data set is a self-administered questionnaire study aimed at gaining insight into characteristics and motivation of the Dutch donor population (Atsma et al., 2011). Our analysis comprises whole blood donors who were registered as a new donor in the period 1 January 2005 to 31 December 2012. To be included in the study, they should have at least one visit after the first donation. A total of 5388 donors (1902 male and 3486 female donors) fulfilled these criteria. The Donor InSight study was approved by the Medical Ethical Committee Arnhem-Nijmegen in the Netherlands, and all participants gave their written informed consent.

Data

Capillary Hb is measured as part of routine donor health assessments. All measured Hb levels from the first Hb measurement up to and including the last donation in 2012 were used for the analysis, with the following exceptions: 1) If the donor was deferred for low Hb at one or more of the visits, then Hb levels up to and including the Hb level measured at the first visit that resulted in a deferral were used. 2) Whole blood donors may change from giving whole blood to donating plasma, or vice versa. In that case, only Hb levels measured at whole blood donations before the first plasma donation were used in the analyses. 3) Donors may quit donating, and register as a new donor again after some period. For these donors only the Hb measurements from the first donor career were used.

Statistical methods

To capture the longitudinal trajectories of Hb levels, and the variation in these trajectories between donors, we implement a growth mixture model (Wang and Bodner, 2007; Nagin and Odgers, 2010). This model assumes that each donor belongs to one of several subgroups (known as latent classes in statistical terminology). Using this method, it is inferred from the data to which class each donor belongs.
The model assigns each donor to one of several groups, in such a way that donors with

similar Hb trajectories are in the same group, and that the groups are most different from each other in terms of the Hb trajectory. The classes typically do not capture the entire variation in the Hb trajectories. To capture the remaining heterogeneity between donors in the same class, the Hb trajectory follows a linear mixed model. The outcome in the linear mixed model is the Hb level. The predictors are age (Yip et al., 1984) at the first visit, season (Hoekstra et al., 2007; Kristal-Boneh et al., 1993) of the visit (a binary covariate, i.e., the cold season includes fall and winter and the warm season includes spring and summer), a linear and quadratic effect of the time since the previous donation (Rikhtehgaran et al., 2012), and the number of donations in the last two years (Baart et al., 2012). Male and female donors have different Hb profiles, therefore the data for men and women are analyzed using separate statistical models.
Random intercepts and random effects of the number of donations in the last two years are used to capture the heterogeneity between different donors in the same class. In one of the latent classes, we impose that there is no effect of the number of donations during the last two years, so that this class consists of donors with a stable Hb trajectory. This constraint enables us to estimate the percentage of donors with a stable trajectory. Finally, we allow latent class membership to depend on age and the Hb level at the screening visit.
The models are estimated using a Bayesian approach with Markov chain Monte Carlo (MCMC) sampling. The number of latent classes in the growth mixture model is based on the Bayesian information criterion (BIC) (Hawkins et al., 2001; Schwarz et al., 1978). Kaplan-Meier analysis is used to compare the classes with respect to the number of donations until the first deferral due to low Hb.
Further technical details regarding the linear mixed model, the growth mixture model and the model for latent class membership are given in the Appendix. Statistical analyses are performed using Jags (3.4.0) (Plummer, 2011) and R (3.1.0) (R Development Core Team, 2010).

Predicting Hb trajectory

A useful feature of the growth mixture model is that it can be used to predict to which latent class a donor belongs, and these predictions can be made in a dynamic way. At the screening visit, the prediction is based only on the observed Hb level and age. At each subsequent visit, the prediction is updated by taking into account the newly observed Hb level for that person. The donors are assigned at each visit to the class with the highest probability, i.e., the class that best fits the observed Hb levels given the donor's age and sex. See the Appendix for more explanation of this procedure.

Table 3.1: Descriptive statistics of the Donor InSight data set based on 1902 male and 3486 female donors.

Variable	Male donors	Female donors
Donors deferred at least once due to low Hb	18.4%	32.3%
Donations in cold season (fall and winter)	49.9%	49.5%
Age at screening visit (years)	34.8 (24.1, 45.8)	29.4 (21.6, 42.2)
Number of donations	9 (4, 16)	5 (2, 9)
Hb level at screening visit (mmol/l)	9.4 (9.0, 9.9)	8.4 (8.0, 8.8)
Inter-donation interval (days)	90 (74, 126)	138 (120, 188)

Dichotomous variables are presented as percentages, and the other variables are presented using medians and interquartile ranges.

3.3 Results

Table 3.1 presents descriptive statistics of the Donor InSight data set. Figures 3.1a and 3.1b show the Hb level profiles of male and female donors, respectively. These graphs show the heterogeneity of Hb level trajectories. We emphasized the profiles of two donors with different trajectories (donor I: a donor with a fast Hb recovery or stable Hb trend after several successive donations, donor II: a donor with a slow Hb recovery or declining trajectory in Hb levels after several successive donations). Based on the BIC, at least 4 classes are needed for both sexes. A model with 5 classes has slightly better BIC values, but the additional class has a very small size (1% for males and 5% for females). Therefore, we selected a model with 4 classes for both sexes. Figures 3.2 and 3.3 show the Hb trajectories for the model with 4 latent classes, for men and women, respectively. Table 3.2 presents the main characteristics of the different latent classes, and Table 3.3 presents the parameter estimates of the growth mixture models with 4 latent classes. The latent classes can be interpreted using the results of these two tables. To allow for an easy comparison of classes between sexes, we sorted the classes based on the predicted Hb level at the first visit. For both sexes, class I represents donors with a stable Hb level. 13.5% of the male donors are in class I. Since these donors have a relatively low average Hb level at the screening visit (8.7 mmol/l), there are 39.3% deferrals in this group. Class II shows the slowest average decline of Hb level, but also a lower initial average Hb level, resulting in 21.6% deferrals. In class III the male donors (37.7%) show a moderate average decline with successive donations. However, because the initial average Hb level of this group is relatively high (9.7 mmol/l), only 10.6% of the donors are deferred. Finally, class IV (10.9%) has the lowest percentage of deferrals (8.2%), due to the very high initial average Hb level (10.4 mmol/l). On the other hand, 15.4% of the female donors are in the stable class (class I). These donors have a low average Hb level at the screening visit (7.8 mmol/l), resulting 66.6% deferrals. Class II exhibits the slowest average decline of Hb level, but also a lower initial average Hb

Table 3.2: Descriptive statistics of the Donor InSight data set by sex and latent class of the growth mixture model.

	Male donors				Female donors			
Variable	Class I	Class II	Class III	Class IV	Class I	Class II	Class III	Class IV
Size of the class	13.5%	37.9%	37.7%	10.9%	15.4%	43.5%	28.4%	12.7%
Donors deferred at least once due to low Hb	39.3 %	21.6 %	10.6%	8.2 %	66.6%	35.4%	16.6%	15.8%
Inter-donation interval (days)	91(76, 140)	89(74, 124)	90(74, 126)	87(74, 119)	137(119, 178)	139(120, 189)	137(120, 185)	138(120, 194)
Number of donations	7(2, 13)	9(5, 16)	10(5, 17)	11(5, 18)	2(1, 5)	4(2, 9)	6(3, 11)	6(3, 9)
Hb level at screening visit (mmol/l)	8.7(8.5, 8.9)	9.2(9.0, 9.4)	9.7(9.4, 10.0)	10.4(10.1, 10.7)	7.8(7.6, 8.0)	8.2(8.0, 8.5)	8.7(8.5, 9.0)	9.3(9.1, 9.6)
Age at screening visit (year)	39(29, 51)	36(26, 48)	32(24, 43)	33(24, 44)	43.5(33, 52)	29(22, 41)	27(21, 40)	23(20, 31)

Dichotomous variables are presented as percentages, and the other variables are presented using medians and interquartile ranges.

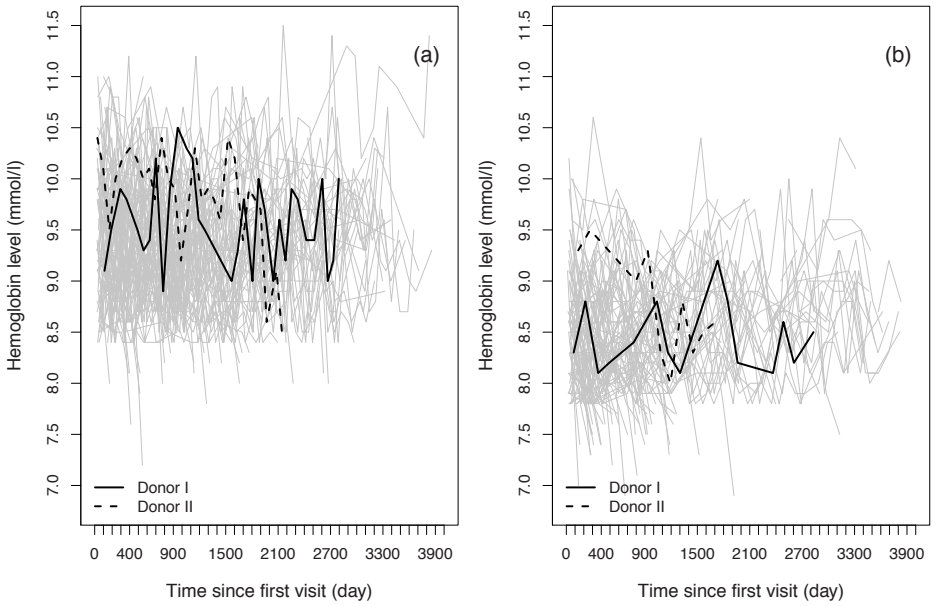


Figure 3.1: Hb level profiles for male and female donors are presented in panels (a) and (b), respectively. The solid and dashed lines indicate different types of Hb trajectories, i.e., donor I represents a stable Hb trajectory and donor II represents an unstable trajectory.

level, 35.4% deferrals. Class III (28.4%) shows a moderate average decline with successive donations, but because the initial average Hb level is relatively high (8.7 mmol/l), there are only 16.6% deferrals. Finally, class IV (12.7%) has the lowest percentage of rejected donors (15.8%), due to the very high initial average Hb level (9.3 mmol/l). From Table 3.3 we conclude that, for male donors, age at baseline is less strongly associated with Hb level than for female donors where age at baseline has a positive effect on Hb level. The average Hb level is higher in cold seasons than in warm seasons for both sexes. Furthermore, time since previous donation (TSPD) has a non-linear, quadratic, effect on Hb level in both sexes, with the fastest recovery of Hb occurring shortly after a donation.

To illustrate the different deferral patterns in the four latent classes, Kaplan-Meier (K-M) curves exhibiting the proportion of deferral in each of the latent classes for each sex separately are shown in Figure 3.4. The log-rank test indicates significant difference between these curves (p-value < 0.001, for both sexes), although similar K-M curves are seen for classes III and IV. Figure 3.5 shows the prediction of the latent class for a male donor with Hb level of 8.9 mmol/l and age of 29 years at the screening visit. Using the information available at the screening visit, this donor would be predicted to belong to class II (with 54% probability). However, using Hb levels measured at the first few visits, it is clear that this donor more likely belongs to class I (i.e., the class with a stable trajectory). For

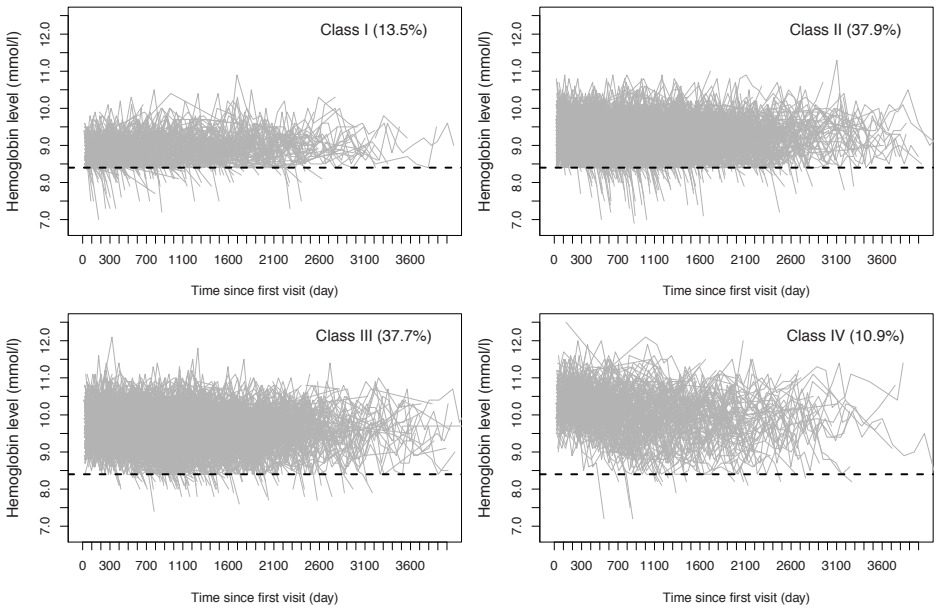


Figure 3.2: Hb level trajectories of male donors split up into 4 latent classes. Class sizes are reported in parentheses. The Hb threshold for the eligibility of donation is shown by a dashed line.

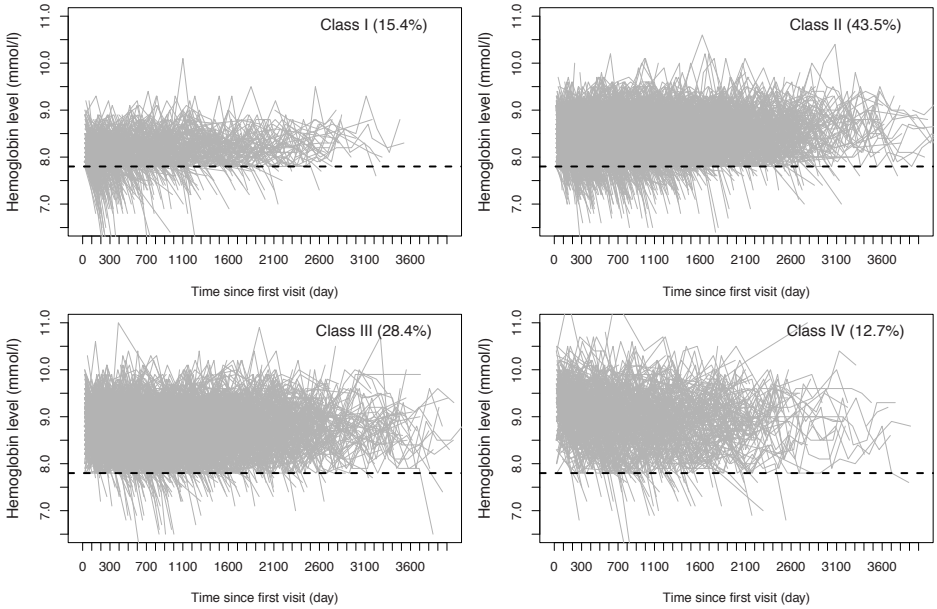


Figure 3.3: Hb level trajectories of female donors split up into 4 latent classes. Class sizes are reported in parentheses. The Hb threshold for the eligibility of donation is shown by a dashed line.

Table 3.3: Parameter estimates for male and female donors in the Donor InSight study based on the growth mixture model with 4 latent classes.

	Male donors			Female donors		
Parameter	Estimate	95% CI		Estimate	95% CI	
Intercept_{I}	8.88	8.81	8.89	7.93	7.85	8.00
Intercept_{II}	9.34	9.27	9.44	8.30	8.24	8.36
Intercept_{III}	9.84	9.78	9.94	8.77	8.68	8.84
Intercept_{IV}	10.38	10.23	10.59	9.13	9.02	9.25
NODY2_{II}	-0.05	-0.06	-0.04	-0.02	-0.04	-0.01
NODY2_{III}	-0.06	-0.08	-0.05	-0.06	-0.10	-0.03
NODY2_{IV}	-0.09	-0.12	-0.07	-0.14	-0.17	-0.10
Age_0(year)	1.7×10^{-4}	-2.8×10^{-3}	3.1×10^{-3}	9.6×10^{-3}	6.4×10^{-3}	1.3×10^{-2}
Warm season	-7.7×10^{-2}	-8.9×10^{-2}	-6.5×10^{-2}	-5.2×10^{-2}	-6.2×10^{-2}	-4.2×10^{-2}
TSPD (month)	2.5×10^{-2}	1.9×10^{-2}	3.0×10^{-2}	1.9×10^{-2}	1.4×10^{-2}	2.3×10^{-2}
TSPD^2(month)	-1.0×10^{-4}	-1.3×10^{-4}	-7.3×10^{-5}	-4.5×10^{-5}	-6.2×10^{-5}	-2.8×10^{-5}
BFD	-8.7×10^{-4}	-3.2×10^{-2}	3.1×10^{-2}	8.3×10^{-2}	5.4×10^{-2}	11.1×10^{-2}

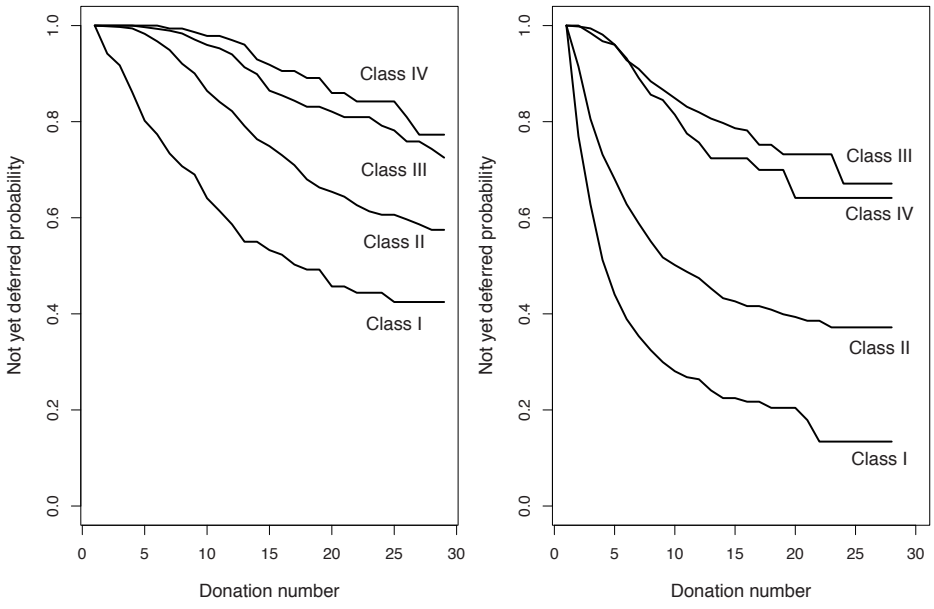


Figure 3.4: Kaplan-Meier (K-M) curves exhibiting the proportion of deferral in each of the latent classes for male donors (left panel) and female donors (right panel) separately.

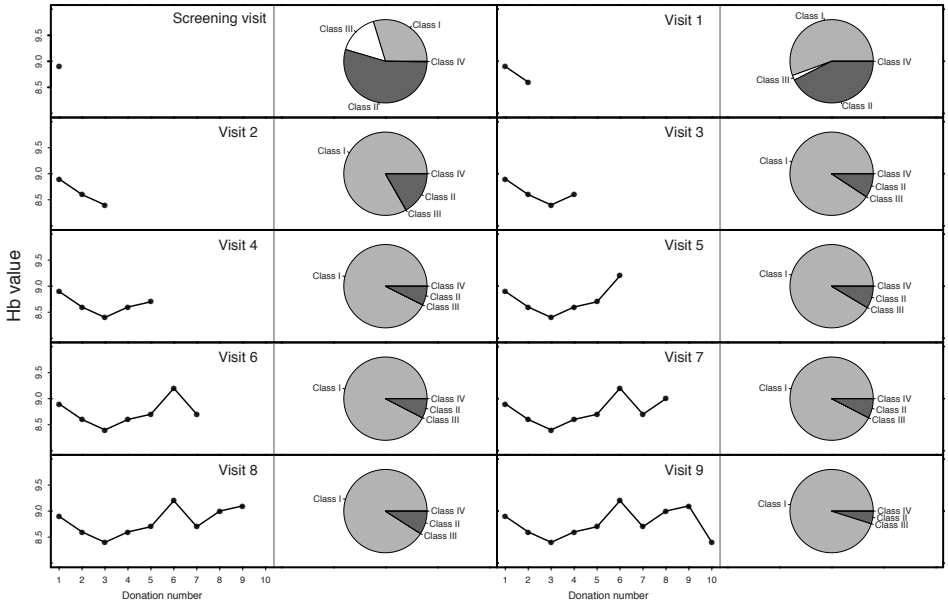


Figure 3.5: Class-membership probabilities at the first nine visits of a male donor with a Hb level of 8.9 and age of 29 years at the screening visit. The first and third columns show the longitudinal trajectory of observed Hb levels. The second and fourth columns give the corresponding class-membership probabilities using a pie chart.

example, after the fifth visit the donor has a probability of 91% to belong to class I. In the primary analysis, we used the number of donations during the last two years to model the decline in Hb level associated with successive blood donations. In a sensitivity analysis, we fitted the same model with the total number of previous donations. This variable did not improve the fit of the model. The same is true for the binary variable indicating whether the person ever had a donation prior to the previous two years.

3.4 Discussion

The results of this study have shown that, for both male and female whole blood donors, Hb trajectories vary among donors. Our growth mixture model identified four types of Hb level trajectories. A minority of male and female donors (13.5% and 15.4%) are in class I, which has a stable Hb trajectory. The donors in the other classes have a declining Hb trajectory, so that their Hb level shows a significant decline after successive blood donations. Although their Hb trajectories are stable, the donors in class I have a relatively low initial Hb level. With increasing class number, the initial Hb level increases, as well as the speed of the decline. Donors in classes with a relatively low initial Hb level (classes I and

II) are deferred more frequently than donors in the other classes. A donors latent class can be predicted and updated after each donation by taking into account the Hb values measured at subsequent visits.

To the best of our knowledge, the present study is the first to identify different long-term trajectories of whole-blood donors using longitudinal data from blood banks. Several studies have shown that the risk of Hb deferral is higher for donors with lower Hb levels (Baart et al., 2011, 2012; Pasricha et al., 2011). Our data add that many blood donors show declining Hb trajectories. In previous studies, prediction models for Hb values in whole blood donors have been developed. These prediction models were proposed as mixed-effects models, transition models, or a combination of these two approaches for predicting Hb values in blood donors (Rikhtehgaran et al., 2012; Nasserinejad et al., 2013). The current findings suggest that describing the total donor population using a single trajectory oversimplifies the complex growth patterns of this population. Instead, a growth mixture modeling approach, which accounts for different subgroups of donors, seems to be an appropriate method for capturing differences in Hb trajectories between donors.

Our results showed that higher age at baseline was associated with higher Hb levels in female donors. This is consistent with earlier results and can be explained by the effect of menopause: women stop losing iron with menstruation (Baart et al., 2012). Furthermore, we showed that on average Hb level is higher in cold seasons than in warm seasons for both sexes. These findings are consistent with earlier results as well (Hoekstra et al., 2007; Kristal-Boneh et al., 1993).

Some donors appear to have high initial Hb levels and others do not, and some show faster declines in Hb than others. This may be due to differences in lifestyle, iron status, iron metabolism, and/or erythropoiesis (Baart et al., 2012; Soranzo et al., 2009; van der Harst et al., 2012; Cvejic et al., 2013). Including more of this information in the models might improve the precision of the prediction of latent classes at the first few visits. For this reason we aim to include other relevant predictors, including lifestyle and genetic factors, in future research.

Individual donors belonging to different classes should potentially be approached differently. For donors with a low but stable Hb trajectory (class I), delaying the next invitation may not help to decrease the probability of deferral. Donors with a normal initial Hb level (class II) become at risk for deferral only with a high donation frequency, because the estimated Hb decline per donation is fairly small in this group (at most 0.09 mmol/l per donation for men and 0.14 mmol/l per donation for women). Thus for this group, the advice could be to increase donation intervals. Donors with high initial Hb levels (classes III and IV) do not have a very high risk of Hb deferral. Changing their donation intervals may therefore not be very effective in preventing Hb deferral. Apart from Hb deferral, donors in different classes may also differ in what is healthy for them. The fact that fast Hb declines do not necessarily lead to Hb levels below the cut-off for donation, does not mean that it does no harm. Blood donation causes a loss of iron and blood cells,

which can lead to depleted iron stores and lowering of Hb levels (Amrein et al., 2012). Potential symptoms of iron deficiency include fatigue, decreased physical endurance and work capacity, and impairment in attention, concentration and other cognitive functions (Schiepers et al., 2010; Dallman, 1986; Newman et al., 2006; Popovsky, 2012). Restless legs syndrome, a neurological disorder with irresistible need to move the legs, and pica, a disorder in which a person is craving and consuming nonnutritive substances have also been repeatedly linked to blood donation-related iron deficiency (Ulfberg and Nyström, 2004; Birgegård et al., 2010; Bryant et al., 2013; Spencer et al., 2013). Future research should indicate whether adverse health effects of donation are different for donors with stable or declining Hb levels.

A major strength of the study is the large amount of routinely measured data, including many repeated measurements per donor. This allowed for detailed insights into Hb trajectories from the initial level to the end of follow up, including the relationships with donor deferral, age, and sex. A limitation of the study is that Hb is measured by photometry in capillary blood instead of more reliable hematology analyses in venous blood (Ziemann et al., 2011). Although this probably increased measurement error for single measurements, the large amount of repeated measurements likely smoothened this error. Furthermore, our results may not be generalizable to different ethnic populations or to blood banks where policies regarding Hb measurement and deferral are different. The Dutch donor population includes relatively low numbers of people from ethnic minority groups (Atsma et al., 2011), but this is very common in donor populations (Murphy et al., 2009). Policies regarding Hb deferral mainly differ in what exactly is measured, capillary or venous Hb or copper sulfate testing, and in the timing, before or after the donation. Nonetheless, cut-off values are quite similar throughout the world and the methods used in the Netherlands are very common (Ziemann et al., 2011).

In conclusion, we found subgroups of donors with stable and declining Hb trajectories. These subgroups were associated with the probability of Hb deferral and can be predicted based on initial Hb levels and age. These findings are of high importance for identification of donors who could benefit from tailored donation intervals to prevent iron deficiency and donor deferrals. Future research replicating our findings and investigating health effects of declining Hb levels will help to further unravel clinical implications.

Appendix

Linear mixed-effects model

Linear mixed-effects models are an important class of statistical models for analyzing longitudinal data, such as repeated measurements of a donor's Hb. The random effects in these models capture the heterogeneity among individuals. The mixed-effects models for the data of the Donor InSight study can be expressed as:

$$\text{Hb}_{it} = \beta_0 + b_{i0} + \beta_1 \text{Age}_{i0} + \beta_2 \text{Season}_{it} + \beta_3 \text{TSPD}_{it}$$
$$+\beta_4 \text{TSPD}_{it}^2 + \beta_5 \text{BFD}_{it} + (\beta_6 + b_{i1})\text{NODY2}_{it} + \epsilon_{it},$$

where Hb_{it} is the Hb level at tth observation of the ith individual. TSPD_{it} denotes the time since previous donation at tth observation of the ith individual and TSPD_{it}^2 is the square of TSPD_{it}. BFD_{it} indicates whether the visit t of donor i precedes the first donation (Yes=1 and No=0). NODY2_{it} stands for the number of donations in the last two years up to visit t of donor i. β_0, β_1, β_2, β_3, β_3, β_5, and β_6 are unknown regression coefficients of intercept, Age (i.e., age at first visit/donation), Season (i.e., the season in which the donation took place), TSPD, TSPD^2, BFD, and NODY2, respectively. The b_{i0} and b_{i1} are random intercepts and random slopes, respectively, in the model. We assumed that the random intercepts and slopes are bivariate normally distributed with mean zero and different variances. The residuals ϵ_{it} are assumed to be normally distributed and mutually independent with mean zero and constant variance, and independent of the random effects.

Growth mixture models

Mixture modeling refers to modeling with categorical latent variables to represent different classes in a population (Muthén and Shedden, 1999; McLachlan and Peel, 2004). In these models, class membership is not known in advance but is inferred from the data. In growth mixture models (Muthén and Shedden, 1999; Wang and Bodner, 2007) (also known as latent class mixed models), the principle of mixture modeling is applied in the context of a linear mixed model, so that the heterogeneity between subjects is captured by different classes and by random effects in the linear mixed models (Wang and Bodner, 2007). The density function of $\mathbf{y}$ in a Gaussian growth mixture model can be expressed as:

$$f(\mathbf{y}|\boldsymbol{\lambda}, \boldsymbol{\theta}) = \sum_{k=1}^{K} \lambda_k f_k(\mathbf{y}|\theta_k),$$

where $f_k(y|\theta_k)$ (k=1, ..., K) are density functions that describe the trajectory for each class. The λ_ks represent the proportion of subjects in each class, and must satisfy $0 \leq \lambda_k \leq 1$ and $\sum_{k=1}^{K} \lambda_k = 1$ (McLachlan and Peel, 2004). The vector θ_k represents the parameters that are associated with the trajectory of class k.

Growth mixture model for Donor InSight data set

The growth mixture model for the trajectory of Hb levels of blood donors who belong to latent class k can be expressed as:

$$\text{Hb}_{it|k} = \theta_{k0} + b_{ik0} + \beta_1 \text{Age}_{i0} + \beta_2 \text{Season}_{it} + \beta_3 \text{TSPD}_{it}$$
$$+\beta_4 \text{TSPD}^2_{it} + \beta_5 \text{BFD}_{it} + (\theta_{k1} + b_{ik1})\text{NODY2}_{it} + \epsilon_{it|k},$$

where $Hb_{it|k}$ is the predicted Hb level at the tth observation of the ith individual, given that this individual is in latent class k. θ_{10}, $\theta_{20}, \ldots, \theta_{K0}$ are the unknown intercepts of K latent classes. Likewise, $\theta_{21}, \theta_{31}, \ldots, \theta_{K1}$ are unknown coefficients of NODY2 in the latent classes. To restrict the Hb trajectory to be stable in the first class, the parameter θ_{11} and the variance of the random effect b_{i11} are set to zero. We assumed that within each latent class the random effects (i.e., b_{ik0} and b_{ik1}) are multivariate normally distributed with mean zero and a class-specific variance-covariance structure. The residuals ϵ_{it} are assumed to be normally distributed, and independent of the random effects.
The probability λ_{ik} that individual i ($i = 1, \ldots, N$) belongs to latent class k ($k = 1, \ldots, K$) is related to the initial Hb level and the age at the first visit using a multinomial logistic regression specification. This probability is calculated as

$$\lambda_{ik} = P(c_i = k|\text{Hb}_{0i}, \text{Age}_{0i}) = \frac{\exp(\gamma_{k0} + \gamma_{k1}\text{Age}_{0i} + \gamma_{k2}\text{Hb}_{0i})}{\sum_{k=1}^{K} \exp(\gamma_{k0} + \gamma_{k1}\text{Age}_{0i} + \gamma_{k2}\text{Hb}_{0i})},$$

where λ_{ik} is the probability that the ith individual belongs to class k given the baseline covariates. The first class is used as a reference class, and therefore the $\gamma_k = (\gamma_{k0}, \gamma_{k1}, \gamma_{k2})$ parameters for the first class are constrained to be 0.
The optimal number of latent classes is determined using the Bayesian information criterion (BIC). The BIC was computed based on the marginal mean posterior of parameters (Hawkins et al., 2001; Schwarz et al., 1978; Fraley and Raftery, 2002).

Model diagnostics

To assess model fit, we used a posterior predictive check (PPC) by computing a Bayesian P-value, i.e., the probability that replicated data from the model could be more extreme than the observed data for an omnibus χ^2 discrepancy measure (Gelman et al., 1996) to test both the distributional and latent class number assumption of the model.
To check whether the number of MCMC iterations is sufficient to obtain accurate estimates, the sampling was continued until the Monte Carlo errors were less than 5% of the posterior standard deviation of each parameter (Lesaffre and Lawson, 2012). The first 5,000 iterations (i.e., burn-in iterations) of each chain were discarded. The posterior means and credible intervals (CI) were calculated using the remaining iterations of each chain, using no thinning factor. We checked the convergence by monitoring trace plots and the Geweke diagnostic (Geweke et al., 1991). Finally, to check if donors were assigned

Table A1: Average class membership probabilities by latent class based on the Donor InSight data.

	Mean of Posterior Probabilities							
	Male donors				Female donors			
Latent Class	Class I	Class II	Class III	Class IV	Class I	Class II	Class III	Class IV
Class I	0.795	0.204	0.001	0.000	0.771	0.227	0.002	0.000
Class II	0.135	0.717	0.144	0.004	0.103	0.744	0.142	0.001
Class III	0.001	0.150	0.736	0.113	0.001	0.163	0.680	0.156
Class IV	0.000	0.001	0.191	0.808	0.000	0.008	0.121	0.780

to latent classes with good discrimination, we computed the mean posterior probability of class membership for donors. These results confirm that our model chooses the class memberships with a high posterior probability for both sexes (68%-81%); see Table A1.

Prior for parameters

In a Bayesian model, prior distributions must be specified for all parameters. For the one-class (mixed-effects) model, non-informative proper priors were assigned. For models with more than one class, to ensure a well-identified model, the priors for the latent class parameters (θ_{k0} and θ_{k1}) were based on the posterior means from the one-class model. To make these priors less informative, we specified large variances for them, i.e., the number of donors times the variance of the posterior distribution from the one-class model (Garrett and Zeger, 2000). In addition, for the class-membership parameters, we used normal prior distributions with mean zero and variance equal to $9/4$, as was suggested by Elliott et al. (2005) and Garrett and Zeger (2000). Finally, for the other model parameters, weakly informative proper priors were assigned.

Predicting latent class membership

The values λ_{ik} describe the probability that the ith donor belongs to class k given the age and the Hb level at the screening visit. As soon as information from subsequent visits becomes available, these probabilities can be updated to incorporate the new information and yield better predictions of a donor's latent class. The updated probability that individual i belongs to the kth latent class can be calculated as

$$P(c_i = k|\text{Hb}_i, \text{Age}_{0i}, \text{Hb}_{0i}, \theta_k) = \frac{\lambda_{ik} f_k(\text{Hb}_i|\theta_k)}{\sum_{k=1}^{K} \lambda_{ik} f_k(\text{Hb}_i|\theta_k)},$$

where $P(c_i = k|\text{Hb}_i, \text{Age}_{0i}, \text{Hb}_{0i}, \theta_k)$ is the probability that donor i belongs to the kth class given the age and the Hb level at the screening visit and the history of Hb levels for that donor. $f_k(\text{Hb}_i|\theta_k)$ is the density of Hb levels for this donor given that he or she belongs to class k.

3.5 Supplementary

Jags codes to implement a latent class mixed-effects model with 4 classes. The first class is restricted to be stable regarding the number of donations in last two years.

```
model {

for(j in 1:n) { ### Begin j loop

for(i in offset[j]:(offset[j+1]-1)){  ### Begin i loop

                Hb[i]~dnorm(mu[i],tau)
                mu[i]<-beta[1]*Age[j]+beta[2]*Season[i]+
                beta[3]*TSPD[i]+beta[4]*TSPD2[i]+
                beta[5]*Indicator[i]+
### 1st class
                equals(g[j],1)*(theta1[1]+b01[j])+

### 2nd class
                equals(g[j],2)*(theta2[1]+b02[j,1]+
                (theta2[2]+b02[j,2])*Donation[i])+

### 3rd class
                equals(g[j],3)*(theta3[1]+b03[j,1]+
                (theta3[2]+b03[j,2])*Donation[i])+

### 4th class
                equals(g[j],4)*(theta4[1]+b04[j,1]+
                (theta4[2]+b04[j,2])*Donation[i])

                }  ### End i loop

### Random effects' priors
        b01[j]~dnorm(0,taub)
        b02[j,1:2]~dmnorm(mub,Omega2[,])
        b03[j,1:2]~dmnorm(mub,Omega3[,])
        b04[j,1:2]~dmnorm(mub,Omega4[,])

### Latent class indicator
        g[j] ~ dcat(psi[j,])
```

```
        for(class in 1:m) {  ### Begin m loop
                psi[j,class]<-phi[j,class]/sum(phi[j,])
                log(phi[j,class])<-gamma[class,1]+
                gamma[class,2]*Age[j]+gamma[class,3]*Hb0[j]
                }   ### End m loop

        }  ### End j loop

## Priors for class assignment parameters
for(c1 in 1:3) {gamma[1,c1]<-0}
for(c2 in 1:3) {gamma[2,c2]~dnorm(0,0.4444)}
for(c3 in 1:3) {gamma[3,c3]~dnorm(0,0.4444)}
for(c4 in 1:3) {gamma[4,c4]~dnorm(0,0.4444)}

R[1, 1]<-1.E-3
R[1, 2]<-0
R[2, 1]<-0
R[2, 2]<-1.E-3

Omega2[1:2, 1:2]~dwish(R, 2)
sigmab2[1:2, 1:2]<-inverse(Omega2[,])

Omega3[1:2, 1:2]~dwish(R, 2)
sigmab3[1:2, 1:2]<-inverse(Omega3[,])

Omega4[1:2, 1:2]~dwish(R, 2)
sigmab4[1:2, 1:2]<-inverse(Omega4[,])

taub~dgamma(1.E-3, 1.E-3)

tau~dgamma(1.E-3, 1.E-3)
sigma2<-1/tau
sigma<-sqrt(sigma2)

for(k in 1:5) {
beta[k]~dnorm(0, 1.E-3)
}
```

```
theta1[1]~dnorm(9.636571, 1.627573)

theta2[1]~dnorm(9.636571, 1.627573)
theta2[2]~dnorm(-0.05665738, 47.71125)I(,0)

theta3[1]~dnorm(9.636571, 1.627573)
theta3[2]~dnorm(-0.05665738, 47.71125)I(,0)

theta4[1]~dnorm(9.636571, 1.627573)
theta4[2]~dnorm(-0.05665738, 47.71125)I(,0)

} ### End
```

Stan codes to implement a latent class mixed-effects model with 4 classes. The first class is restricted to be stable regarding the number of donations in last two years.

```
data {
// Number of latent class
  int k ;
// Number of Observations
  int ntot ;
// Number of individuals
  int n ;
  real Age0[ntot] ;
  real Age[n] ;
  int<lower=1> offset[n+1];
  real <lower=0,upper=1> Season[ntot] ;
  real <lower=0> TSPD[ntot] ;
  real <lower=0> TSPD2[ntot] ;
  real <lower=0> Indicator[ntot] ;
  real Hb[ntot];
  real Hb0[n];
  real freqy2[ntot];
 }

parameters {
//Random intercepts in class I
real b1[n];
```

```
//Random intercepts and slopes class II
vector[2] b2[n];
//Random intercepts and slopes class III
vector[2] b3[n];
//Random intercepts and slopes class III
vector[2] b4[n];

//Scale parameters Random intercepts class I
real <lower=0,upper=0.2> sigma1b;

//Scale parameters Random intercepts class II
real <lower=0,upper=0.2> sigma21b;

//Scale parameters Random intercepts class III
real <lower=0,upper=0.2> sigma31b;

//Scale parameters Random intercepts class IV
real <lower=0,upper=0.2> sigma41b;

//Scale parameters Random slopes class II
real <lower=0,upper=0.2> sigma22b;

//Scale parameters Random slopes class III
real <lower=0,upper=0.2> sigma32b;

//Scale parameters Random slopes class IV
real <lower=0,upper=0.2> sigma42b;

//scale parameters of mixture components
 vector[k-1] gamma[3];

 real <lower=0,upper=0.4> sigma;

// Class specific parameters
 vector[k] theta1;

// Shared parameters
 vector[5] beta;

  real <lower=-0.5,upper=0> theta21;
```

```
  real <lower=-0.5,upper=0> theta22;
  real <lower=-0.5,upper=0> theta23;

}


transformed parameters {
vector[n] pi1; //mixing proportions
vector[n] pi2; //mixing proportions
vector[n] pi3; //mixing proportions
vector[n] pi4; //mixing proportions

vector[n] ppi1; //mixing proportions
vector[n] ppi2; //mixing proportions
vector[n] ppi3; //mixing proportions

for(j in 1:n){
ppi1[j]<-exp(gamma[1,1]+ gamma[2,1]*Age[j]+ gamma[3,1]*Hb0[j]);
ppi2[j]<-exp(gamma[1,2]+ gamma[2,2]*Age[j]+ gamma[3,2]*Hb0[j]);
ppi3[j]<-exp(gamma[1,3]+ gamma[2,3]*Age[j]+ gamma[3,3]*Hb0[j]);
}

for(j in 1:n){
pi1[j]<-ppi1[j]/(ppi1[j]+ppi2[j]+ppi3[j]+1);
pi2[j]<-ppi2[j]/(ppi1[j]+ppi2[j]+ppi3[j]+1);
pi3[j]<-ppi3[j]/(ppi1[j]+ppi2[j]+ppi3[j]+1);
pi4[j]<-1/(ppi1[j]+ppi2[j]+ppi3[j]+1);
}
}


model {

real psi[k];

increment_log_prob(normal_log(gamma[1,1],0,1.5));
increment_log_prob(normal_log(gamma[2,1],0,1.5));
increment_log_prob(normal_log(gamma[3,1],0,1.5));
```

```
increment_log_prob(normal_log(gamma[1,2],0,1.5));
increment_log_prob(normal_log(gamma[2,2],0,1.5));
increment_log_prob(normal_log(gamma[3,2],0,1.5));

increment_log_prob(normal_log(gamma[1,3],0,1.5));
increment_log_prob(normal_log(gamma[2,3],0,1.5));
increment_log_prob(normal_log(gamma[3,3],0,1.5));


//sigma2b~inv_wishart(2,Rb);
//sigma3b~inv_wishart(2,Rb);
//sigma4b~inv_wishart(2,Rb);

sigma1b~inv_gamma(0.01,0.1);
sigma21b~inv_gamma(0.005,0.1);
sigma31b~inv_gamma(0.005,0.1);
sigma41b~inv_gamma(0.005,0.1);
sigma22b~inv_gamma(0.005,0.1);
sigma32b~inv_gamma(0.005,0.1);
sigma42b~inv_gamma(0.005,0.1);

sigma~inv_gamma(0.02,0.1);




for (j in 1:n){
increment_log_prob(normal_log(b1[j],0,sigma1b));
increment_log_prob(normal_log(b2[j,1],0,sigma21b));
increment_log_prob(normal_log(b3[j,1],0,sigma31b));
increment_log_prob(normal_log(b4[j,1],0,sigma41b));

increment_log_prob(normal_log(b2[j,2],0,sigma22b));
increment_log_prob(normal_log(b3[j,2],0,sigma32b));
increment_log_prob(normal_log(b4[j,2],0,sigma42b));

//b2[j]~multi_normal(mub,sigma2b);
//b3[j]~multi_normal(mub,sigma3b);
//b4[j]~multi_normal(mub,sigma4b);
}
```

```
increment_log_prob(normal_log(beta,0,1));

increment_log_prob(normal_log(theta1[1],9.63,0.5));
increment_log_prob(normal_log(theta1[2],9.63,0.5));
increment_log_prob(normal_log(theta1[3],9.63,0.5));
increment_log_prob(normal_log(theta1[4],9.63,0.5));

increment_log_prob(normal_log(theta21,-0.06,0.1));
increment_log_prob(normal_log(theta22,-0.06,0.1));
increment_log_prob(normal_log(theta23,-0.06,0.1));

for(j in 1:n)
{
  for(i in offset[j]:(offset[j+1]-1)){

  psi[1]<- log(pi1[j])+normal_log(Hb[i],theta1[1]+b1[j]+
     beta[1]*Age0[i]+beta[2]*Season[i]+beta[3]*TSPD[i]+
     beta[4]*TSPD2[i]+beta[5]*d[i], sigma);// Class one

  psi[2]<- log(pi2[j])+normal_log(Hb[i],theta1[2]+b2[j,1]+
     beta[1]*Age0[i]+beta[2]*Season[i]+beta[3]*TSPD[i]+
     beta[4]*TSPD2[i]+beta[5]*d[i]+
     (theta21+b2[j,2])*freqy2[i], sigma);// Class two

  psi[3]<- log(pi3[j])+normal_log(Hb[i],theta1[3]+b3[j,1]+
    beta[1]*Age0[i]+beta[2]*Season[i]+beta[3]*TSPD[i]+
    beta[4]*TSPD2[i]+beta[5]*d[i]+
    (theta22+b3[j,2])*freqy2[i], sigma);// Class three

  psi[4]<- log(pi4[j])+normal_log(Hb[i],theta1[4]+b4[j,1]+
    beta[1]*Age0[i]+beta[2]*Season[i]+beta[3]*TSPD[i]+
    beta[4]*TSPD2[i]+beta[5]*d[i]+
    (theta23+b4[j,2])*freqy2[i], sigma);// Class four

increment_log_prob(log_sum_exp(psi));
}
}
}
```

4 Prediction of hemoglobin in blood donors using a latent class mixed-effects transition model

This chapter is published as: Nasserinejad K, van Rosmalen J, de Kort W, Rizopoulos D, and Lesaffre E. Prediction of hemoglobin in blood donors using a latent class mixed-effects transition model. *Statistics in Medicine*, 2016 ;35(4):581-594. doi: 10.1002/sim.6759

Abstract.

BLOOD donors experience a temporary reduction in their hemoglobin (Hb) value after donation. At each visit, the Hb value is measured, and a too low Hb value leads to a deferral for donation. Because of the recovery process after each donation as well as state dependence and unobserved heterogeneity, longitudinal data of Hb values of blood donors provide unique statistical challenges. To estimate the shape and duration of the recovery process, and to predict future Hb values, we employed three models for the Hb value: (i) a mixed-effects models; (ii) a latent-class mixed-effects model; and (iii) a latent-class mixed-effects transition model. In each model, a flexible function was used to model the recovery process after donation. The latent classes identify groups of donors with fast or slow recovery times and donors whose recovery time increases with the number of donations. The transition effect accounts for possible state dependence in the observed data.

All models were estimated in a Bayesian way, using data of new entrant donors from the Donor InSight study. Informative priors were used for parameters of the recovery process that were not identified using the observed data, based on results from the clinical literature. The results show that the latent-class mixed-effects transition model fits the data best, which illustrates the importance of modeling state dependence, unobserved heterogeneity, and the recovery process after donation. The estimated recovery time is much longer than the current minimum interval between donations, suggesting that an increase of this interval may be warranted.

4.1 Introduction

Blood donation helps to save millions of lives each year and is an essential part of modern healthcare. Many blood donors come for donation on a regular basis. A blood donation implies a loss of erythrocytes and iron, resulting in a temporary decrease in hemoglobin (Hb) values. The minimum interval between two donations is internationally set at 8 weeks but this interval seems to be too short for the body to completely recover the Hb value to its pre-donation value. Previous studies have shown that among donors with many visits, there is a decline in their Hb values at subsequent donations (Cable et al., 2011; Brittenham, 2011). Low Hb values may potentially lead to anemia, which should be prevented. In the majority of blood banks, the Hb value is first measured to screen the potential donor for eligibility to give blood (Radtke et al., 2005). To protect potential donors from developing low Hb values, a Hb value of 8.4 mmol/l (135 g/l) and 7.8 mmol/l (125 g/l) for men and women, respectively, is widely accepted as the lower cut-off value of eligibility for donation (Radtke et al., 2005).

Each year a considerable proportion of prospective blood donors are temporarily deferred from donation due to low Hb values (Newman, 2004; Popovsky, 2012). Hb deferrals decrease the cost-effectiveness of blood supply, because a) testing and deferring a donor is expensive, b) for every deferred donor another donor needs to be invited to reach collection targets, and c) lapsing donors need to be replaced by new donors since deferred candidates rarely return for donation (Halperin et al., 1998).

To limit the number of deferrals, it is important that blood banks are able to predict when donors have sufficiently recovered after a blood donation to be invited for a new donation. However, prediction of the Hb value for the subsequent visit of a blood donor is not straightforward. Longitudinal data of Hb values of blood donors have a high within-subject and between-subject variability due to unobserved individual heterogeneity and state dependence (Heckman, 1981). This requires a complex statistical model to capture both sources of variation for proper inference. Predictions of future Hb values of blood donors may be used to improve the invitation policy of potential blood donors.

Hence, for a model to be appropriate, two sources of variation among Hb values must be taken into account. There are various statistical techniques for analyzing longitudinal data. Two well-known approaches are a) the mixed-effects model, which can capture unobserved individual heterogeneity and b) the transition (auto-regressive) model, which can capture state dependence within an individual's observations. However, it may be that neither approach can adequately explain the correlation structure alone, due to the presence of both unobserved heterogeneity and state dependence. For these reasons, we may combine the mixed-effects model and the transition model by including the unobserved individual-specific effects and a lagged endogenous variable in a single regression model. Such a mixed-effects transition model is popular in econometrics for forecasting (Diggle et al., 2002), but less commonly used in medical applications (Funatogawa and

Funatogawa, 2012).
This paper aims to build an appropriate mixed-effects transition model for a longitudinal data set of Hb values collected from new entrant whole blood donors from 2005 to 2012 in the Netherlands. We not only wish to predict the future Hb value but also the recovery time that is needed for Hb to return to its pre-donation value after a blood donation. This will help to improve the planning of the donors' visits to the blood bank. Predicting Hb value has been studied previously by Rikhtehgaran et al. (2012) and Nasserinejad et al. (2013). In the latter study, the authors applied mixed-effect models and transition models separately to predict the Hb value. In the former study, a mixed-effects transition model was applied on a subset of the longitudinal data, excluding all measurements after the candidate was deferred from donation for any reason. Recently, Nasserinejad et al. (2015) used a latent class mixed-effects model to show how the trajectories of Hb values differ between donors.
The main contribution of this paper is the development of a realistic model for longitudinal Hb values. This model takes into account the unique aspects of these data, namely a) heterogeneity of the initial Hb value, b) state dependence of a donor's Hb values, c) varying time intervals between donations, d) the temporary reduction in Hb after blood donation, and e) the fact that the recovery process may change with the number of donations and may differ between donors. To do so, the mixed-effects transition model is combined with a flexible nonlinear specification for the recovery process, which enables us to estimate the shape and duration of the recovery, for which precise estimates are not available in the clinical literature. Latent classes are used to account for the heterogeneity of the recovery process between donors, and a time change point specification is used to determine how the recovery time changes with the number of donations. The observed data do not provide sufficient information to identify all model parameters in a frequentist setting. A further contribution of this paper is that it is shown how the model parameters can be estimated in a Bayesian way, using suitable prior information from the clinical literature. Finally, it is shown how to forecast a future Hb value based on the available history of donations in a fully Bayesian approach.
The remainder of the paper is organized as follows: Section 4.2 presents the motivating data set. Section 4.3 introduces the statistical models for the recovery process and the donation problem described above. Results are presented in Section 4.4. Section 4.5 deals with prediction in the latent class mixed-effects transition model. We discuss the findings of the study in Section 4.6.

4.2 Sanquin blood donor data set

The Donor InSight data set was collected by Sanquin Blood Supply in the Netherlands (Atsma et al., 2011). This data set is based on a self-administered questionnaire study

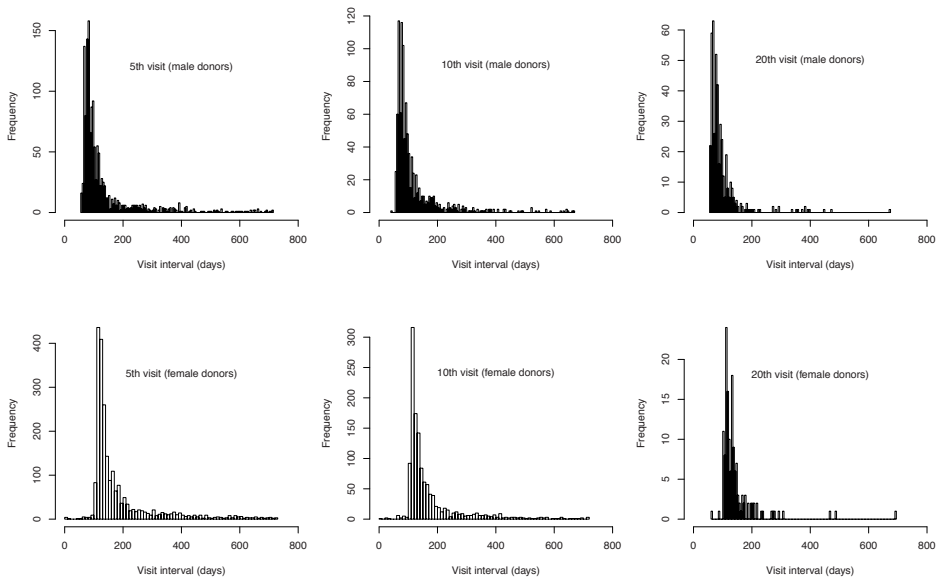


Figure 4.1: Interval distributions for three different visits (i.e., 5th, 10th, and 20th visit) are shown for each sex, separately.

aimed at gaining insight into characteristics and motivation of the Dutch donor population (Atsma et al., 2011). Our analysis comprises whole blood donors who were registered as a new donor from 1 January 2005 to 31 December 2012. Whole blood is a term used in transfusion medicine for a standard blood donation as opposed to plasma and platelet donation. To be included in the study, the donors must have had at least one visit after a first donation. A total of 4461 donors (1552 male and 2909 female donors) fulfilled these criteria. The descriptive statistics of these donors are presented in Table 4.1. Figure 4.1 shows the interval distributions for the 5th, the 10th, and the 20th visit for each sex separately.

Table 4.1: Descriptive statistics of the Donor InSight data set.

	Male	Female
Age at 1st visit* (year)	35 (25 - 47)	30 (22 - 43)
Number of donations*	12 (6 - 19)	7 (4 - 13)
Visit intervals* (day)	89 (73 - 119)	130 (116 - 168)
Deferral due to low Hb	5.29%	11.38%
Donors with at least one deferral due to low Hb	36.92%	54.40%

* Median and inter-quantile range.

For blood collection, all measured data were entered into the blood bank computer system (e)PROGESA (MAK-SYSTEM International Group, France). Prior to donation, Hb and other parameters undergo a check to determine whether the prospective donor is eligible. In the Netherlands, a newly registered donor is not allowed to donate blood at the first visit, the screening visit, which consists of a health check only. At every subsequent visit, donors who pass all eligibility checks may donate 500 ml of whole blood. Finally, guidelines impose a minimum interval of 56 days between donations, with a yearly maximum of 5 donations for men and 3 for women (Baart et al., 2011).

In Figure 4.2, profiles of the Hb value after the screening visit for a subset of male and female donors are displayed. The horizontal lines represent the eligibility thresholds for donation.

Several factors are known to be associated with the Hb value and hence may be used as predictors for Hb, i.e., sex (Yip et al., 1984), season (Hoekstra et al., 2007), age (Yip et al., 1984), and body mass index (BMI) (Yip et al., 1984). In this study, we take into account the effect of age at first visit and the season of the visit (a binary covariate, i.e., the cold season includes fall and winter and the warm season includes spring and summer). Because male and female donors have different Hb profiles, we analyze the data for men and women separately. The Donor InSight study was approved by the Medical Ethical Committee Arnhem-Nijmegen in the Netherlands and all participants gave their written informed consent.

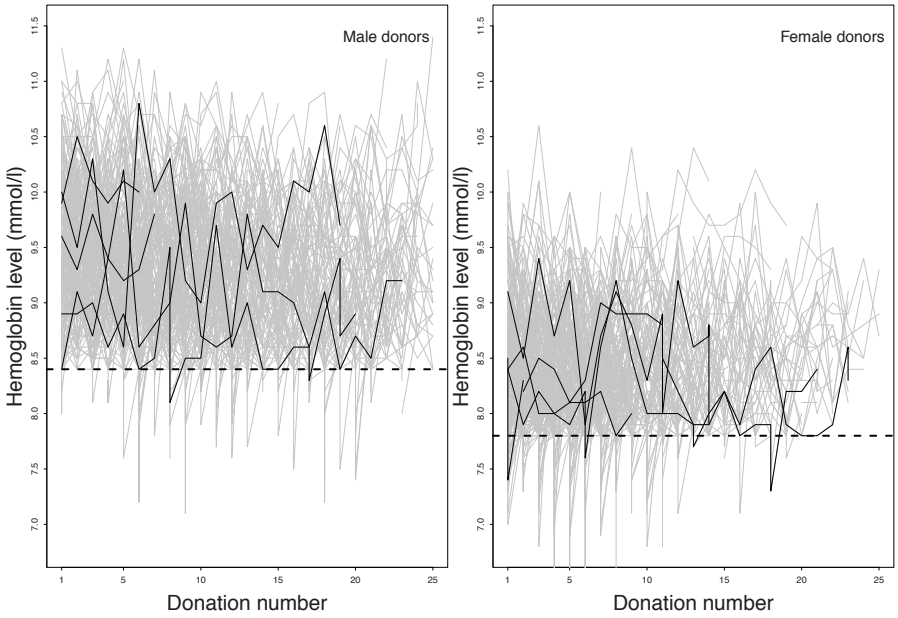


Figure 4.2: Hb profiles after the screening visit for a subset of male and female donors. Five random profiles are highlighted for both sexes. The bold dashed lines show the corresponding thresholds of eligibility for donation.

4.3 Statistical model for Hb values

In this section, we propose a mixed-effects model, a latent class mixed-effects model and a latent class mixed-effects transition model for the Donor InSight data. Formally, let Hb_{it} denote the Hb recorded for the ith individual $(i = 1, \ldots, N)$ at T_i different times $(t = 1, \ldots, T_i)$, together with a set of p strictly exogenous covariates $x_{it} = \{x_{i1t}, x_{i2t}, \ldots, x_{ipt}\}'$. Since the donation intervals are not equal, the data set is unbalanced (see Figure 4.1). Furthermore, there is a reduction in Hb value after donation. Since the current Hb value is associated with the Hb value observed at the previous visit (Rikhtehgaran et al., 2012; Nasserinejad et al., 2013) and the time the since previous donation (Rikhtehgaran et al., 2012; Nasserinejad et al., 2015), we need to take into account the time interval since the previous donation and the reduction in Hb due to donation. All of these aspects must be incorporated into the statistical model for predicting Hb.

Hb recovery after blood donation

The Hb recovery process after blood donation is illustrated in Figure 4.3. In this figure δ indicates the time that Hb reaches its minimum value after donation. After donation of

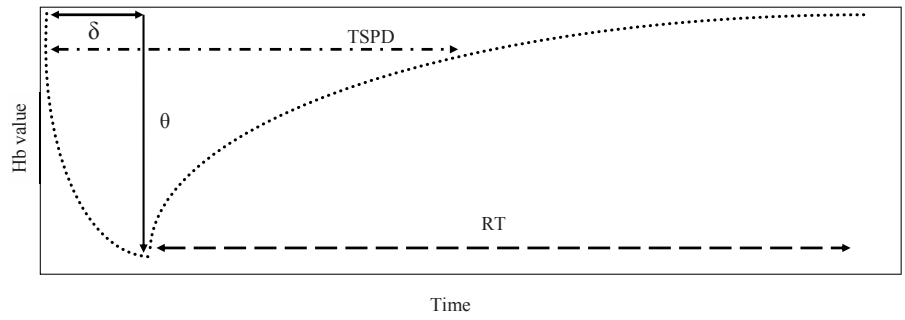


Figure 4.3: Hb recovery process after blood donation: δ indicates the time that Hb reaches its minimum value after donation, RT indicates the recovery time that is needed for Hb to return to its pre-donation value, and θ shows the amount of reduction in Hb after donation.

500 ml of blood, on average a male donor loses 242 mg and a female donor 217 mg of iron (Simon, 2002). This will cause Hb to decrease and to reach its nadir a few days after donation (Boulton, 2004; Wadsworth, 1955; Kiss et al., 2015). The Hb value will then gradually recover to its pre-donation value (Boulton, 2004; Wadsworth, 1955; Kiss et al., 2015). The recovery time needed for Hb_{it} to return to its pre-donation value is given by RT. Unfortunately in this study we have only the Hb value prior to donation at each visit, and there is no information on the Hb value between two invitations. The observed interdonation interval is at least 56 days in our data set. Therefore, we cannot accurately estimate the trajectory of Hb value during the first 56 days after donation. We incorporate into our model that the reduction of Hb after giving blood takes around three days (Boulton, 2004; Kiss et al., 2015). To estimate the amount of reduction after donation, we use the results of a recent randomized clinical trial on iron supplementation after blood donation. The donor characteristics and the donation policy in this trial were similar to the Dutch setting of the Donor InSight study; the amount of blood given per donation and the minimum interval between donations are the same for both data sets. The data of this trial showed that, 3 through 8 days after donation of 500 mL of whole blood, the amount of reduction in the Hb value had a mean (95% CI) of 0.68 (0.59, 0.77) and 0.96 (0.89, 1.03) mmol/l for male and female donors, respectively (Kiss et al., 2015). We use the following Hb recovery function (HRF):

$$\text{HRF}_{it} = \theta \left[\max \left(\frac{\text{RT} - (\text{TSPD}_{it} - \delta)}{\text{RT}}, 0 \right) \right]^{\psi}, \tag{4.1}$$

where HRF_{it} is the amount Hb that needs to be recovered for person i at time t. δ and RT were defined above and θ is the amount of Hb reduction. TSPD_{it} represents the time since the previous donation for the ith donor at time t. The parameter ψ indicates the shape of the recovery function of the Hb value after donation. Values of ψ smaller than one indicate a convex trajectory for Hb recovery, while values greater than one indicate a concave

trajectory for Hb recovery (as in Figure 4.3). For $\psi = 1$, the Hb recovery is linear.
The recovery time may differ between donors due to diet, genetic factors, and other unobserved characteristics. In addition, some donors experience a reduction of their iron reserve after a few donations, which may cause the recovery time to increase with the number of donations. Therefore, the postulated function based on (4.1) can oversimplify the recovery process, because the recovery time (RT) may depend on the number of previous donations as well as the donor. However, a too flexible function (e.g., a different recovery time for each donor) is likely not to be estimable from the data. So, we use finite mixture modeling to capture the heterogeneity in recovery time by assuming that donors belong to two different latent classes. Namely, one class of donors is assumed to have a constant recovery time and the other class is assumed to exhibit a nonconstant recovery time. The function could be:

$$\text{HRF}_{it,g(i),\kappa} = \theta \left[\max \left(\frac{\text{RT}_{g(i),\kappa} - (\text{TSPD}_{it} - \delta)}{\text{RT}_{g(i),\kappa}}, 0 \right) \right]^{\psi}, \tag{4.2}$$

where $g(i)$ is the latent class membership of donor i, which has to be estimated from the data. It is assumed that donors who belong to (latent) class I ($g(i) = 1$) have a constant recovery time. For donors in class II, it is assumed that the recovery time changes after a certain number of donations (κ). This change point κ is assumed equal for all people in class II and must be estimated from the data.

Mixed-effects model

To accumulate the effects of the donation and recovery process a non-linear model making use of (4.1) or (4.2) is needed. To also take into account the unobserved individual heterogeneity a random intercept is included. Therefore a possible model with (4.1) as recovery function could be:

$$\text{Model 1:} \quad \text{Hb}_{it} = \beta_0 + b_{i1} + \beta_1 \text{Age}_{i1} + \beta_2 \text{Season}_{it} - \text{HRF}_{it} + \varepsilon_{it}, \quad t = 1, \ldots, T_i \tag{4.3}$$

where b_{i1} is the random intercept in the model to control heterogeneity. It is assumed that $\epsilon_{it} \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$ and $b_{i1} \overset{\text{iid}}{\sim} N(0, \sigma_{b1}^2)$ and that b_{i1} and ϵ_{it} are mutually independent.

Latent class mixed-effects model

To take into account the heterogeneity in recovery time in Model 1, a possible model based on (4.2) as recovery function could be:

$$\text{Model 2:} \quad \text{Hb}_{it} = \beta_0 + b_{i1} + \beta_1 \text{Age}_{i1} + \beta_2 \text{Season}_{it} - \text{HRF}_{it,g(i),\kappa} + \varepsilon_{it}, \quad t = 1, \ldots, T_i \tag{4.4}$$

Latent class mixed-effects transition model

To take into account the state dependence among Hb values of the same individual, one can add a first lag of Hb as an autoregressive term to Model 2. Because the recovery process depends on the time since the last donation, we also use the Hb value at the last donation (instead of the last visit) as an autoregressive term. This model is given by:

$$\begin{aligned} \text{Model 3:} \quad \text{Hb}_{it} &= \beta_0 + b_{i1} + \beta_1 \text{Age}_{i1} + \beta_2 \text{Season}_{it} + \\ &\gamma \text{Hb.pd}_{it} - \text{HRF}_{it,g(i),\kappa} + \varepsilon_{it}, \quad t = 2, \ldots, T_i \end{aligned} \tag{4.5}$$

where b_{i1} controls the heterogeneity now partly explaining the intra-subject correlation, γ is the lagged impact of Hb at the previous donation (Hb.pd_{it}). Note that Hb.pd_{it} is equal to $\text{Hb}_{i,t-1}$ if the last visit was a donation. For a stationary process, i.e., $|\gamma| < 1$, the correlation between two subsequent measurements, can be expressed as (Rikhtehgaran et al., 2012):

$$\rho_{Hb_{it},Hb_{it-1}} = \gamma + \frac{1-\gamma}{1 + (1-\gamma)\sigma_\varepsilon^2/[(1+\gamma)\sigma_{b1}^2]}, \tag{4.6}$$

when the lag impact γ is negligible, this correlation reduces to the intra-class correlation (ICC) in the mixed-effects model. On the other hand, when there is no heterogeneity between individuals, i.e., $\sigma_{b1}^2 \approx 0$, the correlation is equal to the lag impact only.

The initial conditions problem

One of the assumptions in classical mixed-effects models is that the covariates in the model are exogenous, i.e., the covariance between covariates and the random effects are zero. But this assumption is violated in mixed-effects transition models where one of the covariates is the lagged variable, which is endogenous. This issue relates to the initial conditions problem (ICP), which is well-known in the econometric literature. The ICP occurs due to the fact that the individual effects, b_{i1}, that capture the unobserved heterogeneity are correlated with the initial Hb values, i.e., $\text{cov}(\text{Hb}_{i1}, b_{i1}) \neq 0$ (Kazemi and Davies, 2002; Kazemi and Crouchley, 2006). Ignoring the ICP and thus the endogeneity of Hb_{i1} results in inconsistent estimates in the model (Kazemi and Davies, 2002; Kazemi and Crouchley, 2006), i.e., an upward bias of the estimated state dependence and a downward bias in the estimated coefficients of explanatory variables (Kazemi and Crouchley, 2006).
A possible solution to the ICP problem is to incorporate the association of the initial value and the random effects jointly into the model for the subsequent Hb values. The model is assumed to be similar to the dynamic equations (Models 2 and 3), but without the lagged response variables (Kazemi and Crouchley, 2006). Using this solution, the regression parameters as well as the residual variance are allowed to differ between the initial and the subsequent observations. The joint modeling approach enables one to capture the correlation between the individual effects, b_{i1}, and the initial Hb values and provides reliable

estimates for the regression parameters (Kazemi and Crouchley, 2006). A possible model for the initial Hb values could be:

$$\text{Model } 3_0: \quad Hb_{i1} = \beta_{00} + \beta_{01}\text{Age}_{i1} + \beta_{02}\text{Season}_{i1} + b_{i0}, \tag{4.7}$$

where $b_{i0} \sim N(0, \sigma^2_{b0})$. Furthermore, we allow for a correlation ρ_{01} between b_{i0} and b_{i1} in Model 3.

Due to the complexity of the models as well as the lack of information on the first 56 days of the recovery process, we opted for a Bayesian approach with Markov chain Monte Carlo (MCMC) sampling to estimate the parameters in these models.

Prior specification

Vague normal priors were specified for the β's in Model 1, Model 2, and Model 3_0, and for the β's and γ in Model 3, i.e., $N(0, 10^3)$. Since Hb is measured only at visits to the blood bank, and the minimum interval between donations is 56 days, the current data provide little information on the parameters θ and δ. Therefore, we used informative priors for the amount of Hb reduction after donation, θ, and the time at which Hb value reaches its minimum value, δ. The parameters of the informative prior distributions are based on previous clinical studies (Boulton, 2004). For θ, we specified a normal prior with mean 0.68 and standard deviation 0.038 for male donors and a normal prior with mean 0.96 and standard deviation 0.045 for female donors, based on the results of a recent clinical trial (Kiss et al., 2015). For δ, a normal prior with mean 3 and variance equal to 0.1 was specified. For the recovery time RT, a positive uniform distribution with a wide range, i.e., $U(0, 10^4)$ was specified. For the parameter ψ, which indicates the shape of the recovery trend, an exponential of normal distribution, i.e., $N(0, 10)$ was specified. An IG$(\varepsilon, \varepsilon)$ prior with small value for ε, i.e., 10^{-3}, was specified for the variance of the residuals. An Inv-Wishart(R, df) distribution was specified for the variance-covariance structure of the residuals b_{i0} in Model 3_0 and the random intercept b_{i1} in Model 3. We set the degrees of freedom, df, to 3 and the scale parameter matrix, R, to a diagonal matrix with small values, i.e., 10^{-3} (Lesaffre and Lawson, 2012). For the class membership probability a Dirichlet distribution with small values (i.e., 0.1) for the mixing distribution was specified (Rousseau and Mengersen, 2011). We chose a discrete uniform distribution for κ with range from 1 to the maximum number of donations, i.e., 43 and 26 for male and female donors, respectively. Finally, prior sensitivity analyses were performed for the non-informative prior distributions.

Model fit and assessment

To be able to validate the model, we randomly divided the available donors into a training data set (2,231 donors) and a validation data set (2,230 donors). All statistical models were estimated using the training data set only. Parameter estimates were obtained using

the JAGS (Plummer, 2011) interface to R (R Development Core Team, 2010); the JAGS syntax is shown in the Appendix. The first 10,000 iterations (i.e., burn-in iterations) of each chain were discarded. The posterior medians and 95% HPD credible intervals (CI) were calculated using the remaining iterations of each chain, using a thinning factor of 10. We checked the convergence by running two chains from dispersed initial values and examining standard Bayesian diagnostics, such as trace plots, the Brooks-Gelman-Rubin statistics (Gelman et al., 2014), and the Geweke diagnostic (Geweke et al., 1991). To check whether the number of MCMC iterations was sufficient to obtain accurate estimates, sampling was continued until the Monte Carlo errors were less than 5% of the posterior standard deviation of each parameter (Lesaffre and Lawson, 2012). The label switching problem can not occur in Models 2 and 3, because one class in these models was restricted to have a stable recovery time.

To find the best fitting model for the data, we computed the deviance information criterion (DIC) (Spiegelhalter et al., 2002). The DIC was computed outside Jags using a self-written R program (R Development Core Team, 2010) based on MCMC and the data used to fit the models. For this calculation the expectations of the class-membership and time change point parameters were chosen based on the mode and for the other parameters this expectation was based on the median. Finally, to assess goodness-of-fit, we used an omnibus posterior predictive check (PPC) (Gelman et al., 1996). We generated Hb values given the parameters (Φ) from a random sample of draws from the posterior distribution and calculated the chi-square statistic

$$X^2 = \sum (\text{Hb}_{it} - \text{E}(\text{Hb}_{it}|\Phi))^T \text{Var}(\text{Hb}_{it}|\Phi)^{-1} (\text{Hb}_{it} - \text{E}(\text{Hb}_{it}|\Phi))$$

for both replicated and observed Hb values at each iteration. Then we computed a Bayesian P-value, i.e., the probability that X^2 based on the replicated data from the model is more extreme than the X^2 based on the observed data. Small or large values of this P-value (e.g., < 0.05 or > 0.95) indicate a poor fit of the model to the data (Gelman et al., 1996).

Simulation study

To evaluate the ability of the proposed model to estimate the true parameters, we performed a simulation study using 20 artificially generated data sets. In each data set Hb values and covariates of 200 male donors were generated according to Model 3 and Model 3_0, using the posterior medians of the parameters estimated using the Sanquin data set. In the simulated data, donors were only allowed to donate if the Hb value was above the cut-off for eligibility. The distributions of age, season and the number of visits per donor were simulated from the observed data for male donors. For each model parameter, the posterior median and the exceedance probability (i.e., the posterior probability that the es-

timated parameter is greater than the true value) were calculated for each generated data set.

4.4 Results

Donor InSight study results

The DICs for different models are presented in Table 4.2. The results show that Model 3 is the best model for both sexes. For this model, the Bayesian P-value of the PPC is 0.48 and 0.43 for males and females, respectively, which indicates that the model assumptions appear to be satisfied. The parameter estimates based on Model 3 (and Model 3_0) for both sexes are presented in Table 4.3. Based on the highest posterior probability, donors can be

Table 4.2: The effective number of parameters (p_D), and the deviance information criterion (DIC) for different models for each sex.

	Male		Female	
Model	p_D	DIC	p_D	DIC
Model 1	691	16189	1237	19576
Model 2	809	15360	1475	19211
Model 3 (and Model 3_0)	557	14230	1185	17397

assigned to latent classes. For Model 3 (and Model 3_0) 48.7% and 45.7% of the male and female donors, respectively, are assigned to the class with a nonconstant recovery time (class II). In class II the recovery time change point (κ) for male and female donors is at 7th and 4th donation, respectively. The profiles of the different classes for male and female donors are displayed in Figure 4.4 and Figure 4.5, respectively. The arrows in these figures indicate the recovery time change points. To contrast the two latent classes, sex-specific descriptive statistics including the average of the time since previous donation (day), age (year), percentage of deferral due to low Hb values, percentage of donors with at least one deferral due to low Hb values, and number of donations are presented in Table 4.4. These results show that the two classes are dramatically different in their percentage of deferral and the percentage of donors with at least one deferral due due to low Hb values. To determine the class membership discrimination, we computed the mean posterior probability of class membership for donors and report the results in Table 4.5. The mean posterior probability of the class to which a donor is assigned is approximately 70%. The estimated recovery time (95% CI) in class I is 100 (69, 145) and 54 (20, 129) days for male and female donors, respectively. These values are 117 (80, 168) and 343 (270, 450) days for male and female donors, respectively prior to the change point in class II and increase to 419 (293, 663) and 503 (394, 665) days after the change, much larger than the minimum

Table 4.3: The posterior medians and 95% HPD CIs based on Model 3 and Model 3_0 for male and female donors, separately.

	Male donors			Female donors		
Parameter	Estimate	95% CI		Estimate	95% CI	
β_0	7.87	7.65	8.09	6.96	6.78	7.15
β_1	-0.004	-0.006	-0.002	0.003	0.002	0.005
β_2	-0.082	-0.100	-0.064	-0.068	-0.084	-0.053
γ	0.19	0.17	0.21	0.17	0.15	0.19
θ	0.70	0.64	0.76	0.89	0.82	0.96
δ	2.98	2.39	3.60	2.98	2.32	3.58
ψ	2.18	1.49	3.61	3.61	2.26	4.04
κ	7	7	8	4	4	5
RT_{g_1}	100	69	145	54	20	129
$\text{RT}_{g_{21}}$	117	80	168	343	270	450
$\text{RT}_{g_{22}}$	419	293	663	503	394	665
σ_{b1}	0.34	0.31	0.35	0.29	0.28	0.31
σ	0.46	0.45	0.47	0.44	0.43	0.45
$\rho_{b_{i1},Hb_{i1}}$	0.67	0.64	0.69	0.61	0.58	0.63

Table 4.4: Sex-specific descriptive statistics for the two classes pertaining to Model 3 (and Model 3_0).

	Male		Female	
	Class I	Class II	Class I	Class II
Class size	54.3%	45.7%	51.3%	48.7%
Age at 1st visit* (year)	36 (27 - 46)	33 (25 - 45)	31 (22 - 46)	27 (21 - 41)
Number of donations*	11 (6 - 17)	12 (6 - 20)	8 (4 - 13)	6 (4 - 11)
Visit intervals* (day)	90 (73 - 124)	86 (72 - 115)	133 (119 - 175)	134 (119 - 180)
Deferral due to low Hb	0.9%	4.7%	4.5%	13.7%
Donors with at least one deferral due to low Hb	6%	31%	25%	57%

* Median and inter-quantile range.

(56 days) intra-donation interval. These results show a longer recovery time for female donors than for male donors in class II, though the uncertainty in the estimates is large. The lag impact of Hb for male and female (0.19 and 0.17, respectively) is highly significant, which shows that there is strong evidence for the state dependence between successive Hb values. This can also be induced from the estimated correlation between two successive observations in Models 2 and 3. For Model 2 the correlation (95% CI) is estimated as 0.45 (0.42, 0.48) and 0.42 (0.39, 0.44) between two successive observations in male and female

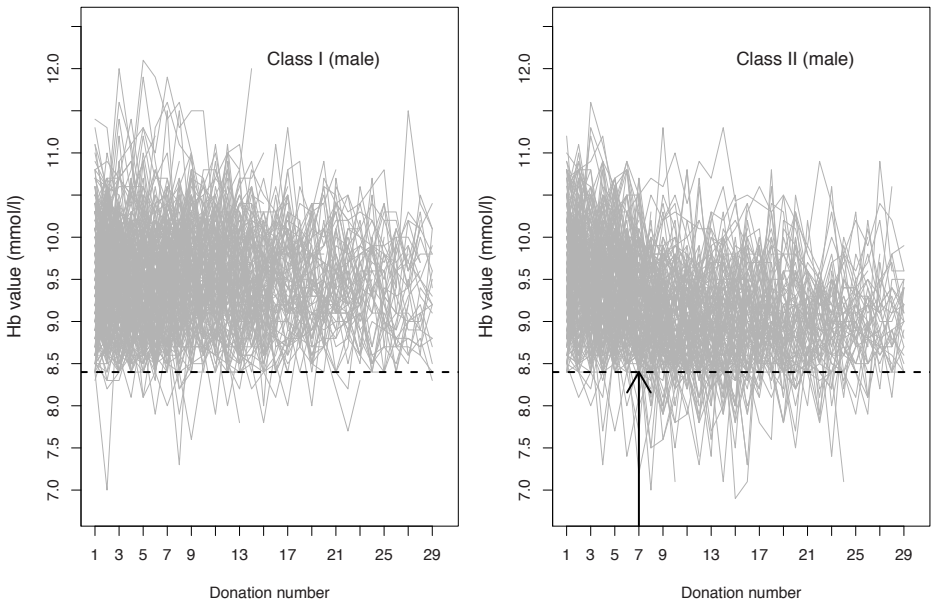


Figure 4.4: Male Hb profiles after screening visit separately for each class obtained from Model 3 (and Model 3_0). The donors assigned to the highest class probability. The horizontal lines show the threshold for donation eligibility. The arrow shows the recovery time change point.

Table 4.5: Average class probabilities by latent classes for both sexes based on Model 3 (and Model 3_0).

	Mean of posterior probabilities			
	Male		Female	
Latent Class	Class I	Class II	Class I	Class II
Class I	0.70	0.30	0.72	0.28
Class II	0.28	0.72	0.29	0.71

donors, respectively (not shown here). For Model 3, these correlations increase to 0.54 (0.51, 0.57) and 0.49 (0.46, 0.52) for male and female donors, respectively.

The posterior distributions for θ and δ are very close to the corresponding prior distributions. The prior and posterior means (95% CI) of θ for male donors were 0.68 (0.59, 0.77) and 0.70 (0.64, 0.76), respectively; these values for female donors were 0.96 (0.82, 1.03) and 0.89 (0.82, 0.96), respectively. The prior and posterior means (95% CI) of δ for male donors were 3.00 (2.38, 3.62) and 2.98 (2.39, 3.60), respectively; these values for female donors were 3.00 (2.38, 3.62) and 2.98 (2.32, 3.58), respectively. These results indicate that there is little information in the data to estimate these parameters, so that the posterior results are

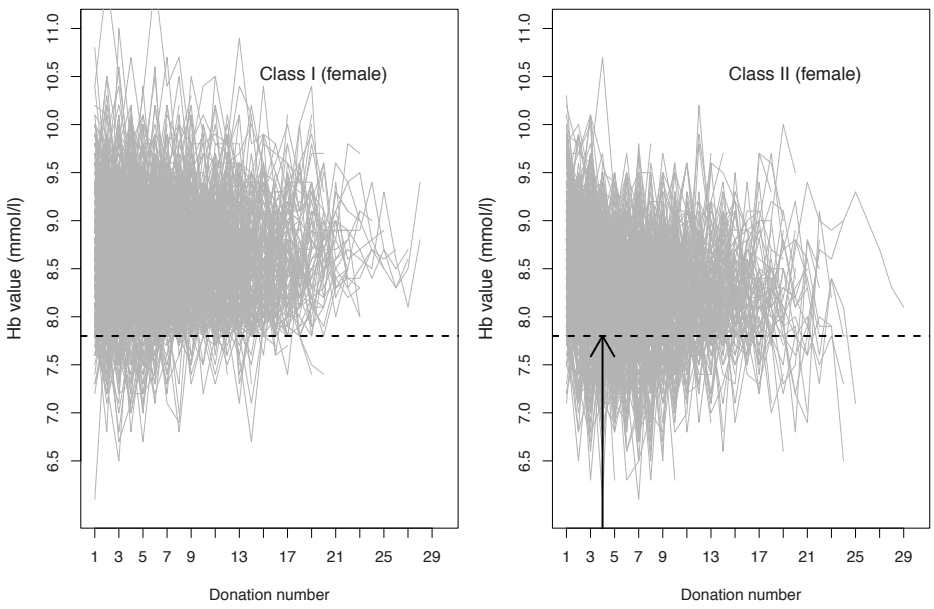


Figure 4.5: Female Hb profiles after screening visit separately for each class obtained from Model 3 (and Model 3_0). The donors assigned to the highest class probability. The horizontal lines show the threshold for donation eligibility. The arrow shows the recovery time change point.

determined by the informative priors.

The shape parameter ψ is estimated greater than one in both sexes, which means that the estimated recovery process is a concave curve. That is, the Hb recovery at the beginning is fast and becomes slower over time. This corresponds to the function exhibited in Figure 4.3. Finally, the Pearson correlation of initial Hb values and random intercepts in the main model is 0.67 (0.64, 0.69) and 0.61 (0.58, 0.63) for male and female donors respectively, which shows the importance of the ICP.

To compare the prediction accuracy between different models, we used the results of the training data set to predict the Hb values in the validation data set. Figure 4.6 shows the mean square error (MSE) values during successive donations across these donors for different models. The graph illustrates the superiority of Model 3 compared with Model 2 and Model 1. The prior sensitivity analyses showed that the posterior results are stable with respect to reasonable choices for the non-informative priors (e.g., a truncated normal distribution with mean zero and standard deviation of 1000 for the recovery times yielded similar results). We also checked the sensitivity of the posterior results by using a less informative prior for θ (i.e., prior standard deviation increased by a factor 4) and δ (i.e., prior standard deviation increased by a factor 10). Since there is little information in the data about δ and θ, using less informative priors for these parameters, different results might be

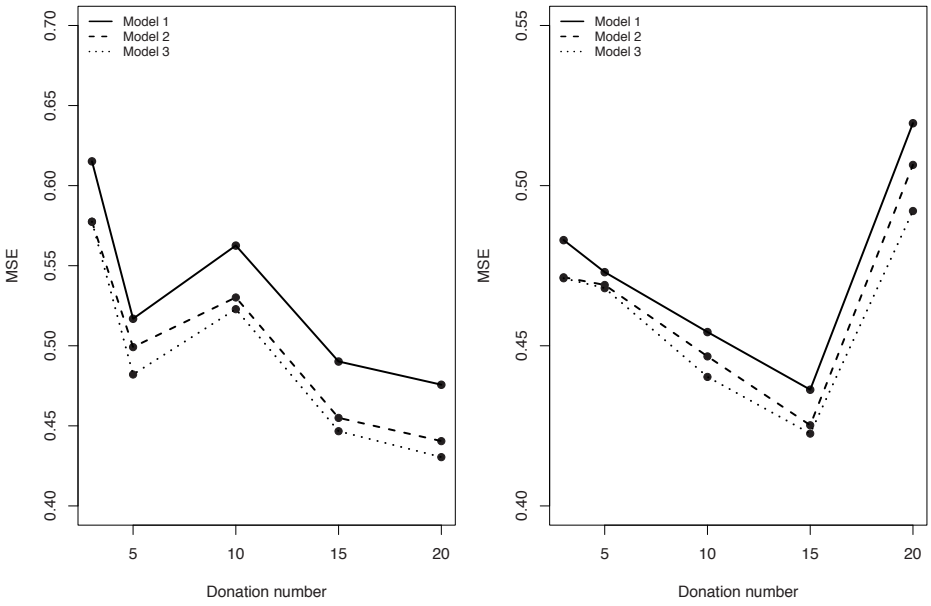


Figure 4.6: The mean square error (MSE) values for different models at successive donation numbers for male and female donors (male donors in the left panel, female donors in the right panel) in the validation data set.

gotten for parameter estimates of the recovery function. This would have been treated as a weakness of the analysis, but this is actually the strength of a Bayesian analysis. Namely it is the approach that allows to include such information in an elegant manner. Detailed results of these sensitivity analyses are presented in Table 4.6 and Table 4.7.

Simulation results

The results of Model 3 (and Model 3_0) for 20 artificially generated data sets of male donors are presented in Table 4.8. This table shows the true parameter values used to generate these data, the overall mean (over all 20 simulations) of the posterior medians of the parameters, and the mean (over all simulations) of the corresponding exceedance probabilities. Extreme values for the exceedance probability (i.e., <0.05 or >0.95) indicate a significant difference. These simulation results indicate that the proposed model is able to estimate the true parameters without bias. The percentage of subjects correctly assigned to their true class was estimated to be 67% in this simulation study.

Table 4.6: The posterior medians and 95% HPD CIs based on Model 3 and Model 3_0 for male and female donors, separately. These results are based on less informative priors for δ and θ, i.e., $\delta \sim N(\mu = 3, \sigma = 3.2)I(0,)$ (truncated) and $\theta \sim N(\mu = 0.68, \sigma = 0.18)$ for males and $\theta \sim N(\mu = 0.96, \sigma = 0.15)$ for females.

	Male donors			Female donors		
Parameter	Estimate	95% CI		Estimate	95% CI	
β_0	7.85	7.61	8.07	6.97	6.79	7.15
β_1	-0.004	-0.006	-0.002	0.003	0.002	0.005
β_2	-0.082	-0.100	-0.064	-0.068	-0.084	-0.053
γ	0.19	0.17	0.21	0.17	0.15	0.19
θ	0.72	0.63	0.83	0.65	0.55	0.79
δ	3.52	0.20	9.17	3.09	0.18	8.72
ψ	2.09	1.40	3.05	1.68	1.01	2.88
κ	7	7	8	4	4	5
RT_{g_1}	93	57	123	54	10	113
$\text{RT}_{g_{21}}$	113	83	147	278	207	399
$\text{RT}_{g_{22}}$	380	269	530	427	338	594
σ_{b1}	0.34	0.31	0.35	0.29	0.28	0.31
σ	0.46	0.45	0.47	0.44	0.43	0.45

4.5 Predicting future Hb values

As mentioned in the introduction, the ultimate aim of this study is to improve the planning of the donors' visits to the blood bank. Predicting the Hb level after donation and forecasting the appropriate time for inviting again the prospective donor for the next donation may improve the planning of the donors' visits to the blood bank. Prediction of a future observation ($\text{Hb}_{i_{\text{new}_{t-1}}}$) is based upon the chosen model and the estimated parameters via the current data (Data_{ref}) together with any available observations ($\text{Hb}_{i_{\text{new}_{t-1}}}, \ldots, \text{Hb}_{i_{\text{new}_1}}$) for the prospective donor. The prediction consists of two stages. The first stage is to determine the class membership of the prospective donor. The class membership probability given all available information for a donor can be computed by applying Bayes's theorem:

$$P(g_i = k|\text{Data}_{\text{ref}}, \text{Hb}_{i_{\text{new}_{t-1}}}, \ldots, \text{Hb}_{i_{\text{new}_1}}) = \frac{P(\text{Hb}_{i_{\text{new}_{t-1}}}, \ldots, Hb_{i_{new_1}}|\text{Data}_{\text{ref}}, g_i = k)P(g_i = k)}{\sum_{k=1}^{2} P(\text{Hb}_{i_{\text{new}_{t-1}}}, \ldots, \text{Hb}_{i_{\text{new}_1}}|\text{Data}_{\text{ref}}, g_i = k)P(g_i = k)},$$

where $P(g_i = k|\text{Data}_{\text{ref}}, Hb_{i_{\text{new}_{t-1}}}, \ldots, \text{Hb}_{i_{\text{new}_1}})$ is the marginal probability that donor i belongs to the kth class given the history of Hb values and the other covariates for that

Table 4.7: The posterior medians and 95% HPD CIs based on Model 3 and Model 3_0 for male and female donors, separately. These results are based on a truncated normal distribution with mean zero and σ equals to 10^3 for RT_{g_1}, $\text{RT}_{g_{21}}$, and $\text{RT}_{g_{22}}$.

	Male donors			Female donors		
Parameter	Estimate	95% CI		Estimate	95% CI	
β_0	7.85	7.63	8.08	6.96	6.78	7.14
β_1	-0.004	-0.006	-0.002	0.003	0.002	0.005
β_2	-0.082	-0.100	-0.064	-0.068	-0.084	-0.053
γ	0.19	0.17	0.21	0.17	0.15	0.19
θ	0.70	0.63	0.76	0.89	0.82	0.96
δ	3.00	2.38	3.62	2.97	2.36	3.62
ψ	2.10	1.31	3.58	3.02	2.23	4.44
κ	7	7	8	4	4	5
RT_{g_1}	96	58	143	58	15	122
$\text{RT}_{g_{21}}$	114	83	167	346	275	485
$\text{RT}_{g_{22}}$	395	257	617	508	395	690
σ_{b1}	0.34	0.31	0.35	0.29	0.28	0.31
σ	0.46	0.45	0.47	0.44	0.43	0.45

Table 4.8: Results of Model 3 (and Model 3_0) for 20 artificially generated data sets of male donors

Parameter	β_0	β_1	β_2	γ	θ	δ	ψ	RT_{g_1}	$\text{RT}_{g_{21}}$	$\text{RT}_{g_{22}}$	Class I size
True simulated value	7.87	-0.004	-0.082	0.19	0.70	2.98	2.18	100	117	419	54%
Estimated value	7.99	-0.004	-0.082	0.18	0.67	3.00	1.91	114	122	371	52%
Exceedance probability	0.65	0.51	0.47	0.38	0.25	0.52	0.38	0.46	0.50	0.32	

donor (Komárek et al., 2010). $P(\text{Hb}_{i_{\text{new}_{t-1}}}, \ldots, \text{Hb}_{i_{\text{new}_1}} | \text{Data}_{\text{ref}}, g_i = k)$ is the density of Hb values for this donor given that the donor belongs to class k. At this stage we assign the donor to the class with the highest probability.

The second stage, which is predicting the random intercept for the donor $b_{i_{\text{new}}}$, also can be performed by applying Bayes's theorem:

$$P(b_{i_{\text{new}}} | \text{Data}_{\text{ref}}, \text{Hb}_{i_{\text{new}_{t-1}}}, \ldots, \text{Hb}_{i_{\text{new}_1}}, g_i) \propto$$
$$P(\text{Hb}_{i_{\text{new}_{t-1}}}, \ldots, \text{Hb}_{i_{\text{new}_1}} | \text{Data}_{\text{ref}}, b_{i_{\text{new}}}, g_i) P(b_{i_{\text{new}}}). \quad (4.8)$$

Since the model is linear in the random effects, the posterior distribution for the random effects has the following closed-form expression (Harville, 1976):

$$b_{i_{\text{new}}} | \text{Data}_{\text{ref}}, \text{Hb}_{i_{\text{new}_{t-1}}}, \ldots, \text{Hb}_{i_{\text{new}_1}},$$

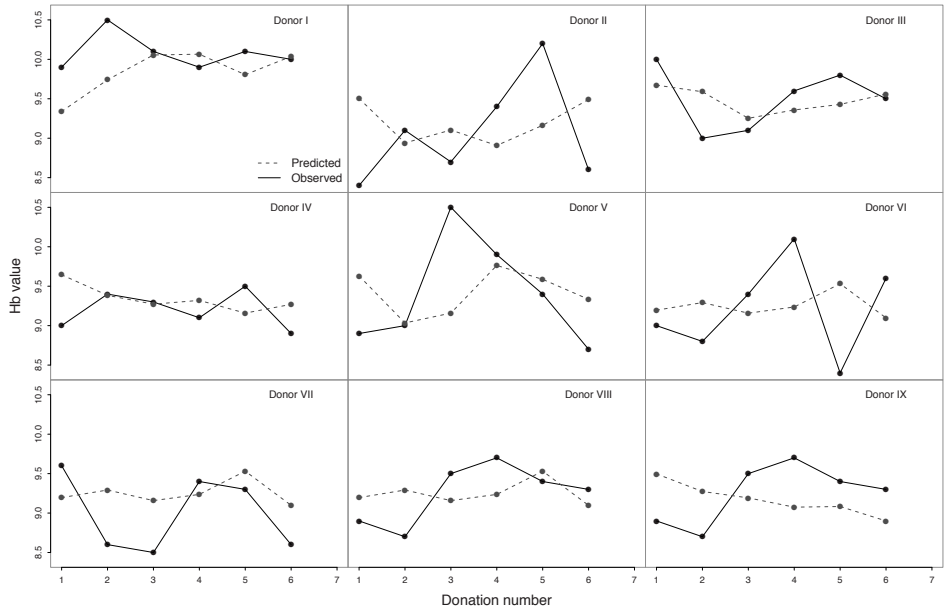


Figure 4.7: Predicted (from Model 3 (and Model 3_0)) versus observed Hb values for 9 randomly chosen male donors. To predict the Hb value at each time point the information up to this point was used.

and,

$$g_i \sim N(\sigma_{b1}^2 Z_i^T V_i^{-1}(Hb_i - X_i\beta), \sigma_{b1}^2 Z_i^T K_i Z_i \sigma_{b1}^2),$$

where $K_i = V_i^{-1} - V_i^{-1} X_i \left(\sum_{i=1}^N X_i^T V_i^{-1} X_i\right)^{-1} X_i^T V_i^{-1}$, $V_i = \sigma_{b1}^2 1_{T_i} + \sigma^2 I_{T_i}$, and Z_i is a vector of 1s of length T_i. Finally, to compute the predicted Hb value, one can apply the equation for the corresponding model, with b_{i1} estimated as the mean of the posterior distribution in (4.8), and with the other parameters obtained from Data_{ref}. This is a dynamic prediction in the sense that it can be updated as soon as information from subsequent donations becomes available. We predicted the first six Hb values for nine randomly chosen male donors in Figure 4.7. The predicted Hb values were computed dynamically. That is, at each time point the information up to this point was used to predict the subsequent Hb value.

4.6 Conclusion

In this study, we have considered the prediction of a future Hb value for a potential blood donor given the previous observations, and the estimation of the recovery time after a donation. The prediction of a donor's Hb value is complicated due to a) heterogeneity of the initial Hb value, b) state dependence of a donor's Hb values, c) varying time intervals

between donations, d) the temporary reduction in Hb after blood donation, and e) the fact that the recovery process may change with the number of donations and may differ between donors. To account for these aspects of the data, we developed a mixed-effects model (Model 1), a latent class mixed-effects model (Model 2), and a latent class mixed-effects transition model (Model 3). The advantage of the mixed-effects transition model is that it simultaneously captures heterogeneity and state dependence. In all these models, the temporary reduction of Hb after donation was modeled using a flexible function. In the models with latent classes, the heterogeneity in the recovery process was controlled using latent classes and the dynamics of the recovery process using a change point model. The latent class mixed-effects transition model was preferred over the simpler models according to the DIC and based on an evaluation of model fit. This finding shows that it is important to account for both unobserved individual heterogeneity and state dependence among the Hb values for an individual.

The flexible function (4.2) enables us to estimate the recovery time, which is the time needed for Hb to return to its pre-donation value. In the model, the estimated pre-donation Hb value depends on the Hb value at the previous donation via the transition effect. If the value of the transition effect γ is small, the recovery time can be interpreted as approximately the time needed to return to the original Hb value, before the person started donating.

The model results show that the estimated recovery time is considerably longer than the mandatory interval between donations (i.e., 56 days). Also, our findings point to a concave Hb recovery process. That is, the recovery process is fastest at the beginning and becomes slower over time. The estimated recovery time should be seen as the ultimate recovery time, i.e., the time by which a donor's Hb value has fully recovered. Due to the concave shape of the recovery process, most of the recovery occurs before half of the recovery time has passed, which partially explains the long estimated recovery times in our data set. Furthermore, it should be noted that a recovery time that is longer than the average interval between donations is in line with the observed data, as there is a decline in the Hb trajectories with the number of donations.

Another interesting finding is that there is heterogeneity between donors in the recovery time, i.e., 54.3% and 51.3% of male and female donors have a constant recovery time during successive donations. The remaining donors have a longer recovery time and their recovery time increases after a number of donations. This increase in recovery time might be attributed to a reduction of the iron reserves in these donors. In a previous study, a relatively faster Hb recovery was observed in donors with high pre-donation iron reserves (Kiss et al., 2015), so these results require further investigation.

Our models also showed that male donors on average have a shorter recovery process than female donors. This findings is consistent with previous studies (Custer et al., 2014; Wadsworth, 1955). The effect of age was estimated negative for male donors and positive for female donors, which is again consistent with previous studies (Nasserinejad et al.,

2013; Baart et al., 2011) and can be explained by the effects of menopause: women stop losing iron after menstruation (Baart et al., 2011). The results also showed that the Hb value is lower on average in warm seasons than in cold seasons, see also Baart et al. (2011); Hoekstra et al. (2007) for other evidence of this finding.

The model comparison (using DIC) hints that the latent class mixed-effects model with random intercepts is not sufficient to capture the entire within-subject correlation structure. An alternative approach would be to use a more complicated random-effects model, e.g., using a random slope of the time since the first visit or the number of donations. Although adding additional random effects to the model may improve the fit, the resulting parameters may be hard to interpret (Funatogawa and Funatogawa, 2012). In our study, assuming that the time since the first donation or the number of donations affects the Hb values would imply that a part of the reduction in Hb value is permanent. This assumption would not be realistic from a clinical perspective, therefore we did not include such random effects in the model. Also, adding random or fixed effects of time since the first donation or the number of donations to the model would have affected the estimated recovery time. Namely, the recovery time could no longer be interpreted as the time needed for Hb to return to its original value. Finally, yet another alternative model could be a random effects model using an autoregressive structure in the residuals instead of in the mean structure, which is not affected by previous covariates.

The transition model is not very often used in medical applications. One of the reasons is that the associated covariance matrix is more restrictive than for the mixed-effects model (Funatogawa and Funatogawa, 2012). Furthermore, also transition models with random effects are not really popular in the medical area. The fact that the ICP condition must be dealt with implies that standard estimation techniques cannot be applied. To handle the ICP in the latent class mixed-effect transition model, we used here a reduced form equation for the initial period similar to the dynamic equation, but excluding the lag effect from the model. We let the model take into account the correction between the unobserved individual effects, b_i, and the initial state Hb_{i1}.

Although we designed our model to be very flexible, it is only one out of many possible models. For instance, our model could be improved by incorporating more covariates that can affect Hb value such as physical activity (Beard and Tobin, 2000), race (Johnson-Spear and Yip, 1994), nutrition (Brussaard et al., 1997), BMI (Skjelbakken et al., 2006), and smoking status (Kristal-Boneh et al., 1997). However, due to a lack of information, we could not incorporate them into our model. Class membership could be modeled to depend on some time-independent covariates such as genetic information of donors. This is the aim for a subsequent paper, once this information is available.

In conclusion, we developed a statistical modeling approach that allows to classify donors in two subgroups based on their Hb recovery time. This is of high practical importance because identification of the class for a donor could improve the planning of donors' visits to the blood banks and help to tailor donation intervals and prevent iron deficiency and

donor deferrals.
Furthermore, our results support a donation interval longer than 56 days for both sexes, which has also been recommended in previous literature (Rikhtehgaran et al., 2012; Kiss et al., 2015; Brittenham, 2011; Simon et al., 1981). The U.S. Food and Drug Administration (FDA) is currently considering revising this interval to better protect donors (Brittenham, 2011). For the candidates belonging to the group with non-constant recovery time, we suggest appropriate interventions (e.g., postponing the next invitation or using an iron-rich diet or taking over-the-counter vitamin supplements that contain iron or specific iron supplements) to prevent rejection at the next visit.

4.7 Supplementary

Jags codes to implement a latent class mixed-effects model with 4 classes. The first class is restricted to be stable regarding the number of donations in the last two years. In this syntax

```
model {

for(j in 1:n) { ### Begin j loop

    for(i in offset[j]: (offset[j+1]-1)){ ### Begin i loop

        Hb[i]~dnorm(mu[i], tau)

        mu[i]<-b[j]+beta[1]+beta[2]*Age[j]+beta[3]*Season[i]-
        Indicator[i]*theta*
        pow((max(RT[g[j],Time[i]]-(TSPD[i]-delta),0)/
                RT[g[j],Time[i]]),psi)
        } ### End i loop

### Time[i] is number of donations in this visit
### Indicator[i] is a binary variable (visit till 1st donation=0)
 ### Latent class indicator
    g[j] ~ dcat(dsi[j,])
    dsi[j, 1:2] ~ ddirch(alpha[])
    } ### End j loop
####
for(i in 1:n){
b[i]~dnorm(0, tau0)
}

for(i in 1:TT){
### 1st class, class with stable recovery time
    RT[1,i]<-RT11
### 2nd class
    RT[2,i]<-(1-step(i-ChangeTime))*RT21+step(i-ChangeTime)*RT22
}
### Priors
for (n in 1:20) {
    prior[n] <- 1/20
```

```
}

for (n1 in 1:2) {
    alpha[n1] <- 4
}

ChangeTime~dcat(prior[])
 for (k in 1:3)
 {
     beta[k] ~ dnorm(0, 1.E-3)
     beta0[k] ~ dnorm(0, 1.E-3)
 }

### Recovery time in 1st class
RT11~dunif(0,1.E+3)

### Recovery time before change point time in 2nd class
RT21~dunif(0,1.E+3)

### Recovery time after change point time in 2nd class
RT22~dunif(0,1.E+3)

tau~dgamma(1.E-2 , 1.E-2)

tau0~dgamma(1.E-2 , 1.E-2)

theta~dnorm(0.67645,493.07)

psi<-exp(lgpsi)

lgpsi~dnorm(0,10)

delta~dnorm(3,10)

} ### End
```

5 Comparison of criteria for choosing the number of classes in Bayesian finite mixtures

This chapter is submitted as: Nasserinejad K, van Rosmalen J, de Kort W, and Lesaffre E. Comparison of criteria for choosing the number of classes in Bayesian finite mixtures. 2016.

Abstract.

IDENTIFYING the number of classes in Bayesian finite mixture models is a challenging problem. Several criteria have been proposed, such as adaptations of the deviance information criterion, marginal likelihoods, Bayes factors, and reversible jump MCMC techniques. It was recently shown that in overfitted mixture models, the overfitted latent classes will asymptotically become empty under specific conditions for the prior of the class proportions. This result may be used to construct a criterion for finding the true number of latent classes, based on the removal of latent classes that have negligible proportions. Unlike some alternative criteria, this criterion can easily be implemented in complex statistical models such as latent class mixed-effects models and multivariate mixture models using standard Bayesian software. We performed an extensive simulation study to develop practical guidelines to determine the appropriate number of latent classes based on the posterior distribution of the class proportions, and to compare this criterion with alternative criteria. The performance of the proposed criterion is illustrated using a data set of repeatedly measured hemoglobin values of blood donors.

5.1 Introduction

Finite mixture models can be used to capture unobserved heterogeneity in the population by assuming that the population consists of K homogeneous subgroups. These models also allow to represent non-standard distributions by an appropriate mixture of standard distributions. However, identifying the number of latent classes (K) remains a challenging problem (Lee et al., 2008; McGrory and Titterington, 2007; Richardson and Green, 1997; Rousseau and Mengersen, 2011). Several criteria exist for choosing the number of latent classes in mixture models in both the frequentist and the Bayesian setting. Whereas information criteria such as the Akaike information criterion (AIC) (Akaike, 1973) and the Bayesian information criterion (BIC) (Schwarz et al., 1978) seem to be the most popular criteria in a frequentist setting (Steele and Raftery, 2010; Pan and Huang, 2014; Keribin, 2000), no clear consensus on the optimal criterion in a Bayesian setting has yet emerged.
Although the deviance information criterion (DIC) (Spiegelhalter et al., 2002) is a well-established criterion for comparing different Bayesian models, unfortunately this criterion is not suited to the case of mixture models (Steele and Raftery, 2010). Several adaptations of this criterion to mixture models have been proposed (Celeux et al., 2006). Alternatively, models with different numbers of latent classes can be compared by computing marginal likelihoods, Bayes factors, or by using reversible jump Markov chain Monte Carlo (RJMCMC) techniques (Green, 1995).
The appropriate number of latent classes is obtained by optimizing one of the criteria by fitting several mixture models with different numbers of classes. However, this procedure is often not easy to apply, as estimating a finite mixture model for different numbers of classes can be time consuming. Furthermore, some of these criteria cannot be calculated using standard software for Bayesian analyses such as WinBUGS, JAGS, or Stan, so that the researcher often has to compute the criteria outside these software packages. RJMCMC sampling is another approach with its own drawbacks. In this algorithm the Markov chain moves between mixture models with different numbers of classes based on carefully selected proposal densities (Frühwirth-Schnatter, 2004; Dellaportas and Papageorgiou, 2006). It can be difficult to derive appropriate proposal densities, especially for complex hierarchical models. Alternative choices such as marginal likelihood approaches, which are generally not available in closed form in mixture models, also yield challenging numerical issues even for mixture models with a moderate number of classes (Frühwirth-Schnatter, 2004).
Rousseau and Mengersen (2011) (hereafter R&M) showed that in overfitted mixture models (i.e., a mixture model fitted with more latent classes than present in the data), the superfluous latent classes will asymptotically become empty if the Dirichlet prior on the class proportions is sufficiently uninformative. Rousseau and Mengersen (2011) indicated that their result may lead to a criterion for finding the true number of latent classes by simply excluding latent classes that are negligible in proportion. A subsequent study by

Malsiner-Walli et al. (2016) proposed a specific implementation of this criterion, and used simulated data to investigate its performance in finding the true number of latent classes. In their implementation, the mixture model is first estimated with a relatively large number of latent classes. The true number of latent classes is then estimated as the mode of the number of non-empty classes, where a class is defined as empty if no subject is assigned to it in a specific MCMC iteration.
The advantage of R&M criterion is that it is simple to implement using standard Bayesian software, even for complex statistical models, because the latent class proportions are an automatic byproduct of the estimation.
In this study, we use a criterion that resembles the criterion used by Malsiner-Walli et al. (2016). However, we relax the rather conservative criterion used by Malsiner-Walli et al. (2016) that a class is only empty if it contains zero observations, and instead assess the effects of different cut-offs for the proportions in a class. This is more logical, because Rousseau and Mengersen (2011) only showed that the class proportions converge to 0 if the sample size approaches infinity, not that they should be 0 with any data set of finite size. The simulation study of Malsiner-Walli et al. (2016) only used data sets with well-separated latent classes and did not compare the criterion with alternative methods for choosing the number of latent classes. In our simulation study, we considered various scenarios with different degrees of separation between latent classes as well as longitudinal data, to assess how this criterion performs in a more realistic setting. We also compared the R&M criterion with alternative criteria for estimating the number of latent classes.
We show that both the prior for the class-specific parameters as well as the hyperparameter of the Dirichlet prior distribution for the class proportions have to be chosen carefully to ensure a good performance of this method. We use the simulation results to provide recommended settings, and apply these settings in the analysis of longitudinal hemoglobin (Hb) values of blood donors.
In the next section, background on finite mixture models is presented including a discussion of priors for mixture models. In Section 5.3, methods for choosing the appropriate number of classes in this study are presented. Section 5.4 deals with the simulation study in both a univariate and a longitudinal setting. In Section 5.5, a practical example of longitudinal mixture modeling is presented. Finally, the results are discussed and practical recommendations are given in Section 5.6.

5.2 Background on finite mixture models

Definition of mixture models

A finite mixture model is defined as:

$$f(y|\boldsymbol{\lambda}, \theta, \gamma) = \sum_{j=1}^{K} \lambda_j f_j(y|\theta_j, \gamma), \tag{5.1}$$

where $f(y|\boldsymbol{\lambda}, \theta, \gamma)$ is the density of the observed data, $f_j(y|\theta_j, \gamma)$ is the density of the observed data in latent class j, K is the true number of latent classes and the vector $\boldsymbol{\lambda}$ represents the class proportions, which are non-negative and sum to 1. θ_j is a vector of parameters for the distribution of the data in class j, and γ is a vector of parameters common to all classes. The observed data y can be either univariate or multivariate, and $f_j(y|\theta, \gamma)$ may correspond to e.g., a simple Gaussian model or a complex hierarchical model.
Since we use a Bayesian setting, priors need to be chosen for λ_j, θ_j ($j = 1, \ldots, K$), and γ. A challenging issue that arises in Bayesian mixture models is the nonidentifiability of the latent classes. The problem is caused by the invariance of the posterior distribution with respect to permutations of class labeling under symmetric priors and likelihood (Dellaportas and Papageorgiou, 2006). This leads to so-called label switching in the MCMC output, and the posterior distributions of class-specific parameters θ_j will be identical and thus useless for inference (Jasra et al., 2005).

Priors for mixture models

If no relevant prior information for the parameters is available, many researchers prefer to use noninformative (vague) prior distributions whose impact on the posterior distribution of the model parameters is minimal. The most commonly used prior for the class proportions, λ_j, is a symmetric Dirichlet distribution, i.e., $\boldsymbol{\lambda}|K \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_K)$, and $\alpha_k = \alpha$ for $k = 1, \ldots, K$. Smaller values of α correspond with a less informative prior. A flat prior distribution is obtained with $\alpha = 1$, whereas setting $\alpha = 0$ leads to an improper Dirichlet distribution, and also to an improper posterior result.
The choice of α is important, as its value can strongly affect the posterior results. Although large values of α lead to informative prior distributions, some researchers have suggested to use values larger than 1 (e.g., $\alpha = 4$ or $\alpha = 10$) to avoid solutions with empty classes (Asparouhov and Muthén, 2011). When using the marginal likelihood as a criterion (i.e., choosing the number of latent classes that yields the highest value of the marginal likelihood), it has been shown that more informative Dirichlet distributions lead to a lower probability of overestimating the number of latent classes in the data (Nobile, 2004; Frühwirth-Schnatter, 2006).
In contrast, Rousseau and Mengersen (2011) have suggested to use smaller values of α,

with $\alpha < d/2$, where d is the number of class-specific parameters, i.e., θ_j. This recommendation is based on a mathematical proof showing that with a sufficiently uninformative Dirichlet prior distribution, the proportions of overfitted latent classes will converge to zero as the sample size increases. For α greater than $d/2$, the class proportions of overfitted classes will asymptotically converge to nonnegligible values, even if the data are homogeneous. This is one of the few examples in Bayesian statistics where less informative priors lead to better results (Rousseau and Mengersen, 2011), as the more informative Dirichlet distributions will overestimate the number of latent classes. Rousseau and Mengersen (2011) further argued that with $\alpha < d/2$, the posterior distribution of the class proportions has a much more stable behavior than the maximum likelihood estimator. Another disadvantage of using informative Dirichlet priors is that the posterior distributions of the class proportions may be biased, especially in small data sets.

An alternative approach to fixing α in advance would be to let the data determine the optimal value for alpha, which means to use a hyperprior specification for α, so that α is an unknown parameter that is estimated using the data. The prior for α could for example be a gamma prior, with $\alpha \sim \Gamma(\epsilon_1, \epsilon_2)$ (Ishwaran et al., 2001; Malsiner-Walli et al., 2016), where ϵ_1 and ϵ_2 are the shape and rate parameters of the gamma distribution, respectively. Priors must also be chosen for the class-specific parameters θ_j. In many cases, there is no relevant prior information available for the class-specific parameters, so that the use of noninformative priors seems appropriate. However, it is generally not possible to use improper priors for the class-specific parameters in finite mixture models, because there is a nonzero posterior probability that at least one of the classes is empty, leading to improper posteriors for the class-specific parameters (Wasserman, 2000). Instead one can use minimally informative but diffuse proper priors which lead to diffuse posterior distributions of the class-specific parameters, but the posterior results may be sensitive to the spread of the prior (Wasserman, 2000).

Data-dependent priors, which are prior distributions that are a function of the observed data, have been proposed instead (Richardson and Green, 1997; Wasserman, 2000; Raftery et al., 1996). Wasserman (2000) showed that these prior distributions may have better frequentist properties.

Another approach would be to use a hierarchical prior. For example in a mixture model one can specify a hierarchical prior for the class-specific means as $\mu_k|b_0 \sim N(b_0, B_0)$, where $b_0 \sim N(m_0, M_0)$. The aim of these hierarchical priors is to minimize the impact of the prior on the posterior.

In many finite mixture models the distribution within each class is assumed to be normal, conditional on the observed covariates. Different priors have been proposed in the literature for the class-specific parameters, namely the priors proposed by Nobile and Fearnside (2007), and the normal-gamma prior (Griffin et al., 2010) for class-specific means used in Malsiner-Walli et al. (2016) combined with the approach of Rousseau and Mengersen (2011). Previous literature showed that the choice of prior has a strong effect

on choosing the number of latent classes in mixture models (Jasra et al., 2005).

5.3 Methods for choosing the number of classes

Various approaches have been proposed in the literature for choosing the number of latent classes in mixture models, in both frequentist and Bayesian settings. However, no consensus has emerged regarding which of these methods performs best. In this study we compare a number of well-known Bayesian approaches for choosing the number of latent classes in mixture models. These approaches are described below.

Deviance information criterion (DIC)

The deviance information criterion (DIC) is a well-known Bayesian criterion for the assessment and comparison of different Bayesian models (Spiegelhalter et al., 2002). The DIC involves a trade-off between goodness of fit (deviance) and model complexity (the effective number of parameters $\mathrm{p_D}$), and can be calculated as follows:

$$\mathrm{D}(\theta) = -2 \log f(y|\theta) + 2 \log h(y),$$

where $h(y)$ is a standardizing term that is a function of the data alone. Then the estimated effective number of parameters is defined as:

$$\mathrm{p_D} = \overline{\mathrm{D}(\theta)} - \mathrm{D}(\hat{\theta}),$$

where $\overline{\mathrm{D}(\theta)}$ is the posterior mean deviance and $\hat{\theta} = \mathbb{E}[\theta|y]$ is the posterior mean of the model parameters. DIC is then defined as:

$$\mathrm{DIC} = -4\mathbb{E}_\theta[\log f(y|\theta)|y] + 2 \log f(y|\hat{\theta}), \tag{5.2}$$

$\hat{\theta} = \mathbb{E}[\theta|y]$ ensures that $\mathrm{p_D}$ is positive when the density is log-concave in θ, but it is not appropriate for discrete parameters θ (Spiegelhalter et al., 2002; Celeux et al., 2006). In mixture models, the parameters θ are not identifiable if the prior and likelihood are invariant with respect to the labeling of classes. Therefore, $\hat{\theta} = \mathbb{E}[\theta|y]$ can be a very poor estimator and $\mathrm{p_D}$ may become negative (Celeux et al., 2006). A more relevant choice for $\hat{\theta}$ would be the mode of the posterior distribution (Celeux et al., 2006). Several adaptations of this criterion were proposed by Celeux et al. (2006), such as DIC_3 and DIC_4. Namely,

$$\mathrm{DIC}_3 = -4\mathbb{E}_\theta[\log f(y|\theta)|y] + 2 \log \hat{f}(y), \tag{5.3}$$

where $\hat{f}(y) = \prod_{i=1}^{n} \hat{f}(y_i)$, $\hat{f}(y_i) = \frac{1}{M} \sum_{m=1}^{M} \sum_{j=1}^{K} \lambda_j^m f_j(y|\theta_j^m)$, M denotes the number of MCMC iterations, λ_j^m and θ_j^m are the results of the mth MCMC iteration, and

$$\mathrm{DIC}_4 = -4\mathbb{E}_{\theta,Z}[\log f(y, Z|\theta)|y] + 2\mathbb{E}_Z[\log f(y, Z|\mathbb{E}_\theta[\theta|y, Z]|y], \tag{5.4}$$

where $Z = (z_1, \ldots, z_n)$ is the class assignment vector of observations (individuals). To compute DIC_4, it is necessary to calculate the posterior expectation for each possible value of z. Among various DICs studied Celeux et al. (2006), these two DICs were found to be the most reliable criteria by the authors (Celeux et al., 2006).

Reversible jump MCMC algorithm

Another fully Bayesian approach is the reversible jump MCMC algorithm (RJMCMC), as introduced by Green (1995), which is an extension of the standard MCMC. RJMCMC allows sampling of the posterior distribution on spaces of varying dimensions. In this algorithm the Markov chain moves between finite mixture models with different number of classes based on carefully selected degenerated proposal densities, but which are in general not easy to design (Brooks et al., 2003; Dellaportas and Papageorgiou, 2006).

Rousseau and Mengersen's criterion

Rousseau and Mengersen (2011) proved that the posterior behavior of an overfitted mixture model depends on the chosen prior on the proportions λ_j. They showed that an overfitted mixture model converges to the true mixture, if the Dirichlet-parameters α_j of the prior are smaller than $d/2$ (d is the dimension of the class-specific parameters). This result can be used to define a criterion for choosing the true number of latent classes in a mixture model. Basically, a deliberately overfitted mixture model with K_{max} ($K_{\text{max}} > K$) latent classes is fitted to the data. A sparse prior (Dirichlet distribution with $\alpha_j < \text{d}/2$) on the proportions is then assumed to empty the superfluous classes ($K_{\text{max}} - K$) during MCMC sampling.
Various criteria can be used for a class to be declared empty. For instance, one could declare a class empty if the number of observations assigned to that class is smaller than a certain proportion of the observations in the data set (e.g., ψ). In other words, the (assumed) true number of non-empty classes (K) could be computed in each MCMC iteration as:

$$K^{(m)} = K_{\text{max}} - \sum_{j=1}^{K_{\text{max}}} I\{\frac{N_j^{(m)}}{N} \leq \psi\}, \tag{5.5}$$

where $K^{(m)}$ is the number of non-empty classes in iteration m of MCMC sampling, $N_j^{(m)}$ is the number of observations allocated to class j at iteration m, N is the total number of observations and I denotes the indicator function. ψ can be set to a predefined value, e.g., 0, 0.01, 0.02, or 0.05. Then one can derive the number of non-empty classes based on the posterior mode of the number of non-empty classes based on all MCMC iterations.

Bayesian information criterion

The Bayesian information criterion (BIC) (Schwarz et al., 1978) is a well-known frequentist criterion, which has been shown to be consistent for choosing the number of latent classes in mixture models (Keribin, 2000). BIC is defined as follows:

$$\text{BIC} = -2[\log f(y|\hat{\theta})] + g \log(n), \tag{5.6}$$

where $\hat{\theta}$ is the maximum-likelihood estimate of the parameter θ, g is the number of free parameters in the model, and n is the number of observations in the data.

5.4 Simulation studies

To investigate the performance of the criterion proposed by R&M compared to other well-known approaches, we set up two simulation studies with different scenarios. The first simulation study is based on one-dimensional data, whereas the second simulation study uses longitudinal data.

Simulation study A: univariate Gaussian mixture

In this simulation study, we consider a univariate Gaussian mixture, i.e., a location-scale mixture of univariate normal distributions:

$$f(y_i|\boldsymbol{\lambda}, \mu, \sigma^2) = \sum_{j=1}^{K} \lambda_j N(y_i|\mu_j, \sigma_j^2), \tag{5.7}$$

where $f(y_i|\boldsymbol{\lambda}, \mu, \sigma^2)$ is the density of the observed data y_i ($i = 1, \ldots, n$), n is the number of independent observations, $N(y_i|\mu_j, \sigma_j^2)$ is the density of the normal distribution with mean μ_j and variance σ_j^2, K is the true number of latent classes and λ_j is the proportion of latent class j.
We simulate data from this model using $n = 500$ observations with different numbers of latent classes, and different degrees of separation (i.e., *low*, *moderate*, and *high* separation). Our definition of *low*, *moderate*, and *high* separation is somewhat subjective, and is based on the percentage of variation in the data that can be explained by the clustering structure, i.e., $\sigma^2_{E(Y|Z)}/\sigma^2_Y$ where σ^2_Y denotes the marginal variance of the data, Z is an indicator variable for the latent class, and $\sigma^2_{E(Y|Z)}$ is the between-class variance. We also assessed the degree of separation between latent classes using the overlapping coefficient (OVL), which is the overlapping area that is below the density functions (Inman and Bradley Jr, 1989). The following four scenarios were considered:

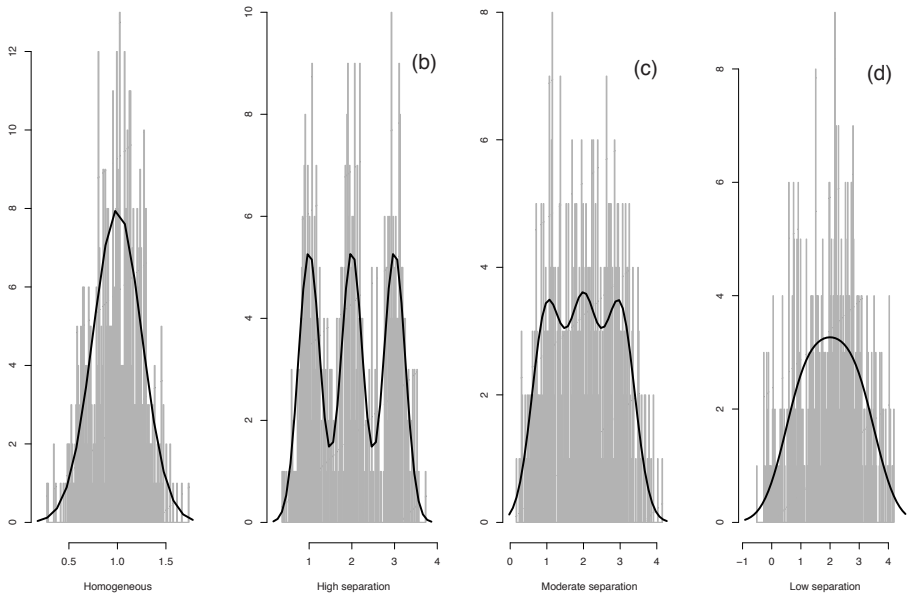


Figure 5.1: Univariate simulated data study: Histograms of randomly selected generated data sets. The solid lines represent the true marginal densities.

- **Scenario A1:** No clustering structure: $K = 1$ class with $\mu_1 = 1$ and $\sigma_1 = 0.25$, see Figure 5.1(a).

- **Scenario A2:** High separation ($\sigma^2_{E(Y|Z)}/\sigma^2_Y = 0.80$, OVL $= 0.06$): $K = 3$ classes with $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 3$, and $\sigma_1 = \sigma_2 = \sigma_3 = 0.25$, see Figure 5.1(b).

- **Scenario A3:** Moderate separation ($\sigma^2_{E(Y|Z)}/\sigma^2_Y = 0.70$, OVL $= 0.29$): $K = 3$ classes with $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 3$, and $\sigma_1 = \sigma_2 = \sigma_3 = 0.4$, see Figure 5.1(c).

- **Scenario A4:** Low separation ($\sigma^2_{E(Y|Z)}/\sigma^2_Y = 0.60$, OVL $= 0.74$): $K = 3$ classes with $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 3$, and $\sigma_1 = \sigma_2 = \sigma_3 = 0.7$, see Figure 5.1(d).

In the base-case analysis, the data are simulated using equal class proportions (i.e., $\lambda_j = 1/K$ for each class j). The histograms of these simulated data for a randomly selected data set are displayed in Figure 5.1, together with the true marginal densities. Separation decreases from Figure 5.1(b) to Figure 5.1(d), to end in a unimodal distribution. We implemented the criterion proposed by R&M, and we compared this criterion with the results of RJMCMC (Celeux et al., 2006), DIC_3 and DIC_4 Celeux et al. (2006), and BIC. To

establish whether a class is empty under the R&M criterion we used different values for the cut-off (ψ) i.e., 0, 0.01, 0.02, and 0.05 of observations in the sample, and the maximum number of latent classes was set to $K_{\text{max}} = 10$.
The prior for the class proportions λ was chosen to be a symmetric Dirichlet distribution with hyper-parameter equal to $\alpha = 0.00001$, 0.001, 0.01, 0.05, 0.1, 0.3, 0.5, 0.9.
For the priors of the class-specific means, we considered both a normal-gamma prior and a relatively uninformative prior. The relatively uninformative prior was $\mu_j \sim N(0, 31.6^2)$. The normal-gamma prior is a hierarchical data-dependent prior that places a normal prior on the prior mean and a shrinkage prior on the prior variance (Malsiner-Walli et al., 2016). This prior for a univariate mixture model can be defined as follows:

$$\mu_k|\lambda\,, b_0 \sim N(b_0, \eta\, R^2),$$

where $\eta \sim \Gamma(\nu_1, \nu_2)$ and $b_0 \sim N(m_0, M_0)$, m_0 and R are the median and range of the data, respectively. M_0^{-1} is set to 0. The hyper-parameters ν_1 and ν_2 are set to 0.5 to allow considerable shrinkage of the prior variance of class means (Malsiner-Walli et al., 2016).
For the priors of the class-specific variance, we also considered a hierarchical data-dependent prior and a relatively uninformative prior. The hierarchical data-dependent prior on the class-specific variances was implemented by Malsiner-Walli et al. (2016) in a multivariate mixture model, and is given by:

$$1/\sigma_k^2 \sim \Gamma(\beta_1 = 1.25, \beta_2 = 1/(2C_0)),$$

where $C_0 \sim \Gamma(\epsilon_1 = 0.25, \epsilon_2 = 20/R^2)$. The relatively uninformative prior on the class-specific variances was $\sigma_j^2 \sim U(0, 10)$.
We used a full factorial design to vary a) the number of latent classes and the degree of separation (using the four scenarios described above), b) the criterion for determining the number of latent classes (i.e., the R&M criterion with different cut-off values, RJMCMC, DIC_3, DIC_4, and BIC), and c) the value of α in the Dirichlet distribution (i.e., α =0.00001, 0.001, 0.01, 0.05, 0.1, 0.3, 0.5, or 0.9).
Three additional factors were varied in sensitivity analyses. In these sensitivity analyses, only the scenario with high separation between classes was simulated, but the other factors in the full factorial design were not fixed.
Two sensitivity analyses consisted of a) changing the sample size of the data set (i.e., 100 - 600 observations) and b) simulating data with unequal proportions of the latent classes, including one small class, using $\lambda_1 = 0.475$, $\lambda_2 = 0.475$, $\lambda_3 = 0.05$. Finally, we also performed a sensitivity analysis for the number of latent classes, with K ranging from $K = 1$ to $K = 6$, with $n = 100 \times K$ and means chosen as $\mu_j = j$ for $j = 1, \ldots, K$ and $\sigma_j = 0.25$ and also $\sigma_j = 0.40$.
We generated 50 data sets for each setting in the base-case analysis and the sensitivity analyses, except for the sensitivity analyses with varying number of classes, which used only 20 data sets. The low number of simulated data sets for these sensitivity analyses

was necessary to limit the total computation time. MCMC sampling is run for each data set for 50,000 iterations after discarding the first 5,000 iterations (burn-in). Computations were performed using the JAGS interface in R, Rmixmod, lcmm, and mixAK packages in R. To be able to compute DIC_3 and DIC_4, an MCMC sampler for the model parameters and the class assignments in the univariate mixture model was programmed in R.

Simulation study A: results

Table 5.1 shows the simulation results of Scenario A1. This table presents the success rate (the percentage of data sets in which the true number of clusters was obtained) of the different approaches, the mode of the estimated number of classes is presented in parentheses. The criterion of Rousseau and Mengersen (2011) is denoted as $R\&M^{NG}$ if the normal-gamma prior is used, and as $R\&M^{NI}$ if the relatively uninformative prior is used, with the cut-off value for defining a class to be empty as a subscript. For example, $R\&M^{NI}_{0.02}$ represents the Rousseau and Mengersen (2011) criterion with the relatively uninformative prior where $\psi = 0.02$.
In this scenario, the models cannot underestimate the number of classes. Small values for α for both a normal-gamma prior and a relatively uninformative prior in the R&M criterion result in a better estimation of the true number of latent classes. However, the R&M criterion with a normal-gamma prior requires much lower values of α (i.e., $\alpha < 0.1$) to obtain adequate results compared to the R&M criterion with the relatively uninformative prior, in which any value of α below 0.5 leads to good results. The other approaches (i.e., RJMCMC, DIC_3, and DIC_4) show better results with larger values for α.

Table 5.1: The results of Scenario A1. Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses.

α	RJMCMC	$R\&M^{NG}_{0}$	$R\&M^{NG}_{0.01}$	$R\&M^{NG}_{0.02}$	$R\&M^{NG}_{0.05}$	$R\&M^{NI}_{0}$	$R\&M^{NI}_{0.01}$	$R\&M^{NI}_{0.02}$	$R\&M^{NI}_{0.05}$	DIC_3	DIC_4
0.00001	68%(1)	100%(1)	100%(1)	100%(1)	100%(1)	100%(1)	100%(1)	100%(1)	100%(1)	0%(5)	0%(5)
0.001	18%(10)	100%(1)	100%(1)	100%(1)	100%(1)	100%(1)	100%(1)	100%(1)	100%(1)	0%(3)	0%(3)
0.01	28%(1)	98%(1)	98%(1)	98%(1)	98%(1)	100%(1)	100%(1)	100%(1)	100%(1)	98%(1)	20%(5)
0.05	90%(1)	22%(2)	80%(1)	84%(1)	92%(1)	100%(1)	100%(1)	100%(1)	100%(1)	96%(1)	72%(1)
0.1	98%(1)	2%(4)	10%(3)	18%(2)	40%(2)	100%(1)	100%(1)	100%(1)	100%(1)	96%(1)	100%(1)
0.3	98%(1)	0%(8)	0%(6)	0%(5)	0%(3)	98%(1)	100%(1)	100%(1)	100%(1)	94%(1)	100%(1)
0.5	98%(1)	0%(9)	0%(7)	0%(6)	0%(5)	96%(1)	98%(1)	98%(1)	100%(1)	94%(1)	100%(1)
0.9	98%(1)	0%(10)	0%(9)	0%(8)	0%(6)	96%(1)	98%(1)	98%(1)	100%(1)	90%(1)	100%(1)

The success rate of BIC using a frequentist approach was 100%.

Table 5.2 shows the simulation results of Scenario A2 (high separation), Scenario A3 (moderate separation), and Scenario A4 (low separation). In Scenario A2, small values for α (i.e., $\alpha < 0.05$) in the R&M criterion result in a perfect estimation of the true number of latent classes. The number of classes is overestimated by the R&M criterion with the normal-gamma prior for higher values of α. No such overestimation is observed for the

noninformative prior. Similar results were obtained in the sensitivity analysis for the number of latent classes (see Tables S1 and S2). In that sensitivity analysis, the normal-gamma prior yielded good results with values of $\alpha < 0.1$, but the noninformative only gave good results for larger values of α, with $\alpha > 0.05$. RJMCMC and DIC_3 gave the best results with larger values for α ($\alpha > 0.1$). The performance of DIC_4 does not seem to depend on the value of α, but it is not very good, with the probability of finding the true number of latent classes ranging from 50 to 70%.

When looking at Scenario A3 (moderate separation) and Scenario A4 (low separation), a different picture emerges. These results show that the R&M criterion may underestimate the true number of latent classes for low values of α. Namely, the R&M criterion with the normal-gamma prior underestimates the number of classes with low values of α and overestimates this number with high values of α. There is a narrow range around values of $\alpha = 0.05$ in which the performance of this criterion is good, and this range seems to depend on the cut-off for defining a class to be empty. On the other hand, the R&M criterion with the noninformative prior almost always underestimates the number of latent classes in Scenario A3 and Scenario A4. Underestimation rarely occurs with higher values of α, but a large value for α may result in overestimating the true number of latent classes. In Scenario A4, in which the distribution of the data looks unimodal, all approaches except $R\&M^{NI}_{0.02}$ and $R\&M^{NI}_{0.05}$ (for α=0.9) perform poorly, and most methods detect only a single class.

As a sensitivity analysis, we simulated a heterogeneous population with three unequal proportions, i.e., $\lambda_1 = 0.475$, $\lambda_2 = 0.475$, $\lambda_3 = 0.05$, $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 3$ and $\sigma_1 = \sigma_2 = \sigma_3 = 0.25$ (high separation), see Table 5.3 for the results. Here we performed the R&M criterion only with the noninformative prior. These results are consistent with the results of Scenario A2. The performance of the R&M criterion is quite good except for $R\&M^{NI}_{0.05}$ since the smallest class proportion is 5%, the cut-off defined for a class to be empty.

Table 5.2: The results of Scenario A2-A4. Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses.

Scenario	α	RJMCMC	R&M$_0^{NG}$	R&M$_{0.01}^{NG}$	R&M$_{0.02}^{NG}$	R&M$_{0.05}^{NG}$	R&M$_0^{NI}$	R&M$_{0.01}^{NI}$	R&M$_{0.02}^{NI}$	R&M$_{0.05}^{NI}$	DIC$_3$	DIC$_4$
	0.00001	44%(4)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	6%(5)	68%(3)
	0.001	6%(8)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	4%(5)	68%(3)
	0.01	16%(4)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	28%(5)	54%(3)
	0.05	54%(3)	0%(4)	84%(3)	98%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	96%(3)	52%(3)
Scenario A2	0.1	94%(3)	0%(5)	0%(4)	12%(4)	86%(3)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	68%(3)
(high separation)	0.3	100%(3)	0%(8)	0%(6)	0%(5)	0%(4)	100%(3)	100%(3)	100%(3)	100%(3)	100%(3)	50%(3)
	0.5	100%(3)	0%(9)	0%(8)	0%(6)	0%(5)	98%(3)	100%(3)	100%(3)	100%(3)	100%(3)	64%(3)
	0.9	100%(3)	0%(10)	0%(9)	0%(8)	0%(6)	6%(4)	80%(3)	92%(3)	98%(3)	100%(3)	94%(3)
	0.00001	10%(2)	4%(2)	4%(2)	4%(2)	4%(2)	0%(2)	0%(2)	0%(2)	0%(2)	26%(4)	10%(1)
	0.001	10%(10)	6%(2)	6%(2)	6%(2)	6%(2)	2%(2)	2%(2)	2%(2)	2%(2)	22%(5)	8%(1)
	0.01	16%(2)	36%(2)	34%(2)	34%(2)	34%(2)	2%(2)	2%(2)	2%(2)	2%(2)	42%(3)	6%(1)
	0.05	2%(2)	38%(4)	88%(3)	86%(3)	74%(3)	2%(2)	2%(2)	2%(2)	2%(2)	20%(2)	8%(1)
Scenario A3	0.1	4%(2)	0%(5)	4%(4)	32%(3)	94%(3)	4%(2)	4%(2)	2%(2)	2%(2)	16%(2)	8%(1)
(moderate separation)	0.3	6%(2)	0%(8)	0%(6)	0%(4)	0%(4)	28%(2)	28%(2)	28%(2)	28%(2)	10%(2)	4%(2)
	0.5	8%(2)	0%(9)	0%(8)	0%(7)	0%(5)	60%(3)	48%(2)	46%(2)	46%(2)	18%(2)	4%(5)
	0.9	10%(2)	0%(10)	0%(9)	0%(8)	0%(6)	0%(4)	50%(3)	66%(3)	94%(3)	42%(3)	8%(2)
	0.00001	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	2%(5)	8%(5)
	0.001	8%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	78%(3)	46%(3)
	0.01	6%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	0%(1)	16%(1)	34%(3)
	0.05	0%(1)	46%(3)	2%(2)	0%(2)	0%(2)	0%(1)	0%(1)	0%(1)	0%(1)	10%(1)	0%(1)
Scenario A3	0.1	0%(1)	2%(4)	80%(3)	64%(3)	16%(2)	0%(1)	0%(1)	0%(1)	0%(1)	20%(1)	0%(1)
(low separation)	0.3	0%(1)	0%(8)	0%(6)	0%(5)	8%(4)	0%(2)	0%(1)	0%(1)	0%(1)	20%(1)	2%(1)
	0.5	0%(1)	0%(9)	0%(7)	0%(6)	0%(5)	62%(3)	0%(2)	0%(2)	0%(2)	34%(3)	0%(1)
	0.9	0%(1)	0%(10)	0%(9)	0%(8)	0%(6)	0%(5)	6%(4)	44%(4)	98%(3)	22%(2)	0%(1)

The success rates of BIC using a frequentist approach for high, moderate, and low levels of separation were 100%(3), 16%(2), and 0%(1), respectively.

Table 5.3: Unequal proportions heterogeneous scenario; a heterogeneous population with three clusters. $\lambda_1 = 0.475$, $\lambda_2 = 0.475$, $\lambda_3 = 0.05$, $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 3$ and $\sigma_1 = \sigma_2 = \sigma_3 = 0.25$ (high separation). Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses.

α	RJMCMC	R&M$_0^{\text{NI}}$	R&M$_{0.01}^{\text{NI}}$	R&M$_{0.02}^{\text{NI}}$	R&M$_{0.05}^{\text{NI}}$	DIC$_3$	DIC$_4$
0.00001	48%(3)	100%(3)	100%(3)	100%(3)	48%(2)	0%(5)	14%(4)
0.001	6%(7)	98%(3)	98%(3)	98%(3)	48%(2)	2%(5)	14%(4)
0.01	36%(4)	98%(3)	98%(3)	98%(3)	50%(3)	48%(3)	24%(4)
0.05	70%(3)	100%(3)	100%(3)	100%(3)	50%(3)	86%(3)	16%(4)
0.1	98%(3)	100%(3)	100%(3)	100%(3)	50%(3)	98%(3)	20%(4)
0.3	100%(3)	100%(3)	100%(3)	100%(3)	54%(3)	100%(3)	28%(4)
0.5	100%(3)	100%(3)	100%(3)	100%(3)	56%(3)	100%(3)	44%(3)
0.9	100%(3)	20%(4)	88%(3)	96%(3)	80%(3)	100%(3)	92%(3)

The success rate of BIC using a frequentist approach was 98%(3).

Simulation study B: a longitudinal study with a mixture of Gaussian random effects distributions

Simulation study A enabled us to compare different criteria in a simple setting. However, mixtures also appear in more complicated models, where it may be difficult to calculate some of the criteria that were evaluated in Simulation study A. However, the calculation of the R&M criterion should still be feasible in that case. To verify the performance of the R&M criterion we tested its performance based on a simulation study for a mixture model with longitudinal data.

In this simulation study we generate data from a growth mixture model, which is also known as a latent class mixed effects model (Muthén and Shedden, 1999; Wang and Bodner, 2007) with a mixture model on the random effects (Wang and Bodner, 2007). The density function in a Gaussian growth mixture model can be expressed as 5.1, where $f_j(y)$ is the density function that describes the trajectory for class j. The vector θ_j represents the parameters that are associated with the trajectory of class j.

The growth mixture model for individuals that belong to latent class j can be expressed as:

$$y_{it|j} = \theta_{j0} + b_{ij0} + (\theta_{j1} + b_{ij1})\,\text{time}_{it} + \epsilon_{it},$$

where $y_{it|j}$ is the tth observation of the ith individual, given that this individual is in latent class j. θ_{j0} and θ_{j1} are the fixed intercept and slope of the jth latent class, respectively. b_{ij0} and b_{ij1} are the random intercept and slope of the jth latent class that are assumed to be bivariate normally distributed with mean zero and a class-specific variance-covariance structure. The residuals ϵ_{it} are now assumed to be normally distributed, and independent of the random effects.

In this simulation study we computed the R&M criterion and the BIC. To establish whether a class is empty with the R&M criterion we used different values for the cut-off (ψ) as in Section 5.4 (i.e., 0, 0.01, 0.02, and 0.05 of observations in the sample), and the maximum number of latent classes was also set to $K_{\max} = 10$. Here we considered a homogeneous population ($K = 1$) and a heterogeneous population ($K = 3$).

- **Scenario B1:** (homogeneous data with a random intercept and slope): $K = 1$ class with $\theta_{10} = 2$ and $\theta_{11} = -0.2$, $\mathbf{b_{i1}} \sim N_2(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \begin{bmatrix} 0.25^2 & 0 \\ 0 & 0.025^2 \end{bmatrix}$ and the residuals ϵ_{it} are normally distributed with variance of 0.25^2, $\epsilon_{it} \sim N(0, 0.25^2)$, and independent of the random effects, see Figure 5.2(a).

- **Scenario B2:** (heterogeneous data with a random intercept): $K = 3$ classes with $\theta_{10} = 1$, $\theta_{20} = 2$, $\theta_{30} = 3$ and $\theta_{11} = \theta_{21} = \theta_{31} = -0.2$, $b_{ij0} \sim N(0, 0.25^2)$, $b_{ij1} = 0$ (a random intercept model) for $j = 1, 2, 3$ and the residuals ϵ_{it} are normally distributed with variance of 0.25^2, i.e., $\epsilon_{it} \sim N(0, 0.25^2)$, and independent of the random effects, see Figure 5.2(b).

- **Scenario B3:** (heterogeneous data with a random intercept and slope): $K = 3$ classes with $\theta_{10} = 1$, $\theta_{20} = 2$, $\theta_{30} = 3$ and $\theta_{11} = -0.1$, $\theta_{21} = -0.2$, and $\theta_{31} = -0.3$, $b_{ij0} \sim N(0, 0.25^2)$, $\mathbf{b_{ij}} \sim N_2(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \begin{bmatrix} 0.25^2 & 0 \\ 0 & 0.025^2 \end{bmatrix}$ for $j = 1, 2, 3$ and the residuals ϵ_{it} are normally distributed with variance of 0.25^2, $\epsilon_{it} \sim N(0, 0.25^2)$, and independent of the random effects, see Figure 5.2(c).

Relatively noninformative priors were specified for the class-specific parameters θ_{j0} and θ_{j1}, i.e., $N(0, 10^3)$. An $\Gamma^{-1}(10^{-3}, 10^{-3})$ was specified for the variance of the residuals (this prior also used for the variance of random intercept in Scenario B2). An Inv-Wishart(R, df) distribution was specified for the variance-covariance structure of the random intercept and random slope. We set the degrees of freedom, df, to 3 and the scale parameter matrix, R, to a diagonal matrix with small values, i.e., 10^{-3} Lesaffre and Lawson (2012). For the class membership probability a Dirichlet distribution with different values (i.e., $\alpha =$ 0.00001, 0.001, 0.01, 0.05, 0.1, 0.3, 0.5, 1.0, 1.5, 2.0, and 2.5, where $K_{\max} = 10$) for the class proportions was specified. Larger values here were specified since in Scenario B1 and Scenario B3 $d = 5$.
In this analysis, the data are simulated using equal class proportions (i.e., $\lambda_j = 1/K$ for each class j). Figure 5.2 shows a randomly selected generated data set for the three scenarios.

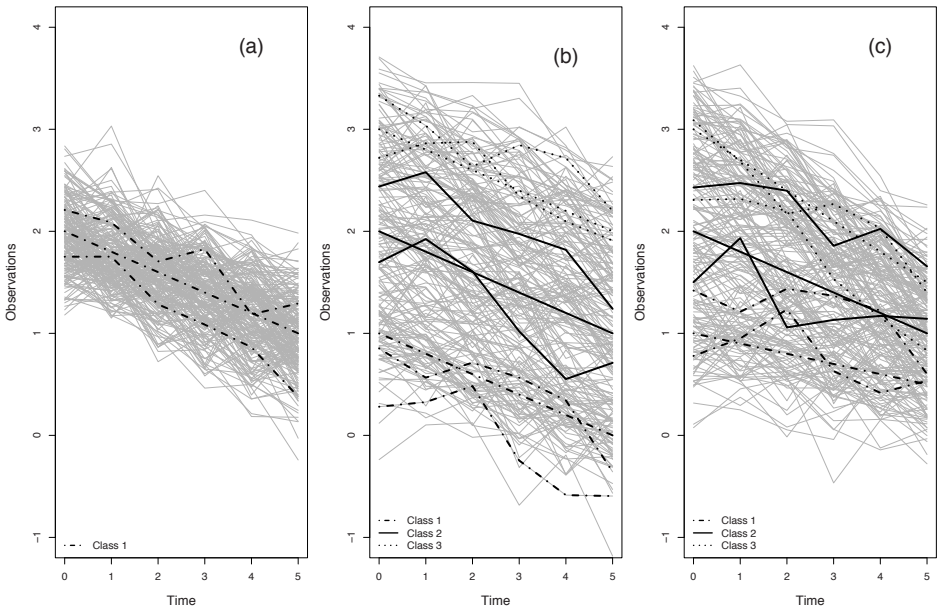


Figure 5.2: Longitudinal simulated data study: The left profile belongs to a homogeneous population with one class. The middle one belongs to a population with three classes where classes differ only in intercept, and the right profile belongs to a heterogeneous population with three classes where classes differ both in intercept and slope.

We generated 50 data sets for each setting consisting of 200 subjects and 6 observations per subject. MCMC sampling is run for each data set for 50,000 iterations after discarding the first 5,000 iterations (burn-in).

Simulation study B: results

The simulation results of Scenario B1 (homogeneous data with a random intercept and slope) show that the R&M criterion with the noninformative prior estimates the true number of classes perfectly. The results of this simulation are presented in Table S3 in Supplementary Material Section 5.7.

Table 5.4: The results of Scenario B2. Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses.

α	R&M$_0^{\text{NI}}$	R&M$_{0.01}^{\text{NI}}$	R&M$_{0.02}^{\text{NI}}$	R&M$_{0.05}^{\text{NI}}$
0.00001	4%(1)	4%(1)	4%(1)	4%(1)
0.001	4%(2)	4%(2)	4%(2)	4%(2)
0.01	18%(2)	18%(2)	18%(2)	18%(2)
0.05	48%(2)	48%(2)	48%(2)	48%(2)
0.1	74%(3)	74%(3)	74%(3)	74%(3)
0.3	90%(3)	90%(3)	90%(3)	90%(3)
0.5	96%(3)	98%(3)	98%(3)	100%(3)
0.9	8%(4)	14%(4)	20%(4)	34%(4)

The success rate of BIC using a frequentist approach was 98(3)%.

Table 5.5: The results of Scenario B3. Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses.

α	R&M$_0^{\text{NI}}$	R&M$_{0.01}^{\text{NI}}$	R&M$_{0.02}^{\text{NI}}$	R&M$_{0.05}^{\text{NI}}$
0.00001	4%(2)	4%(2)	4%(2)	4%(2)
0.001	4%(2)	4%(2)	4%(2)	4%(2)
0.01	6%(2)	6%(2)	6%(2)	6%(2)
0.05	6%(2)	6%(2)	6%(2)	6%(2)
0.1	6%(2)	6%(2)	6%(2)	6%(2)
0.3	10%(2)	10%(2)	10%(2)	10%(2)
0.5	10%(2)	10%(2)	10%(2)	10%(2)
1.0	22%(2)	22%(2)	22%(2)	20%(2)
1.5	36%(2)	36%(2)	36%(2)	32%(2)
2.0	58%(3)	58%(3)	58%(3)	52%(3)
2.5	36%(3)	36%(3)	36%(3)	34%(4)

The success rate of BIC using a frequentist approach was 46%(3).

Table 5.4 shows the simulation results of scenario B2 (heterogeneous data with a random intercept). In this scenario $d = 2$, therefore α should be smaller than 1 to make sure that overfitted classes become empty asymptotically (Rousseau and Mengersen, 2011). In this scenario, large values for α (i.e., $0.1 < \alpha < 0.9$) in the R&M$^{\text{NI}}$ criterion result in an accurate estimation of the true number of latent classes. An underestimation of the number of classes is observed for R&M$^{\text{NI}}$ criterion when a lower value of α is used. In this scenario, different cut-offs lead to the same results.

Table 5.5 shows the simulation results of Scenario B3 (heterogeneous data with a random intercept and slope). In this scenario $d = 5$, therefore α should be smaller than 2.5. In this scenario, setting $\alpha = 2.0$ in the R&M$^{\text{NI}}$ criterion yields the most precise estimation of the true number of latent classes. Using this value for α, the result of the R&M criterion was better than BIC. An underestimation of the number of classes is observed for the R&M$^{\text{NI}}$ criterion when a lower value of α is used. Larger values for α lead to an overestimation of the true number of latent classes.

Simulation study A and B: conclusions

Simulation study A shows how the prior for the class-specific parameters and the Dirichlet prior for the class proportion interact to affect the selection of the correct number of latent class models. Using a hierarchical prior (i.e., a normal-gamma prior) for the class-specific means and variances, values for the Dirichlet hyperparameter α in the range 0.05-0.10 lead to acceptable results with both moderate or high separation between classes. Higher values for α may lead to an overestimation of the number of latent classes, even if α remains well below the threshold value $d/2$ that was given in the proof of Rousseau and Mengersen (2011). For $\alpha < 0.05$ a good performance is observed in the high separation scenario, but the number of classes is underestimated in scenarios with a moderate or low amount of separation. This underestimation of the number of latent classes with a low Dirichlet hyperparameter was not observed in a previous simulation study, however that study simulated only data sets with well separated latent classes (Malsiner-Walli et al., 2016).
With an uninformative prior for the class-specific means and variances, a perfect performance of the R&M criterion is observed in well separated data sets, irrespective of the value of α. An underestimation of the number of classes is observed in the scenarios with a low or moderate separation, especially with low values for α. Setting α to a higher value, while still ensuring that $\alpha < d/2$, led to a considerable improvement in the selection of the number of latent classes in these scenarios. In additional simulations (results not shown), we confirmed that setting α to a value above the threshold (i.e., to $\alpha > d/2$) results in an overestimation of the number of latent classes, as was predicted by the proof in Rousseau and Mengersen (2011).

Using the normal-gamma prior, the performance of the R&M criterion seems quite sensitive to the value of α. In addition, the optimal value of α (i.e., that leads to highest probability of choosing the correct number of classes) depends on the separation between classes and the true number of classes, which are typically not known in practice. In contrast the performance of the R&M criterion with an uninformative prior seems much more stable, as long as the value of α is close to but below the threshold of $d/2$.

Of the 4 possible values for the threshold to determine whether a class is empty (i.e., ψ in 5.5), we found the best performance using a value of 0.01 in the scenario with a moderate separation (see Table S4). In the other scenarios there was no clear difference between the possible values of ψ. Therefore setting $\psi =$ between 0.02 and 0.05 seems reasonable, and a value in this range should allow for the detection of relatively small classes containing a few percent of the population.
Compared to alternative criteria for selecting the number of latent classes, the performance of the R&M criterion was good.
The performance of BIC was generally inferior to that of the R&M criterion, especially in data sets with many latent classes and data sets with moderate or low separation. The performance of DIC_3, DIC_4 and RJMCMC depends on the value of α. Although in some scenarios specific values of α seem to lead to a good performance for these criteria, there is no value of α that leads a good performance across all scenarios.
Simulation study B confirms the conclusions of simulation study A. It shows that the R&M criterion can also be implemented in a more complex and realistic setting such as a growth mixture model for longitudinal data. The R&M criterion using uninformative priors for the class-specific parameters and α smaller than but close to $d/2$ (e.g., between 0.8 and 0.9 $d/2$) yielded the best results, and outperformed BIC. However, the results were generally less good in the Scenario B3, which has a more complex structure with random intercept and slope.

5.5 Hemoglobin longitudinal data

In this section, we apply the R&M criterion to a finite mixture model for hemoglobin (Hb) values of blood donors. Our motivating application is the trajectory of Hb values of blood donors over successive donations. Blood donors experience a temporary reduction in their Hb value after donation. Therefore, a minimum 8 week interval between two donations is set by the blood bank, to allow the donor's Hb value to recover to its pre-donation level. However, this interval seems to be too short since on average there is a declining trajectory in the Hb values for blood donors who donate regularly (Brittenham, 2011; Cable et al., 2011). Therefore, a considerable proportion of prospective blood donors are temporarily deferred from donation each year due to low Hb values (Newman, 2004). A Hb value of 8.4 mmol/l (135 g/l) and 7.8 mmol/l (125 g/l) for men and women, respectively, is widely accepted as the lower cut-off value of eligibility for donation to protect donors from anemia (Radtke et al., 2005). The previous studies showed that some individuals have a fast recovery, which results in a relatively stable trajectory, whereas others have a slow recovery that yields a declining trajectory in their Hb values (Nasserinejad et al., 2015, 2016).
Here, we use a data set of longitudinally observed Hb values from 1 January 2005 to 31 December 2012 collected by Sanquin Blood Supply in the Netherlands. This data set is based

on a self-administered questionnaire study aimed at gaining insight into characteristics and motivation of the Dutch donor population (Atsma et al., 2011). Here we randomly selected 200 new registered male blood donors who have at least 5 visits to blood bank. These data are part of the Donor InSight study, for more details see Atsma et al. (2011).
A mixed-effects model with random intercept and slope may be able to capture the heterogeneity between individuals in these data. However previous studies suggested that describing the total donor population using a single trajectory may oversimplify the complex growth patterns of this population (Nasserinejad et al., 2015, 2016). Therefore, a growth mixture modeling approach, which accounts for different subgroups of donors, seems to be a more appropriate method for capturing differences in Hb trajectories between donors (Nasserinejad et al., 2015, 2016). Here we implemented the R&M criterion with relatively uninformative priors for the parameters. Different cut-offs (i.e., 0, 0.01, 0.02, and 0.05) were used to define a class to be empty.
Several factors are known to be associated with Hb and hence may be used as predictors, i.e., sex (Yip et al., 1984), season (Hoekstra et al., 2007), age (Yip et al., 1984). Here we model Hb trajectory based on number of donations in last two years (NODY2), the season donation took place (a binary value for cold=1 and warm seasons=0), time since previous donation (TSPD), and age of donor (years) at first visit. The class-specific parameters are the intercept and the effect of NODY2. The aim of the model is to assign each donor to one of j groups in such a way that donors with similar Hb trajectories are in the same group, and that the groups are most different from each other in terms of the Hb trajectory.
The growth mixture model for the trajectory of Hb levels of blood donors who belong to latent class j can be expressed as:

$$\text{Hb}_{it|j} = \theta_{j0} + b_{ij0} + \gamma_1 \text{Age}_i + \gamma_2 \text{Season}_{it} + \gamma_3 \text{TSPD}_{it} + (\theta_{j1} + b_{ij1})\text{NODY2}_{it} + \epsilon_{it},$$

where $\text{Hb}_{it|j}$ is the predicted Hb level at the tth observation of the ith individual, given that this individual is in latent class j. θ_{j0} and θ_{j1} are the fixed intercept and slope (coefficients of NODY2) of latent class j. b_{ij0} and b_{ij1} are the random intercept and slope of latent class j that are assumed to be bivariate normally distributed with mean zero and a class-specific variance-covariance structure. The residuals ϵ_{it} are assumed to be normally distributed, and independent of the random effects.

Prior specification

The priors for the model parameters were chosen as follows. Relatively noninformative priors were specified for both the class-specific parameters θ's and the non-class-specific parameters γ's, i.e., $N(0, 10^3)$. An $\Gamma^{-1}(10^{-3}, 10^{-3})$ was specified for the variance of the residuals. An Inv-Wishart(R, df) distribution was specified for the variance-covariance structure of the random intercept and random slope. We set the degrees of freedom, df, to 3 and the scale parameter matrix, R, to a diagonal matrix with small values, i.e., 10^{-3}

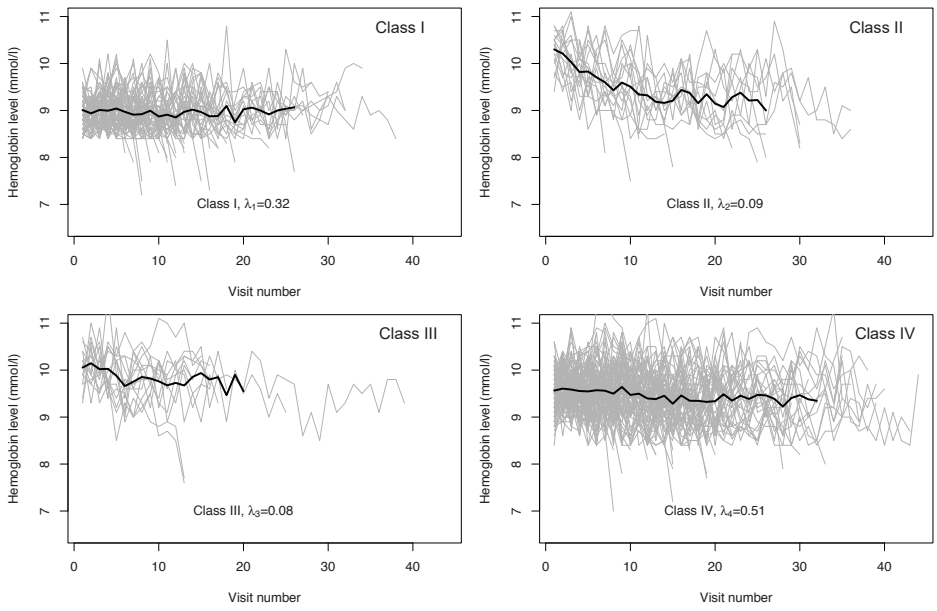


Figure 5.3: Hb profiles for four different classes.

Lesaffre and Lawson (2012). Since the number of class-specific parameters d is 5, for the class membership probability a Dirichlet distribution with different values for alpha (i.e., 1.0, 1.5, 2.0, and 2.5) was specified for the mixing proportions.

To analyze these data, we chose the results with $\alpha = 2$, in view of the results of Scenario B3. Therefore, donors can be assigned to four different classes. Based on the highest posterior probability, individuals were assigned to the latent classes. The profiles of these different classes are displayed in Figure 5.3. This figure shows how trajectories of Hb values for blood donors are different. A group of donors have a low initial Hb value but relatively stable trajectory (Class I), donors in Class II have a very high initial Hb value and a very sharply declining trajectory. Donors in Class III have a high initial Hb value and a moderately declining trajectory, donors in Class IV have moderate initial Hb value and relatively stable trajectory. The results of this study regarding the number of latent classes and the interpretation of each class are supported by a previous study (Nasserinejad et al., 2015).

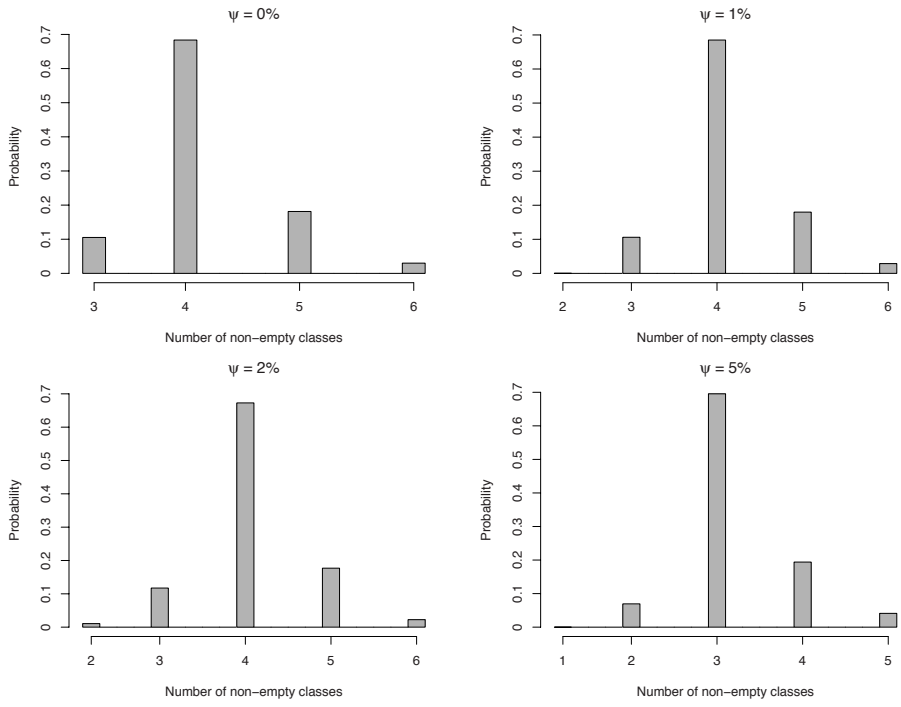


Figure 5.4: Posterior distribution of non-empty classes (K) for different cut-offs (ψ).

Table 5.6: Number of latent classes in Hb data for different α and different cut-offs (ψ).

α	R&M$_0^{\text{NI}}$	R&M$_{0.01}^{\text{NI}}$	R&M$_{0.02}^{\text{NI}}$	R&M$_{0.05}^{\text{NI}}$
0.5	1	1	1	1
1.0	1	1	1	1
1.5	2	2	2	2
2.0	4	4	4	3
2.5	4	4	4	3

BIC using a frequentist approach found 2 classes.

Figure 5.4 shows the posterior distribution of the number of non-empty classes (K) for different cut-offs (ψ) using 50,000 MCMC iterations when $\alpha = 2$. This figure show how the posterior mode of the number of nonempty classes may be affected by changing the ψ.

5.6 Discussion

The results of the simulation studies showed that the R&M criterion has a high probability of estimating the correct number of latent classes, provided that the priors on the proportions and the class-specific parameters are chosen carefully. Despite the simplicity of this criterion, it performs at least as good as alternative selection criteria for the number of latent classes. The application of the R&M criterion to longitudinal data of blood donors further illustrated the practical usefulness of this method.

An important advantage of the R&M criterion is that this approach is straightforward to implement, using MCMC sampling for a mixture model with a large number of latent classes. The number of nonempty latent classes (i.e., classes with a proportion larger than the predefined cutoff value) is then an automatic byproduct of the MCMC sampler. Therefore, this criterion is easily implemented in standard Bayesian software such as WinBugs and JAGS, even for complex statistical models such as latent class mixed-effects models and multivariate mixture models. A further advantage of the R&M criterion is that it is not affected by label switching. Despite the fact that the R&M criterion is relatively easy to implement, this criterion seems to perform better than alternative criteria at estimating the true number of classes. Although only a limited set of statistical models was considered in the simulations, these results suggest that the R&M criterion works well and may be considered for practical use in Bayesian finite mixture models.

A strength of this study is that it is one of the first studies to compare different criteria for selecting the number of latent classes in a Bayesian setting. Although the R&M criterion has been implemented in simulated data previously (Malsiner-Walli et al., 2016), our study adds important insight into how this criterion should be implemented, based on a more elaborate simulation study with several scenarios.

In a previous simulation study, it was shown that using a sufficiently low value of α (e.g., $\alpha < 0.001$) prevents overfitting of the number of latent classes, and that using higher values of α, with $\alpha < d/2$ can lead to overfitting (Malsiner-Walli et al., 2016). In that study, no underestimation of the number of latent classes was observed. In our simulation study we observed that with a slightly lower amount of separation between classes than in the previous study, underestimation of the number of classes often occurs, especially with low values of α. This shows that the value of α should be chosen to provide a trade-off between the probability of overfitting and the probability of underfitting the number of latent classes. Furthermore we observed that if an uninformative prior is used for the class-specific parameters, overfitting of the number of latent classes does not seem to occur, provided that $\alpha < d/2$.

Rousseau and Mengersen (2011) proved that the class proportions converge to 0, not that they should be 0 with any data set of finite size. We therefore used different cut-offs for the proportions in a class to define a class to be empty. Using a cut-off of 0 may be sensitive to outliers in the data and did not perform well in the simulation studies. A cut-off of

between 0.02 and 0.05 should be sufficient to make the criterion robust to outliers, while being small enough to avoid the exclusion of real segments in the population. The interest of finding classes with small proportion may depend on the application and the research questions. Therefore, in practice the value of the cut-off may be chosen by the researcher in advance. Due to the asymptotic nature of the result of Rousseau and Mengersen (2011), larger sample sizes would generally warrant lower values for the cut-off.
Based on the results of the simulation studies, as discussed above, combined with the results of the blood donor data set, we give the following recommendations:

- We recommend to consider the R&M criterion to choose the number of latent classes in Bayesian finite mixture models. This criterion is easy to implement in practice, and its performance compares favorably with alternative criteria.
- To implement the criterion, one should first estimate a mixture model with a large number of classes (e.g., 10 classes), so that some classes will be overfitted.
- The number of classes in the final finite mixture model is then chosen as the posterior mode of the number of classes with a proportion larger than the predefined cut-off, which we recommend to set between 0.02 - 0.05. Lower values of the cut-off should be used if the researcher is specifically interested in the classes with small proportions in the population.
- It seems best to use vague or uninformative priors for the class-specific parameters, and the use of hierarchical priors such as the normal-gamma prior is not recommended.
- The class proportions should be given a Dirichlet prior with α lower than d/2, i.e., the number of class-specific parameters divided by 2. A value of α slightly lower than d/2 (e.g., between 0.8 and 0.9 $d/2$) seems to yield the best results.

A limitation of this study is that only finite mixtures of Gaussian distributions and growth mixtures models were considered in the simulation study. Although the results of the simulation study were similar in these two types of models, it is not certain that the performance of the R&M will be similar in other types of models. Due to the large computation time associated with simulation studies in a Bayesian setting, it was not feasible to consider additional statistical models.
Another limitation is that only predefined settings were evaluated for the priors of both class-specific parameters and the class proportions. It is possible that intermediate values of α or ψ, or also other priors not considered here would lead to a better performance. We further did not consider alternatives to the normal-gamma prior and the noninformative prior for the class-specific parameters.

Conclusion

If appropriate priors are used for both the class-specific parameters and the class proportions, it seems possible to effectively estimate the number of latent classes in a Bayesian finite mixture model using the R&M criterion. This criterion compares favorably to alternative model selection criteria for the number of latent classes in terms of both performance and ease of implementation.

5.7 Supplementary material

Table S1: Scenario Appendix 1; a heterogeneous population with different clusters ($K = 1, \ldots, 6$). $\mu_j = j$ and $\sigma_j = 0.25$, ($j = 1, \ldots, 6$), and ($K_{\max} = 10$). Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses. A relatively uninformative prior was used for the class-specific parameters.

α	Cut-off	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
0.00001	R&M$_0^{\text{NI}}$	100%(1)	100%(2)	95%(3)	70%(4)	40%(4)	10%(4)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	100%(2)	95%(3)	70%(4)	40%(4)	10%(4)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	100%(2)	95%(3)	70%(4)	40%(4)	10%(4)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	100%(2)	95%(3)	70%(4)	40%(4)	10%(4)
0.001	R&M$_0^{\text{NI}}$	100%(1)	100%(2)	95%(3)	70%(4)	30%(4)	0%(4)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	100%(2)	95%(3)	70%(4)	30%(4)	0%(4)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	100%(2)	95%(3)	70%(4)	30%(4)	0%(4)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	100%(2)	95%(3)	70%(4)	30%(4)	0%(4)
0.01	R&M$_0^{\text{NI}}$	100%(1)	100%(2)	95%(3)	75%(4)	55%(5)	5%(4)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	100%(2)	95%(3)	75%(4)	55%(5)	5%(4)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	100%(2)	95%(3)	75%(4)	55%(5)	5%(4)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	100%(2)	95%(3)	75%(4)	55%(5)	5%(4)
0.05	R&M$_0^{\text{NI}}$	100%(1)	100%(2)	100%(3)	75%(4)	60%(5)	35%(5)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	75%(4)	60%(5)	35%(5)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	75%(4)	60%(5)	35%(5)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	75%(4)	60%(5)	35%(5)
0.1	R&M$_0^{\text{NI}}$	100%(1)	100%(2)	100%(3)	80%(4)	80%(5)	55%(6)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	80%(4)	80%(5)	50%(6)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	80%(4)	80%(5)	50%(6)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	80%(4)	80%(5)	45%(5)
0.3	R&M$_0^{\text{NI}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	95%(6)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	90%(6)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	90%(6)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	90%(6)
0.5	R&M$_0^{\text{NI}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	85%(6)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
0.9	R&M$_0^{\text{NI}}$	95%(1)	95%(2)	45%(3)	0%(5)	0%(6)	0%(7)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	70%(4)	25%(6)	10%(7)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	85%(4)	60%(5)	30%(7)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	100%(2)	100%(3)	100%(4)	95%(5)	90%(6)
frequentist	BIC	100%(1)	100%(2)	100%(3)	65%(4)	45%(5)	15%(7)

Table S2: Scenario Appendix 2; a heterogeneous population with different clusters ($K = 1, \ldots, 6$). $\mu_j = j$ and $\sigma_j = 0.25$, ($j = 1, \ldots, 6$), and ($K_{\max} = 10$). Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses. A normal-gamma prior was used for the class-specific parameters.

α	Cut-off	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
0.00001	R&M$_0^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	70%(6)
	R&M$_{0.01}^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	70%(6)
	R&M$_{0.02}^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	70%(6)
	R&M$_{0.05}^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	70%(6)
0.001	R&M$_0^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
	R&M$_{0.01}^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
	R&M$_{0.02}^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
	R&M$_{0.05}^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
0.01	R&M$_0^{\text{NG}}$	95%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
	R&M$_{0.01}^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
	R&M$_{0.02}^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
	R&M$_{0.05}^{\text{NG}}$	100%(1)	100%(2)	100%(3)	100%(4)	100%(5)	100%(6)
0.05	R&M$_0^{\text{NG}}$	30%(2)	5%(3)	0%(3)	5%(5)	40%(6)	80%(6)
	R&M$_{0.01}^{\text{NG}}$	55%(1)	85%(2)	95%(3)	100%(4)	95%(5)	100%(6)
	R&M$_{0.02}^{\text{NG}}$	60%(1)	95%(2)	100%(3)	100%(4)	95%(5)	100%(6)
	R&M$_{0.05}^{\text{NG}}$	80%(1)	95%(2)	100%(3)	100%(4)	100%(5)	100%(6)
0.1	R&M$_0^{\text{NG}}$	0%(3)	0%(4)	0%(5)	0%(6)	0%(6)	0%(7)
	R&M$_{0.01}^{\text{NG}}$	0%(3)	0%(3)	0%(4)	0%(5)	15%(6)	80%(6)
	R&M$_{0.02}^{\text{NG}}$	0%(2)	0%(3)	10%(4)	20%(5)	70%(5)	100%(6)
	R&M$_{0.05}^{\text{NG}}$	10%(2)	45%(3)	80%(3)	100%(4)	100%(5)	100%(6)
0.3	R&M$_0^{\text{NG}}$	0%(7)	0%(7)	0%(8)	0%(8)	0%(9)	0%(9)
	R&M$_{0.01}^{\text{NG}}$	0%(6)	0%(6)	0%(6)	0%(7)	0%(7)	0%(8)
	R&M$_{0.02}^{\text{NG}}$	0%(5)	0%(5)	0%(6)	0%(6)	0%(7)	0%(7)
	R&M$_{0.05}^{\text{NG}}$	0%(4)	0%(4)	0%(4)	0%(5)	0%(6)	75%(6)
0.5	R&M$_0^{\text{NG}}$	0%(8)	0%(9)	0%(9)	0%(9)	0%(10)	0%(10)
	R&M$_{0.01}^{\text{NG}}$	0%(7)	0%(7)	0%(7)	0%(8)	0%(8)	0%(9)
	R&M$_{0.02}^{\text{NG}}$	0%(6)	0%(6)	0%(7)	0%(7)	0%(8)	0%(8)
	R&M$_{0.05}^{\text{NG}}$	0%(5)	0%(5)	0%(5)	0%(6)	0%(6)	5%(7)
0.9	R&M$_0^{\text{NG}}$	0%(9)	0%(10)	0%(10)	0%(10)	0%(10)	0%(10)
	R&M$_{0.01}^{\text{NG}}$	0%(8)	0%(9)	0%(9)	0%(9)	0%(9)	0%(9)
	R&M$_{0.02}^{\text{NG}}$	0%(8)	0%(8)	0%(8)	0%(8)	0%(8)	0%(9)
	R&M$_{0.05}^{\text{NG}}$	0%(6)	0%(6)	0%(6)	0%(6)	0%(7)	0%(7)
frequentist	BIC	100%(1)	100%(2)	100%(3)	65%(4)	45%(5)	15%(7)

Table S3: The results of Scenario B1; Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses.

α	R&M$_0^{\text{NI}}$	R&M$_{0.01}^{\text{NI}}$	R&M$_{0.02}^{\text{NI}}$	R&M$_{0.05}^{\text{NI}}$
0.00001	100%(1)	100%(1)	100%(1)	100%(1)
0.001	100%(1)	100%(1)	100%(1)	100%(1)
0.01	100%(1)	100%(1)	100%(1)	100%(1)
0.05	100%(1)	100%(1)	100%(1)	100%(1)
0.1	100%(1)	100%(1)	100%(1)	100%(1)
0.3	100%(1)	100%(1)	100%(1)	100%(1)
0.5	100%(1)	100%(1)	100%(1)	100%(1)
1.0	100%(1)	100%(1)	100%(1)	100%(1)
1.5	100%(1)	100%(1)	100%(1)	100%(1)
2.0	98%(1)	98%(1)	98%(1)	98%(1)

The success rate of BIC using a frequentist approach was 100%.

Table S4: Scenario Appendix 3; a heterogeneous population with different clusters ($K = 1, \ldots, 6$). $\mu_j = j$ and $\sigma_j = 0.40$, ($j = 1, \ldots, 6$), and ($K_{\max} = 10$). Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses. A relatively uninformative prior was used for the class-specific parameters.

α	Cut-off	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
0.00001	R&M$_0^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
0.001	R&M$_0^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
0.01	R&M$_0^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
0.05	R&M$_0^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
0.1	R&M$_0^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	5%(1)	0%(2)	0%(2)	0%(3)	0%(3)
0.3	R&M$_0^{\text{NI}}$	100%(1)	40%(1)	5%(2)	0%(3)	0%(3)	0%(4)
	R&M$_{0.01}^{\text{NI}}$	100%(1)	35%(1)	5%(2)	0%(3)	0%(3)	0%(3)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	35%(1)	5%(2)	0%(3)	0%(3)	0%(3)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	35%(1)	5%(2)	0%(3)	0%(3)	0%(3)
0.5	R&M$_0^{\text{NI}}$	90%(1)	55%(2)	25%(3)	5%(3)	0%(4)	0%(5)
	R&M$_{0.01}^{\text{NI}}$	90%(1)	55%(2)	15%(3)	0%(3)	0%(3)	0%(4)
	R&M$_{0.02}^{\text{NI}}$	100%(1)	55%(2)	15%(3)	0%(3)	0%(3)	0%(4)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	55%(2)	15%(3)	0%(3)	0%(3)	0%(4)
0.9	R&M$_0^{\text{NI}}$	80%(1)	65%(2)	25%(4)	15%(5)	10%(6)	15%(7)
	R&M$_{0.01}^{\text{NI}}$	95%(1)	95%(2)	95%(3)	95%(4)	95%(5)	95%(6)
	R&M$_{0.02}^{\text{NI}}$	95%(1)	90%(2)	95%(3)	95%(4)	90%(5)	30%(5)
	R&M$_{0.05}^{\text{NI}}$	100%(1)	95%(2)	90%(3)	60%(4)	25%(4)	5%(5)
frequentist	BIC	100%(1)	45%(1)	15%(2)	0%(3)	0%(3)	0%(3)

Table S5: Scenario Appendix 4; a heterogeneous population with different clusters ($K = 1, \ldots, 6$). $\mu_j = j$ and $\sigma_j = 0.40$, ($j = 1, \ldots, 6$), and ($K_{\max} = 10$). Percentage of data sets in which the true number of clusters was found, with the mode of the estimated number of classes in parentheses. A normal-gamma prior was used for the class-specific parameters.

α	Cut-off	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
0.00001	R&M$_0^{\text{NG}}$	100%(1)	15%(1)	0%(2)	0%(2)	0%(2)	0%(3)
	R&M$_{0.01}^{\text{NG}}$	100%(1)	15%(1)	0%(2)	0%(2)	0%(2)	0%(3)
	R&M$_{0.02}^{\text{NG}}$	100%(1)	15%(1)	0%(2)	0%(2)	0%(2)	0%(3)
	R&M$_{0.05}^{\text{NG}}$	100%(1)	15%(1)	0%(2)	0%(2)	0%(2)	0%(3)
0.001	R&M$_0^{\text{NG}}$	100%(1)	20%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.01}^{\text{NG}}$	100%(1)	20%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.02}^{\text{NG}}$	100%(1)	15%(1)	0%(2)	0%(2)	0%(3)	0%(3)
	R&M$_{0.05}^{\text{NG}}$	100%(1)	15%(1)	0%(2)	0%(2)	0%(3)	0%(3)
0.01	R&M$_0^{\text{NG}}$	100%(1)	55%(2)	25%(2)	0%(3)	0%(3)	0%(3)
	R&M$_{0.01}^{\text{NG}}$	100%(1)	50%(2)	20%(2)	0%(3)	0%(3)	0%(3)
	R&M$_{0.02}^{\text{NG}}$	100%(1)	50%(2)	20%(2)	0%(3)	0%(3)	0%(3)
	R&M$_{0.05}^{\text{NG}}$	100%(1)	50%(2)	20%(2)	0%(3)	0%(3)	0%(3)
0.05	R&M$_0^{\text{NG}}$	30%(2)	40%(3)	75%(3)	85%(4)	20%(4)	0%(4)
	R&M$_{0.01}^{\text{NG}}$	65%(1)	85%(2)	75%(3)	10%(3)	0%(3)	0%(4)
	R&M$_{0.02}^{\text{NG}}$	70%(1)	95%(2)	60%(3)	5%(3)	0%(3)	0%(4)
	R&M$_{0.05}^{\text{NG}}$	80%(1)	90%(2)	40%(2)	0%(3)	0%(3)	0%(4)
0.1	R&M$_0^{\text{NG}}$	0%(3)	0%(4)	0%(5)	0%(5)	65%(5)	65%(6)
	R&M$_{0.01}^{\text{NG}}$	0%(3)	5%(3)	15%(4)	100%(4)	40%(5)	0%(5)
	R&M$_{0.02}^{\text{NG}}$	0%(2)	10%(3)	50%(3)	100%(4)	25%(5)	0%(4)
	R&M$_{0.05}^{\text{NG}}$	10%(2)	45%(3)	95%(3)	40%(3)	0%(4)	0%(4)
0.3	R&M$_0^{\text{NG}}$	0%(7)	0%(7)	0%(8)	0%(8)	0%(8)	0%(0)
	R&M$_{0.01}^{\text{NG}}$	0%(6)	0%(6)	0%(6)	0%(7)	0%(7)	0%(7)
	R&M$_{0.02}^{\text{NG}}$	0%(5)	0%(5)	0%(6)	0%(6)	0%(6)	95%(6)
	R&M$_{0.05}^{\text{NG}}$	0%(4)	0%(4)	0%(5)	0%(5)	100%(5)	10%(5)
0.5	R&M$_0^{\text{NG}}$	0%(8)	0%(9)	0%(9)	0%(9)	0%(9)	0%(10)
	R&M$_{0.01}^{\text{NG}}$	0%(7)	0%(7)	0%(8)	0%(8)	0%(8)	0%(8)
	R&M$_{0.02}^{\text{NG}}$	0%(6)	0%(6)	0%(7)	0%(7)	0%(7)	0%(7)
	R&M$_{0.05}^{\text{NG}}$	0%(5)	0%(5)	0%(5)	0%(6)	0%(6)	100%(6)
0.9	R&M$_0^{\text{NG}}$	0%(9)	0%(10)	0%(10)	0%(10)	0%(10)	0%(10)
	R&M$_{0.01}^{\text{NG}}$	0%(8)	0%(9)	0%(9)	0%(9)	0%(9)	0%(9)
	R&M$_{0.02}^{\text{NG}}$	0%(8)	0%(8)	0%(8)	0%(8)	0%(8)	0%(8)
	R&M$_{0.05}^{\text{NG}}$	0%(6)	0%(6)	0%(6)	0%(6)	0%(7)	15%(7)
frequentist	BIC	100%(1)	45%(1)	15%(2)	0%(3)	0%(3)	0%(3)

6 Conclusions

In this chapter, we provide an overview of the statistical techniques and the main findings reported in the preceding chapters with a discussion on possible improvements and future extensions.

6.1 General conclusions

In **Chapter 2**, we showed that the transition models and the mixed effects models provide a much better prediction compared to a multiple linear regression model. In general, the transition model provides a somewhat better prediction than the mixed effects model, especially at high visit numbers. In addition, the transition model offers a better trade-off between sensitivity and specificity when varying the cut-off values for eligibility in predicted values. Hence transition models make the prediction of hemoglobin level more precise and may lead to less deferral from donation in the future.
Moreover, both the transition and the mixed effects models use the data of a person's previous observations for making predictions. In the transition model only the last q observations are used for predicting the current response. However, in the mixed effects model, the empirical Bayes method for estimating a person's random effects uses all previous observations. Therefore, the mixed effects model requires more historical information than the transition model. Since the transition model is convenient in practice and needs less historical information compared to the mixed effects model, blood banks may use this model to predict the future hemoglobin level of a candidate and to determine which candidates should not be invited for the next donation.
In **Chapter 3** our findings suggest that describing the total donor population using a single trajectory oversimplifies the complex growth patterns of this population. Instead, a growth mixture modeling approach, which accounts for different subgroups of donors, seems to be an appropriate method for capturing differences in Hb trajectories between donors. Therefore, individual donors belonging to different classes should potentially be approached differently. For donors with a low but stable Hb trajectory, delaying the next invitation may not help to decrease the probability of deferral. Donors with a normal initial Hb level become at risk for deferral only with a high donation frequency, because the estimated Hb decline per donation is fairly small in this group. Thus for this group, the advice could be to increase donation intervals. Donors with high initial Hb levels do not have a very high risk of Hb deferral. Changing their donation intervals may therefore not be very effective in preventing Hb deferral.
In **Chapter 4**, we show that the estimated recovery time is considerably longer than the mandatory interval between donations (i.e., 56 days). Also, our findings point to a concave Hb recovery process. That is, the recovery process is fastest at the beginning and becomes slower over time. The estimated recovery time should be seen as the ultimate recovery time, i.e., the time by which a donor's Hb value has fully recovered. Due to

the concave shape of the recovery process, most of the recovery occurs before half of the recovery time has passed, which partially explains the long estimated recovery times in our data set. Furthermore, it should be noted that a recovery time that is longer than the average interval between donations is in line with the observed data, as there is a decline in the Hb trajectories with the number of donations.
Another interesting finding is that there is heterogeneity between donors in the recovery time, i.e., 54.3% and 51.3% of male and female donors have a constant recovery time during successive donations. The remaining donors have a longer recovery time and their recovery time increases after a number of donations. This increase in recovery time might be attributed to a reduction of the iron reserves in these donors.
In **Chapter 5**, the results of the simulation studies showed that the Rousseau and Mengersen (2011) (R&M) criterion has a high probability of estimating the correct number of latent classes, provided that the priors on the proportions and the class-specific parameters are chosen carefully. Despite the simplicity of this criterion, it performs at least as good as alternative selection criteria for the number of latent classes. The application of the R&M criterion to longitudinal data of blood donors further illustrated the practical usefulness of this method.
An important advantage of the R&M criterion is that this approach is straightforward to implement, using MCMC sampling for a mixture model with a large number of latent classes. The number of nonempty latent classes (i.e., classes with a proportion larger than the predefined cutoff value) is then an automatic byproduct of the MCMC sampler. Therefore, this criterion can easily be implemented in standard Bayesian software such as WinBugs and JAGS, even for complex statistical models such as latent class mixed-effects models and multivariate mixture models. A further advantage of the R&M criterion is that it is not affected by label switching. Despite the fact that the R&M criterion is relatively easy to implement, this criterion seems to perform better than alternative criteria at estimating the true number of classes.

6.2 Future research

Although we designed our models in this thesis to be very flexible to model longitudinal Hb values, we do not claim that our final models for blood donors are optimal; further research is needed to arrive at a better prediction model. Our proposed models in this thesis are only few models out of many possible models. For instance, models that predict the probability of rejection as a function of previous Hb values could be developed. In addition, it may be possible to develop statistical models that treat the amount of iron reserves as a latent variable. The amount of iron reserves is reduced at every blood donation, but is slowly replenished over time. Such a model may be able to explain why the recovery process depends on the number of previous donations, as the iron reserves are depleted

with successive donations.
Further our models for predicting Hb values could be improved by incorporating more covariates that can affect Hb value such as physical activity, race, nutrition, BMI, and smoking status. However, due to a lack of information, we could not incorporate them into our model.
Based on our latent class model, some donors appear to have high initial Hb levels and others do not, and some show faster declines in Hb than others. This may be due to differences in lifestyle, iron status, iron metabolism, and/or erythropoiesis. Including more of this information in the models may improve the precision of the prediction of latent classes at the first few visits. For this reason class membership could be modeled to depend on some of these time-independent covariates and/or genetic information of donors. Moreover, future research should indicate whether adverse health effects of donation are different for donors with stable or declining Hb levels.
Finally, the ultimate purpose of the prediction exercise is not the prediction of the future Hb value, but rather to determine the best time for the donor to return for donation. Hence, one should focus on predicting the recovery time on an individual level in order to determine the optimal interval between donations from a cost-effectiveness point of view, instead of on predicting future Hb values.

7 Summary/Samenvatting
Acknowledgements
Ph.D. portfolio
Curriculum Vitae

Summary

Blood transfusion is an essential part of modern healthcare, which helps save millions of lives each year. Since blood is a unique resource for which an artificial substitute has yet to be found, blood donations are in great need. Although blood donations and subsequent transfusions are meant to help patients, they might be harmful for both the recipients and the donors. The most important potential harm for donors is that whole blood donation causes a loss of iron and blood cells. Repeated donations could however deplete iron stores, leading to iron depletion and ultimately anemia. To protect donors from developing iron deficiency anemia after blood donations, the Hb value of blood donors is assessed prior to each donation. Donors with too low Hb value are deferred from donation to protect donor health and quality of donated blood. Although deferrals are meant to protect donors and the quality of blood units for transfusion, Hb deferrals decrease the cost-effectiveness of blood supply, because (i) testing and deferring a donor are expensive; (ii) for every deferred donor, another donor needs to be invited to reach collection targets; and (iii) lapsing donors need to be replaced by new donors because deferred candidates rarely return for donation.

In this thesis, we developed and investigated several Bayesian statistical approaches to model longitudinal Hb values in blood donors.

Chapter 1 provides a brief overview to the clinical and statistical issues that were addressed in this thesis. In **Chapter 2**, we compare transition models with different numbers of autoregressive terms and mixed effects models, as plausible models to account for the dependence among subsequent Hb values within a donor and as models to predict the future Hb value. This chapter shows the capabilities of longitudinal models in prediction and our findings may help reduce the number of deferred candidate in the blood banks. In **Chapter 3**, we show that describing the total donor population using a single trajectory oversimplifies the complex growth patterns of donor population. Instead, a growth mixture modeling approach, which accounts for different subgroups of donors, seems to be an appropriate method for capturing differences in Hb trajectories between donors. These findings are of high importance for identification of donors who could benefit from tailored donation intervals to prevent iron deficiency and donor deferrals.

Chapter 4 considers a latent class mixed-effects transition model for the prediction of a future Hb value for a potential blood donor given the previous observations, and the estimation of the recovery time after a donation. The advantage of this model is that it simultaneously captures heterogeneity and state dependence. In this model, the heterogeneity in the recovery process was controlled using latent classes and the dynamics of the recovery process using a change point model. The temporary reduction of Hb after donation was modeled using a flexible function. This flexible function enables us to estimate the recovery time, which is the time needed for Hb to return to its pre-donation value. These results are important for practice, as they may be used to improve the planning of donors'

visits to the blood banks and help to tailor donation intervals and prevent iron deficiency and donor deferrals.
In **Chapter 5** we performed an extensive simulation study to develop practical guidelines to determine the appropriate number of latent classes based on the posterior distribution of the class proportions, and to compare this criterion with alternative criteria. The performance of the proposed criterion is illustrated using a data set of repeatedly measured hemoglobin values of blood donors. **Chapter 6** provides a summary of the findings in the preceding chapters with a discussion on possible improvements and future extensions.

Samenvatting

Bloedtransfusie is een essentieel onderdeel van de hedendaagse gezondheidszorg. Dankzij bloedtransfusie worden er jaarlijks miljoenen levens gered. Aangezien er geen kunstmatig alternatief is voor menselijk bloed, is er nog steeds een grote behoefte aan gedoneerd bloed. Hoewel bloeddonatie en de daaropvolgende bloedtransfusie in het algemeen bijdragen aan de gezondheid van patinten, zijn er ook risicos aan verbonden voor zowel de bloeddonoren als ontvangers. Voor bloeddonoren is dit het verlies aan ijzer en rode bloedcellen. Immers veelvuldige bloeddonatie kan leiden tot een afname van de ijzerreserves in het menselijk lichaam, en zelfs tot een ijzertekort en de ontwikkeling van anemie (bloedarmoede).

Voorafgaand aan elke donatie wordt de hemoglobinewaarde (Hb-waarde) van de donor gemeten. Bij donoren met een te lage Hb-waarde wordt de bloeddonatie uitgesteld, zodat de ijzerreserves de tijd krijgen zich te herstellen en de ontwikkeling van een ijzertekort of anemie wordt voorkomen. Deze strategie moet ook bijdragen tot de kwaliteit van het gedoneerde bloed. Echter, een belangrijk nadeel van het uitstellen van een donatie is de verlaagde kosteneffectiviteit. Deze wordt veroorzaakt door a) de kosten van het testen van de Hb-waarde en het maken van een nieuwe afspraak voor bloeddonatie, b) het zoeken naar een andere donor om de donor met een te lage Hb-waarde te vervangen, en c) een verminderde motivatie bij de uitgestelde donoren om terug te keren naar de bloedbank.

In dit proefschrift hebben we diverse, en vooral Bayesiaanse, statistische methoden ontwikkeld voor het modelleren van longitudinale waarnemingen van Hb-waarden bij bloeddonoren. **Hoofdstuk 1** geeft een introductie tot bloeddonatie en een overzicht van de klinische en statistische onderzoeksvragen van het onderzoek in dit proefschrift. In **Hoofdstuk 2** vergelijken we verschillende statistische modellen (transitiemodellen met verschillende aantallen autoregressieve termen en gemengde modellen) voor het modelleren van de afhankelijkheid van opeenvolgende Hb-waarden van een bloeddonor. Tevens worden deze modellen gebruikt voor het voorspellen van toekomstige Hb-waarden van een donor. Dit hoofdstuk illustreert hoe voorspellingen kunnen worden gedaan met longitudinale modellen, en hoe de conclusies van dergelijke methodes kunnen bijdragen tot een reductie van het aantal uitgestelde bloeddonaties.

In **Hoofdstuk 3** laten we zien dat het verloop van de Hb-waarden over de tijd in de donorpopulatie niet met slechts één patroon of wiskundige functie kan worden gemodelleerd. Hiervoor hebben we een gemengd groeimodel (growth mixture model) ontwikkeld dat ermee rekening houdt dat de donorpopulatie bestaat uit verschillende groepen donoren met een verschillend verloop van de Hb-waarden. Deze statistische methode lijkt geschikt voor het modelleren van interindividuele verschillen in het longitudinale verloop van Hb-waarden van donoren. De methode kan worden gebruikt voor het identificeren van donoren voor wie de tijdsintervallen tussen donaties moeten worden aangepast om alzo ijzertekorten en uitgestelde donaties te voorkomen.

Hoofdstuk 4 beschrijft een gemengd transitiemodel met latente klassen voor het voorspellen van de toekomstige Hb-waarde van een potentiële bloeddonor uitgaande van eerder gemeten Hb-waarden. Een belangrijk voordeel van dit model is dat de heterogeniteit en de afhankelijkheid van de Hb-waarden gezamenlijk worden gemodelleerd. De heterogeniteit in het herstelproces van de Hb-waarde wordt beschreven met latente klassen, en de dynamiek van het herstelproces wordt beschreven met een flexibele functie en een kantelpunt (change point) dat afhangt van het aantal donaties. Deze flexibele functie stelt ons in staat de hersteltijd, dat wil zeggen de tijd totdat de Hb-waarde hersteld is tot het niveau voor de donatie, te schatten. Dit is van groot praktisch belang, omdat schattingen van de hersteltijd kunnen worden gebruikt om de planning van de bezoeken van bloeddonoren aan de bloedbank te optimaliseren, en dus ijzertekorten en uitgestelde donaties te voorkomen.
In **Hoofdstuk 5** hebben we een uitgebreide simulatiestudie gedaan voor het bepalen van het aantal latente klassen in latenteklassenmodellen. Met andere woorden, we wensten na te gaan hoe goed subgroepen in een populatie ontdekt kunnen worden met behulp van Bayesiaanse technieken. Deze simulatiestudie is gebruikt voor het ontwikkelen van praktische richtlijnen voor het kiezen van het aantal latente klassen met een criterium op basis van de posteriorverdeling van de klassengroottes in een Bayesiaans latenteklassenmodel. In de simulatiestudie zijn de prestaties van dit criterium vergeleken met andere criteria. Het nut en de praktische toepasbaarheid van het criterium op basis van de posteriorverdeling van de klassengroottes wordt geïllustreerd aan de hand van een dataset met herhaalde metingen van Hb-waarden van bloeddonoren.
Ten slotte geeft **Hoofdstuk 6** een samenvatting van de belangrijkste bevindingen van dit proefschrift en een discussie van mogelijke verbeteringen en uitbreidingen van de gebruikte statistische technieken.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my both promotors Emmanuel Lesaffre and Wim de Kort for the continuous support during my Ph.D. project, for their encouragement, insightful comments, and hard questions. Emmanuel, *a priori* I am forever indebted that you accepted to be my promotor, you provided an extremely friendly atmosphere on a professional and personal level, *a posteriori* words are insufficient to express my gratitude. Wim, I am forever thankful for your encouragement and support, your guidance and ideas were always invaluable inputs for this thesis.

Besides my promotors, I would like to express my very great appreciation to my co-promotor Joost van Rosmalen. Joost, thank you for your consistent support throughout my research period in Biostatistics department. Also a very special thanks goes to Dimitris Rizopoulos, who taught me advanced R programming and was always supportive throughout my Ph.D. project. It is a pleasure to thank Mireille Baart and Katja van den Hurk for their generous support for providing the data and answering my clinical questions.

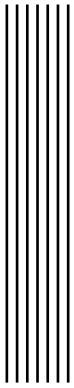
I would like to express my sincere thanks to my paranymphs, Johan de Rooi and Bianca Knotter. I really appreciate that you accepted to be my paranymphs. I would like also to thank all colleagues and friends at Erasmus MC and Sanquin, especially Nicole Erler, Ehsan Motazedi, Sanaz Sedaghat, Bayoue Li, Susan Bryan, Saber Akhondi, Elrozy Andrinopoulou, Sten Willemsen, Greg Papageorgiou, Tiffany Timmer, and of course the secretary of Biostatistics department, Eline van Gent. Thanks goes also to the staff employee of human resources department, Andreas Engelhard and Solange Degenaar thanks for your patience and support.

Veel dank ook aan Ronnie van der Holt, voor zijn steun binnen mijn huidige functie als statisticus bij het clinical trial center.

My special thanks are extended to Norbert Welu and Di Iulio Andrea in Luxembourg Taekwondo federation for their continuous and unconditional support.

I wish to thank my parents(-in-law) for their support and encouragement throughout my study. And last but not the least, I would like to thank my wife, Nahid Mostafavi, who has been extremely supportive throughout my Ph.D. project and has made countless sacrifices to help me get to this point.

Thank you very much, everyone!
Kazem Nasserinejad
July 2016 - Rotterdam



Ph.D. portfolio and Curriculum Vitae

Ph.D. Portfolio

Name Ph.D. student: **drs. K. Nasserinejad**	Ph.D. period: **2011-2016**	
Erasmus MC Department: **Biostatistics**	Promotors: **Prof. dr. E. Lesaffre and Prof. dr. W. de Kort**	
	Co-promotor: **dr. J. van Rosmalen**	
1) Ph.D. training	**Year**	**ECTS**
Courses		
Joint latent class mixed models	2015	0.3
GAMLSS in action	2014	0.9
Joint models for longitudinal and survival data	2013	0.7
Intro. to Bayesian methods in clinical research	2012	1.4
Repeated measurements	2012	1.4
Bayesian statistics	2012	6.0
Bayesian variable selection	2012	0.3
Bayesian adaptive methods for clinical trials	2012	0.7
Statistical computing with R	2012	6.0
Seminars and workshops		
Erasmus MC biweekly Biostatistics seminar	2012-2015	1.0
Presentations		
Sanquin monthly Bioinformatics seminar	2015	1.0
Shahid Beheshti monthly Biostatistics seminar	2014	1.0
Erasmus MC biweekly Biostatistics seminar	2012-2015	1.0
International conferences		
15th CEB - EIB conference (oral presentation)	2015	1.0
Bayes2015 conference (oral presentation)	2015	1.0
8th EMR-IBS conference (oral presentation)	2015	1.0
Bayes2014 conference (oral presentation)	2014	1.0
35th ISCB conference (oral presentation)	2014	1.0
1st ECDHM conference (oral presentation)	2014	1.0
34th ISCB conference (oral presentation)	2013	1.0
Bayes2013 conference	2013	1.0
33rd ISCB conference (poster presentation)	2012	1.0
2) Teaching		
Assistant teaching		
Repeated measurements, NIHES program	2013/2014/2015	2.1
Intro. to Bayesian methods in clinical research, NIHES program	2013/2014/2015	2.1
Biostatistical methods II, NIHES program	2012/2013/2014/2015	2.8
Biostatistical methods I, NIHES program	2012/2013/2014/2015	2.8
Introduction to epidemiology, NIHES program	2012	0.7
SPSS practicals for medical student, Erasmus MC	2012/2013/2014/2015	2.8
Consulting		
Statistical consultant in the Erasmus MC CPO system	2012-2015	10.0
Total		**54.0**

Curriculum Vitae

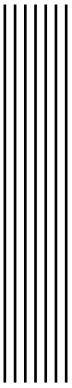
Kazem Nasserinejad	**2 November 1984, Iran**
Education	**Year**
BSc. Statistics, Shahid Beheshti University, Tehran	2003-2007
MSc. Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran	2008-2011
Ph.D. Biostatistics, Erasmus University Medical Center, Rotterdam	2011-2016
Scientific membership	
International Biometric Society (IBS)	since 2014
International Society for Clinical Biostatistics (ISCB)	since 2012
Iranian statistical society (ISC)	since 2015
Statistical advisor	
Developmental Medicine and Child Neurology Journal	since 2015
Honors & Awards	
Student award 15th CEB - EIB conference, Bilbao, Spain	2015
Sport	
5th Degree Taekwondo Black Belt	2011
Taekwondo international referee	2015

List of publications

- van Lee CB, Ip Vai Ching EE, **Nasserinejad K**, Neumann HA, Bol MG, Dikrama PK, Kelleners-Smeets NW, Koljenović S, Munte K, Noordhoek Hegt V, de Vijlder HC, Nijsten T, van den Bos RR. Reliability of Mohs slides diagnosis: interpersonal and intrapersonal agreement on basal cell carcinoma presence and histological subtype. *British Journal of Dermatology*. 2016.
- Stipdonk LW, Weisglas-Kuperus N, Franken MC, **Nasserinejad K**, Dudink J, Goedegebure A. Auditory brainstem maturation in normal-hearing preterm infants: A meta-analysis. *Developmental Medicine and Child Neurology*. 2016.
- **Nasserinejad K**, van Rosmalen J, de Kort W, Rizopoulos D, Lesaffre E. Prediction of hemoglobin in blood donors using a latent class mixed-effects transition model. *Statistics in Medicine*. 2016 ;35(4):581-94.
- van Tiel J, Siebelt M, Reijman M, Bos PK, Waarsing JH, Zuurmond AM, **Nasserinejad K**, van Osch GJ, Verhaar JA, Krestin GP, Weinans H, Oei EH. Quantitative in vivo CT arthrography of the human osteoarthritic knee to estimate cartilage sulphated glycosaminoglycan content: correlation with ex-vivo reference standards. *Osteoarthritis and Cartilage*. 2016.
- van Tiel J, Kotek G, Reijman M, Bos PK, E. Bron EE, Klein S, **Nasserinejad K**, van Osch GJ, Verhaar JA, Krestin GP, Weinans H, Oei EH. Is T1rho-mapping an alternative to delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) in assessing sulphated glycosaminoglycan content in human osteoarthritic knees? An in vivo validation study. *Radiology*. 2015.
- van Lee CB, Graafland B, Koljenović S, Neumann HA, **Nasserinejad K**, Nijsten TE, van den Bos RR, Munte K. Additional review of Mohs slides to optimize Mohs micrographic surgery. *British Journal of Dermatology*. 2015 ;173(1):123-7.
- **Nasserinejad K**, van Rosmalen J, van den Hurk K, Baart M, Hoekstra T, Rizopoulos D, Lesaffre E, de Kort W. Prevalence and determinants of declining versus stable hemoglobin levels in whole blood donors. *Transfusion*. 2015 ;55(8):1955-63.
- Camfferman FA, Ecury-Goossen GM, La Roche JE, de Jong N, van 't Leven W, Vos HJ, Verweij MD, **Nasserinejad K**, Cools F, Govaert P, Dudink J. Calibrating Doppler imaging of preterm intracerebral circulation using a microvessel flow phantom. *Frontiers in Human Neuroscience*. 2013.
- van den Heuvel D, Jansen MAE, **Nasserinejad K**, Dik WA, van Lochem EG, Bakker-Jonges LE, Bouallouch-Charif H, Jaddoe VWV, Hooijkaas H, van Dongen JJM, Moll HA, van Zelm MC. Effects of external determinants on age-related patterns of innate leukocyte and naive and memory B- and T-lymphocyte numbers in early childhood: The generation R-study. Ph.D. Thesis (Chapter 2), Erasmus MC, 2015.

- Runia TF, van Meurs M, **Nasserinejad K**, Hintzen RQ. No evidence for an association of osteopontin plasma levels with disease activity in multiple sclerosis. *Multiple Sclerosis Journal*. 2014 31:1352458514528765.
- Vahabi N, Salehi M, Zayeri F, Torabzadeh H, **Nasserinejad K**, Razmavar S. Comparison of longitudinal data models for hygroscopic expansion of three common composites. *Razi Journal of Medical Sciences*. 2013.
- Gijsen VM, Dubé MP, Hesselink DA, Ross CJ, Mital S, 't Jong GW, van Schaik RHN, Dalinghaus M, Bogers AJ, Grant D, Provost S, **Nasserinejad K**, van Rosmalen J, Verjee Z, Carleton B, Hayden MR, Tibboel D, Koren G, de Wildt SN. Investigating recipients genetic variability in predicting deterioration of renal function among pediatric heart transplant recipients receiving tacrolimus. Tacrolimus: The balance between good and evil. Ph.D. Thesis (Chapter 9), Erasmus MC, 2013.
- Gijsen VM, Hesselink DA, 't Jong GW, van Schaik RHN, Ng V, Grant D, Feroz-Zada Y, **Nasserinejad K**, van Rosmalen J, Verjee Z, Ross CJ, Carleton B, Hayden MR, Tibboel D, Dubé MP, Koren G, de Wildt SN. Genetic variation in relation to renal function among pediatric liver transplant recipients receiving tacrolimus. Tacrolimus: The balance between good and evil. Ph.D. Thesis (Chapter 8), Erasmus MC, 2013.
- Gijsen VM, Hesselink DA, Ross C, Dubé MP, van Schaik RH, Van Gelder T, Cransberg K, 't Jong GW, Feroz-Zada Y, **Nasserinejad K**, van Rosmalen J, Hebert D, Grant D, Verjee Z, Carlton B, Hayden MR, Tibboel D, de Wildt SN, Koren G. Recipient genetic variation does not influence renal function in pediatric kidney transplantation patients receiving tacrolimus. Tacrolimus: The balance between good and evil. Ph.D. Thesis (Chapter 7), Erasmus MC, 2013.
- **Nasserinejad K**, de Kort W, Baart M, Komrek A, van Rosmalen J, Lesaffre E. Predicting hemoglobin levels in whole blood donors using transition models and mixed effects models. *BMC Medical Research Methodology*. 2013;13(1):62.
- Zayeri F, Razmavar S, **Nasserinejad K**, Akbarzadeh-Baghban A, Torabzadeh H, Salehi M. Experimental Comparison of Hygroscopic Expansion in Three Different Composite Resins. *Journal of Islamic Dental Association of Iran*. 2012.
- Kuo W, de Bruijne M, Petersen J, **Nasserinejad K**, Ozturk H, Chen Y, Perez-Rovira A, Tiddens AWM. Diagnosis of bronchiectasis and airway wall thickening in children with cystic fibrosis: Objective airway-artery quantification. Submitted.
- Gijsen VM, van Rosmalen J, Dalinghaus M, Hesselink DA, Mital S, Ng V , Cransberg K, Andrews LM, Finkelstein Y, Pollock-BarZiv SM, Hebert D, 't Jong GW, **Nasserinejad K**, Bogers AJ, Grant D, Verjee Z, Tibboel D, Koren G, de Wildt SN. Long-term renal function in pediatric heart, liver or kidney transplant recipients receiving tacrolimus. Submitted.

- van den Heuvel D, Jansen MAE, **Nasserinejad K**, Dik WA, van Lochem EG, Bakker-Jonges LE, Bouallouch-Charif H, Jaddoe VWV, Hooijkaas H, van Dongen JJM, Moll HA, van Zelm MC. Effects of external determinants on age-related patterns of innate leukocyte and naive and memory B- and T-lymphocyte numbers in early childhood: The generation R-study. Submitted.
- **Nasserinejad K**, van Rosmalen J, de Kort W, Lesaffre E. Comparison of criteria for choosing the number of classes in Bayesian finite mixture models. Submitted.
- Hywood J, Teixeira-Pinto A, **Nasserinejad K**, McCusker E, Loy C. Predicting individual disease progression in huntington disease using mixed models and transition models. Submitted.
- Motazedi E, **Nasserinejad K**, van Rosmalen J, Lesaffre E. The effect of the prior and computational method in Bayesian variance component models. Submitted.



Bibliography

Aitkin, M. and Alfò, M. (2003). Longitudinal analysis of repeated binary data using autoregressive and random effect modelling. *Statistical Modelling*, 3(4):291–303.

Akaike, H. (1973). Information theory and extension of maximum likelihood principle. *Second International, Symposium on Information Theory Pctrov B Csaki F (cd)*.

Alstead, S. (1943). Rate of blood-regeneration after haemorrhage. *The Lancet*, 241(6240):424–426.

Amrein, K., Valentin, A., Lanzer, G., and Drexler, C. (2012). Adverse events and safety issues in blood donationa comprehensive review. *Blood Reviews*, 26(1):33–42.

Anderson, T. W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82.

Asparouhov, T. and Muthén, B. (2011). Using bayesian priors for more flexible latent class analysis. In *Proceedings of the 2011 Joint Statistical Meeting, Miami Beach, FL*.

Atsma, F., Veldhuizen, I., de Vegt, F., Doggen, C., and de Kort, W. (2011). Cardiovascular and demographic characteristics in whole blood and plasma donors: results from the donor insight study. *Transfusion*, 51(2):412–420.

Baart, A., de Kort, W., Atsma, F., Moons, K., and Vergouwe, Y. (2012). Development and validation of a prediction model for low hemoglobin deferral in a large cohort of whole blood donors. *Transfusion*, 52(12):2559–2569.

Baart, A., de Kort, W., Moons, K., and Vergouwe, Y. (2011). Prediction of low haemoglobin levels in whole blood donors. *Vox Sanguinis*, 100:204–211.

Baart, M. (2013). *Prediction models for hemoglobin deferral in whole blood donors*. PhD thesis, Utrecht University, Utrecht, the Netherlands.

Beard, J. and Tobin, B. (2000). Iron status and exercise. *The American Journal of Clinical Nutrition*, 72(2):594s–597s.

Bernardo, J. M. (1979). Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 113–147.

Bhargava, A., Sargan, J. D., et al. (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica*, 51(6):1635–1659.

Birgegård, G., Schneider, K., and Ulfberg, J. (2010). High incidence of iron depletion and restless leg syndrome (rls) in regular blood donors: intravenous iron sucrose substitution more effective than oral iron. *Vox Sanguinis*, 99(4):354–361.

Boulton, F. (2004). Managing donors and iron deficiency. *Vox Sanguinis*, 87(s2):22–24.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. CRC press.

Brittenham, G. M. (2011). Iron deficiency in whole blood donors. *Transfusion*, 51(3):458–461.

Brooks, S. P., Giudici, P., and Roberts, G. O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):3–39.

Brussaard, J., Brants, H., Bouman, M., and Löwik, M. (1997). Iron intake and iron status among adults in the netherlands. *European Journal of Clinical Nutrition*, 51:S51–8.

Bryant, B. J., Yau, Y. Y., Arceo, S. M., Hopkins, J. A., and Leitman, S. F. (2013). Ascertainment of iron deficiency and depletion in blood donors through screening questions for pica and restless legs syndrome. *Transfusion*, 53(8):1637–1644.

Cable, R. G., Glynn, S. A., Kiss, J. E., Mast, A. E., Steele, W. R., Murphy, E. L., Wright, D. J., Sacher, R. A., Gottschall, J. L., Vij, V., and Simon, T. L. (2011). Iron deficiency in blood donors: analysis of enrollment data from the reds-ii donor iron status evaluation (rise) study. *Transfusion*, 51(3):511–522.

Carlin, B. P. and Louis, T. A. (1997). Bayes and empirical bayes methods for data analysis. *Statistics and Computing*, 7(2):153–154.

Celeux, G. (1998). Bayesian inference for mixture: The label switching problem. In *Compstat*, pages 227–232. Springer.

Celeux, G., Forbes, F., Robert, C. P., Titterington, D. M., et al. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4):651–673.

Celeux, G., Hurn, M., and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970.

Custer, B., Bravo, M., Bruhn, R., Land, K., Tomasulo, P., and Kamel, H. (2014). Predictors of hemoglobin recovery or deferral in blood donors with an initial successful donation. *Transfusion*, 54(9):2267–2275.

Custer, B., Chinn, A., Hirschler, N., Busch, M. P., and Murphy, E. (2007). The consequences of temporary deferral on future whole blood donation. *Transfusion*, 47(8):1514–1523.

Cvejic, A., Haer-Wigman, L., Stephens, J. C., Kostadima, M., Smethurst, P. A., Frontini, M., van den Akker, E., Bertone, P., Bielczyk-Maczyńska, E., Farrow, S., et al. (2013). Smim1 underlies the vel blood group and influences red blood cell traits. *Nature Genetics*, 45(5):542–545.

Dallman, P. R. (1986). Biochemical basis for the manifestations of iron deficiency. *Annual Review of Nutrition*, 6(1):13–40.

Dellaportas, P. and Papageorgiou, I. (2006). Multivariate mixtures of normals with unknown number of components. *Statistics and Computing*, 16(1):57–68.

Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002). *Analysis of Longitudinal Data 2nd ed*. New York: Oxford University Press.

Earnest, A., Chen, M. I., Ng, D., and Sin, L. Y. (2005). Using autoregressive integrated moving average (arima) models to predict and monitor the number of beds occupied during a sars outbreak in a tertiary hospital in singapore. *BMC Health Services Research*, 5(1):36.

Elliott, M. R., Gallo, J. J., Ten Have, T. R., Bogner, H. R., and Katz, I. R. (2005). Using a bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction. *Biostatistics*, 6(1):119–143.

Feller, W. (1943). On a general class of contagious distributions. *The Annals of Mathematical Statistics*, 14(4):389–400.

Finch, C. A., Cook, J., Labbe, R., and Culala, M. (1977). Effect of blood donation on iron stores as evaluated by serum ferritin. *Blood*, 50(3):441–447.

Fowler, W. M. and Barer, A. P. (1942). Rate of hemoglobin regeneration in blood donors. *Journal of the American Medical Association*, 118(6):421–427.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, 97(458):611–631.

Frühwirth-Schnatter, S. (2001). Markov chain monte carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453):194–209.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and markov switching models using bridge sampling techniques. *The Econometrics Journal*, 7(1):143–167.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Science & Business Media.

Funatogawa, I. and Funatogawa, T. (2012). An autoregressive linear mixed effects model for the analysis of unequally spaced longitudinal data with dose-modification. *Statistics in Medicine*, 31(6):589–599.

Funatogawa, I., Funatogawa, T., and Ohashi, Y. (2007). An autoregressive linear mixed effects model for the analysis of longitudinal data which show profiles approaching asymptotes. *Statistics in Medicine*, 26(9):2113–2130.

Garrett, E. S. and Zeger, S. L. (2000). Latent class model diagnosis. *Biometrics*, 56(4):1055–1067.

Garry, P. J., Koehler, K. M., and Simon, T. L. (1995). Iron stores and iron absorption: effects of repeated blood donations. *The American Journal of Clinical Nutrition*, 62(3):611–620.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis*, volume 2. Taylor & Francis.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4):733–760.

Geweke, J. et al. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department.

Gómez-Simón, A., Navarro-Núñez, L., Pérez-Ceballos, E., Lozano, M. L., Candela, M. J., Cascales, A., Martínez, C., Corral, J., Vicente, V., and Rivera, J. (2007). Evaluation of four rapid methods for hemoglobin screening of whole blood donors in mobile collection settings. *Transfusion and Apheresis Science*, 36(3):235–242.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732.

Griffin, J. E., Brown, P. J., et al. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188.

Halperin, D., Baetens, J., and Newman, B. (1998). The effect of short-term, temporary deferral on future blood donation. *Transfusion*, 38(2):181–183.

Hanley, J. and McNeil, B. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148:839–843.

Harvey, A. and Phillips, G. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, 66:49–58.

Harville, D. (1976). Extension of the gauss-markov theorem to include the estimation of random effects. *The Annals of Statistics*, pages 384–395.

Hawkins, D. M. (2001). Fitting multiple change-point models to data. *Computational Statistics and Data Analysis*, 37(3):323–341.

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

Heckman, J. J. (1981). Heterogeneity and state dependence. In *Studies in labor markets*, pages 91–140. University of Chicago Press.

Hillgrove, T., Moore, V., Doherty, K., and Ryan, P. (2011). The impact of temporary deferral due to low hemoglobin: future return, time to return, and frequency of subsequent donation. *Transfusion*, 51(3):539–547.

Hoekstra, T., Veldhuizen, I., van Noord, P., and de Kort, W. (2007). Seasonal influences on hemoglobin levels and deferral rates in whole-blood and plasma donors. *Transfusion*, 47(5):895–900.

Homan, M. D. and Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *The Journal of Machine Learning Research*, 15(1):1593–1623.

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics*, 109(1):107–150.

Inman, H. F. and Bradley Jr, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics-Theory and Methods*, 18(10):3851–3874.

Ishwaran, H., James, L. F., and Sun, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *Journal of the American Statistical Association*, 96(456).

Jasra, A., Holmes, C., and Stephens, D. (2005). Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*, pages 50–67.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 186, pages 453–461. The Royal Society.

Johnson-Spear, M. A. and Yip, R. (1994). Hemoglobin difference between black and white women with comparable iron status: justification for race-specific anemia criteria. *The American Journal of Clinical Nutrition*, 60(1):117–121.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

Kazemi, I. and Crouchley, R. (2006). Modelling the initial conditions in dynamic regression models of panel data with random effects. *Contributions to Economic Analysis*, 274:91–117.

Kazemi, I. and Davies, R. (2002). The asymptotic bias of mles for dynamic panel data models. *Statistical modelling in society, Proceedings of the 17th IWSM, Chania, Greece*, pages 391–395.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 49–66.

Kiss, J. E., Brambilla, D., Glynn, S. A., Mast, A. E., Spencer, B. R., Stone, M., Kleinman, S. H., and Cable, R. G. (2015). Oral iron supplementation after blood donation: a randomized clinical trial. *The Journal of the American Medical Association*, 313(6):575–583.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2 (IJCAI'95), Vol. 2*, pages 1137–1143.

Komárek, A., Hansen, B. E., Kuiper, E. M., van Buuren, H. R., and Lesaffre, E. (2010). Discriminant analysis using a multivariate linear mixed model with a normal mixture in the random effects distribution. *Statistics in Medicine*, 29(30):3267–3283.

Kristal-Boneh, E., Froom, P., Harari, G., and Ribak, J. (1997). Seasonal differences in blood cell parameters and the association with cigarette smoking. *Clinical and Laboratory Haematology*, 19(3):177–181.

Kristal-Boneh, E., Froom, P., Harari, G., Shapiro, Y., and Green, M. (1993). Seasonal changes in red blood cell parameters. *British journal of haematology*, 85(3):603–607.

Lee, K., Marin, J.-M., Mengersen, K., and Robert, C. (2008). Bayesian inference on mixtures of distributions. *Proceedings of the Platinum Jubilee of the Indian Statistical Institute*, 776.

Lesaffre, E. and Lawson, A. B. (2012). *Bayesian biostatistics*. John Wiley & Sons.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337.

Malsiner-Walli, G., Frühwirth-Schnatter, S., and Grün, B. (2016). Model-based clustering based on sparse finite gaussian mixtures. *Statistics and Computing*, 26(1-2):303–324.

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013). Bayesian computing with inla: new features. *Computational Statistics and Data Analysis*, 67:68–83.

McCulloch, C. E. and Neuhaus, J. M. (2001). *Generalized Linear Mixed Models*. Wiley Online Library.

McGrory, C. A. and Titterington, D. (2007). Variational approximations in bayesian model selection for finite mixture distributions. *Computational Statistics and Data Analysis*, 51(11):5352–5367.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Micozzi, M. S., Albanes, D., and Stevens, R. G. (1989). Relation of body size and composition to clinical biochemical and hematologic indices in us men and women. *The American Journal of Clinical Nutrition*, 50(6):1276–1281.

Molenberghs, G. and Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling*, 1(4):235–269.

Murphy, E. L., Shaz, B., Hillyer, C. D., Carey, P., Custer, B. S., Hirschler, N., Fang, J., and Schreiber, G. B. (2009). Minority and foreign-born representation among us blood donors: demographics and donation frequency for 2006. *Transfusion*, 49(10):2221–2228.

Murphy, W. G. (2002). Disease transmission by blood products: Past, present and future. *Pathophysiology of haemostasis and thrombosis*, 32(Suppl. 1):1–4.

Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the em algorithm. *Biometrics*, 55(2):463–469.

Nagin, D. S. and Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, 6:109–138.

Nasserinejad, K., de Kort, W., Baart, M., Komárek, A., van Rosmalen, J., and Lesaffre, E. (2013). Predicting hemoglobin levels in whole blood donors using transition models and mixed effects models. *BMC Medical Research Methodology*, 13(1):62.

Nasserinejad, K., van Rosmalen, J., de Kort, W., Rizopoulos, D., and Lesaffre, E. (2016). Prediction of hemoglobin in blood donors using a latent class mixed-effects transition model. *Statistics in Medicine*, 35(4):581–594.

Nasserinejad, K., van Rosmalen, J., van den Hurk, K., Baart, M., Hoekstra, T., Rizopoulos, D., Lesaffre, E., and de Kort, W. (2015). Prevalence and determinants of declining versus stable hemoglobin levels in whole blood donors. *Transfusion*, 55(8):1955–1963.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change.

Newman, B. H. (2004). Adjusting our management of female blood donors: the key to an adequate blood supply. *Transfusion*, 44(4):591–596.

Newman, B. H., Newman, D. T., Ahmad, R., and Roth, A. J. (2006). The effect of whole-blood donor adverse events on blood donor return rates. *Transfusion*, 46(8):1374–1379.

Nobile, A. (2004). On the posterior distribution of the number of components in a finite mixture. *Annals of Statistics*, pages 2044–2073.

Nobile, A. and Fearnside, A. T. (2007). Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Statistics and Computing*, 17(2):147–162.

Pan, J.-C. and Huang, G.-H. (2014). Bayesian inferences of latent class models with an unknown number of classes. *Psychometrika*, 79(4):621–646.

Pasricha, S.-R., McQuilten, Z. K., Keller, A. J., and Wood, E. M. (2011). Hemoglobin and iron indices in nonanemic premenopausal blood donors predict future deferral from whole blood donation. *Transfusion*, 51(12):2709–2713.

Plummer, M. (2011). rjags: Bayesian graphical models using mcmc. *R package version*, 2(0).

Popovsky, M. A. (2012). Anemia, iron depletion, and the blood donor: it's time to work on the donor's behalf. *Transfusion*, 52(4):688–692.

R Development Core Team (2010). R: A language and environment for statistical computing.

Radtke, H., Polat, G., Kalus, U., Salama, A., and Kiesewetter, H. (2005). Hemoglobin screening in prospective blood donors: comparison of different blood samples and different quantitative methods. *Transfusion and Apheresis Science*, 33(1):31–35.

Raftery, A. E. et al. (1996). Hypothesis testing and model selection via posterior simulation. *Markov chain Monte Carlo in practice*, pages 163–188.

Richardson, S. and Green, P. J. (1997). On bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 731–792.

Rikhtehgaran, R., Kazemi, I., Verbeke, G., de Kort, W., and Lesaffre, E. (2012). Piecewise transition models with random effects for unequally spaced longitudinal measurements. *Statistical Modelling*, 12(6):503–525.

Ripley, B. D. (1988). Uses and abuses of statistical simulation. *Mathematical Programming*, 42(1-3):53–68.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., and Muller, M. (2011). proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Informatics*, 12:77.

Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):689–710.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Schiepers, O. J., van Boxtel, M. P., de Groot, R. H., Jolles, J., de Kort, W. L., Swinkels, D. W., Kok, F. J., Verhoef, P., and Durga, J. (2010). Serum iron parameters, hfe c282y genotype, and cognitive performance in older adults: results from the facit study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, page glq149.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Simon, T. L. (2002). Iron, iron everywhere but not enough to donate. *Transfusion*, 42(6):664–664.

Simon, T. L., Garry, P. J., and Hooper, E. M. (1981). Iron stores in blood donors. *The Journal of the American Medical Association*, 245(20):2038–2043.

Skikne, B., Lynch, S., Borek, D., and Cook, J. (1984). Iron and blood donation. *Clinics in Haematology*, 13(1):271–287.

Skjelbakken, T., Dahl, I., Wilsgaard, T., Langbakk, B., and Lochen, M. (2006). Changes in haemoglobin levels according to changes in body mass index and smoking habits, a 20-year follow-up of a male cohort: the tromso study 1974-1995. *Eur J Epidemiol*, 21(7):493–499.

Soranzo, N., Spector, T. D., Mangino, M., Kühnel, B., Rendon, A., Teumer, A., Willenborg, C., Wright, B., Chen, L., Li, M., et al. (2009). A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the haemgen consortium. *Nature Genetics*, 41(11):1182–1190.

Spencer, B. R., Kleinman, S., Wright, D. J., Glynn, S. A., Rye, D. B., Kiss, J. E., Mast, A. E., and Cable, R. G. (2013). Restless legs syndrome, pica, and iron status in blood donors. *Transfusion*, 53(8):1645–1652.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Stan (2015). Stan: A C++ Library for Probability and Sampling, Version 2.8.0.

Steele, R. J. and Raftery, A. (2010). Performance of bayesian model selection criteria for gaussian mixture models. *Frontiers of Statistical Decision Making and Bayesian Analysis*, pages 113–130.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809.

Thomas, A. and OHara, R. B. (2004). Openbugs.

Tofighi, D. and Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. *Advances in latent variable mixture models*, (Information Age Publishing, Inc):317–341.

Tong, E., Murphy, W., Kinsella, A., Darragh, E., Woods, J., Murphy, C., and McSweeney, E. (2010). Capillary and venous haemoglobin levels in blood donors: a 42-month study of 36 258 paired samples. *Vox Sanguinis*, 98:266–267.

Tsay, R. (2010). *Analysis of Financial Time Series, 3rd Edition*. New York: John Wiley & Sons.

Ulfberg, J. and Nyström, B. (2004). Restless legs syndrome in blood donors. *Sleep Medicine*, 5(2):115–118.

van der Harst, P., Zhang, W., Leach, I. M., Rendon, A., Verweij, N., Sehmi, J., Paul, D. S., Elling, U., Allayee, H., Li, X., et al. (2012). Seventy-five genetic loci influencing the human red blood cell. *Nature*, 492(7429):369–375.

van Havre, Z., White, N., Rousseau, J., and Mengersen, K. (2015). Overfitting bayesian mixture models with an unknown number of components. *PloS One*, 10(7):e0131739.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Wadsworth, G. (1955). Recovery from acute haemorrhage in normal men and women. *The Journal of Physiology*, 129(3):583–593.

Wang, M. and Bodner, T. E. (2007). Growth mixture modeling identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods*, 10(4):635–656.

Wasserman, L. (2000). Asymptotic inference for mixture models using data-dependent priors. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, pages 159–180.

Western, B. and Kleykamp, M. (2004). A bayesian change point model for historical time series analysis. *Political Analysis*, pages 354–374.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20(1):39–54.

Yip, R., Johnson, C., and Dallman, P. R. (1984). Age-related changes in laboratory values used in the diagnosis of anemia and iron deficiency. *The American Journal of Clinical Nutrition*, 39(3):427–436.

Ziemann, M., Lizardo, B., Geusendam, G., and Schlenke, P. (2011). Reliability of capillary hemoglobin screening under routine conditions. *Transfusion*, 51(12):2714–2719.

Zou, S., Musavi, F., Notari, E. P., Rios, J. A., Trouern, J., and Fang, C. T. (2008). Donor deferral and resulting donor loss at the american red cross blood services, 2001 through 2006. *Transfusion*, 48(12):2531–2539.