Familial Tendency to Fetal Loss Analyzed with Bayesian Graphical Models by Gibbs Sampling

Heidi H. Hundborg
Department of Biostatistics, University of Aarhus,
Vennelyst Boulevard 6, DK-8000 Aarhus, Denmark
e-mail: heidi@biostat.au.dk

Malene Højbjerre
Department of Mathematical Sciences, Aalborg University,
Fredrik Bajers Vej 7E, DK-9220 Aalborg, Denmark
e-mail: malene@math.auc.dk

Ole Bjarne Christiansen
Department of Obstetrics and Gynaecology, Aalborg Hospital,
P.O. Box 561, DK-9100 Aalborg, Denmark
e-mail: obchr@post5.tele.dk

Steffen L. Lauritzen
Department of Mathematical Sciences, Aalborg University,
Fredrik Bajers Vej 7E, DK-9220 Aalborg, Denmark
e-mail: steffen@math.auc.dk

This is a preprint of an article published in
Statistics in Medicine 2000; **19**:2147-2168

October 25, 1999

Familial Tendency to Fetal Loss Analyzed with Bayesian Graphical Models by Gibbs Sampling

Summary

This paper presents several models for investigating whether the HLA allogenotypes DR1/Br, DR3 and DR10 are genetic markers for a predisposition of experiencing unexplained recurrent fetal losses. A total of 199 women from 113 families answered questionnaires concerning their pregnancies and 145 of these women were HLA typed. The analysis of the data is complicated as dependencies between pregnancy outcomes are expected.

The main purpose of the paper is to illustrate how such analyses can be performed using Bayesian graphical models and Gibbs sampling. The analyses are made using the programs BUGS and CODA. Markov chain Monte Carlo analyses within a Bayesian framework have become easier with the introduction of these programs. However, experience shows that some caution is required so we recommend making some initial analyses using very simple models and perhaps approximative methods, followed by a model development introducing more and more complexity.

1 Introduction

This paper presents analyses of a data set established by one of the authors (O.B. Christiansen) to evaluate the heritability of an increased risk of fetal loss. A more detailed discussion of the medical issues is given in Christiansen *et al.* [1].

The study focuses on occurrence of fetal losses among sisters and sisters-in-law of patients with idiopathic abortus habitualis (IAH). Abortus habitualis is defined as the phenomenon that a woman has had three or more consecutive fetal losses and idiopathic means that the fetal losses occur for unexplained reasons.

Previous studies of women suffering from IAH provide some evidence of a familial tendency to fetal loss as the frequency of inexplicable fetal losses among sisters of IAH patients is increased [2]. Moreover, it has been suggested that the HLA allogenotypes DR1/Br, DR3 and DR10 are genetic markers for the risk of experiencing unexplained fetal losses [3, 4]. To further clarify the heritability of fetal loss and the role of these HLA allogenotypes, sisters and sisters-in-law of IAH patients have been identified and questioned about the outcome of previous pregnancies. Furthermore, some of these women were also HLA typed.

A statistical analysis of such data is complicated by the fact that some dependencies between pregnancy outcomes from the same woman must be expected and perhaps also dependencies between pregnancy outcomes from different women in the same family. In a conventional approach, using the pregnancy as the unit of analysis, these dependencies are ignored and the inference may therefore be inappropriate. In the present paper we report results from analyses of models that allow for such dependencies. These models are directed graphical models [5, 6, 7] and are analyzed by a Bayesian approach using Gibbs sampling [8, 9]. To perform the analyses we use the programs BUGS (Bayesian analysis Using Gibbs Sampling) [10, 11, 12] and CODA (COnvergence Diagnostics and output Analysis software for Gibbs sampling output) [13].

In Section 2 further information on the data are given. For later comparison a simplified model with no dependencies is presented in Section 3 and in Section 4 this is followed by a brief discussion of the general aspects of the statistical methods used. In Section 5 and Section 6 we construct more realistic models representing the complex relationships in the data and describe the results of the Bayesian analyses performed. The results obtained and some general comments on the statistical methods are given in Section 7. Details concerning the choice of prior distributions and the calculations using the BUGS software are presented in the Appendices.

2 Data

Data were obtained by questionnaires mailed to sisters and sisters-in-law of IAH patients. A total of 199 women from 113 families returned the questionnaires. In general, women who had had at least one confirmed fetal loss or two livebirths were HLA typed. Furthermore, women with only one livebirth and no miscarriages who had been HLA typed in a previous study were also included. Altogether this gives HLA information on a total of 95 sisters and 50 sisters-in-law. Calculations made in this paper are based only on pregnancies of these 145 women who are HLA typed. The pregnancy history and the age history (see below) of the women have been thoroughly validated and confirmed by medical records; for further details see Christiansen *et al.* [1]. Basic information on the data is given in Table 1.

For each woman we shall here consider the following variables

- type of family relationship, i.e. sister or sister-in-law.
- pregnancy history, a list giving the confirmed outcomes of all the pregnancies of the woman; fetal loss or livebirth.
- age history, a list similar to the pregnancy history, but giving the age of the woman at each pregnancy.

- the first HLA-DR allele of the woman.
- the second HLA-DR allele of the woman.

Earlier research [14] has established that the risk of fetal loss is related to maternal age; with a low risk for young women and a much higher risk for older women. We classify women of age less than 35 as young and women of age 35 or more as old.

We wish to investigate whether the suggested HLA allogenotypes DR1/Br, DR3 or DR10 are genetic markers for the predisposition of fetal loss. Hence we classify the women as having the risk allogenotypes if either one or both of the woman's HLA-DR alleles are one of these allogenotypes. If the woman does not have any of these HLA allogenotypes then we classify her with no-risk allogenotypes.

The final data used in the analyses concerning maternal age and HLA allogenotypes are therefore reduced to

- maternal age at each pregnancy as either below or above(and equal to) 35 years.
- the woman's HLA-DR alleles are risk or no-risk allogenotypes.

Table 1: Basic information on the pregnancy outcomes for all the responders and for responders who were HLA typed.

	number of women	total number of confirmed pregnancies	total number of confirmed fetal losses	frequency of fetal loss in percent
returned questionnaries	199	431	88	20.4
HLA typed women	145	361	81	22.4

3 A preliminary analysis based on a simple contingency table

Ignoring dependencies between pregnancies the data can be summarized by a three-way contingency table according to maternal age, HLA allogenotypes and pregnancy outcome as shown in Table 2.

Table 2: The pregnancies cross-classified according to maternal age, HLA allogenotypes and outcome of the pregnancy.

	< 35		≥ 35	
	livebirth	fetal loss	livebirth	fetal loss
no-risk allogenotypes	144	18	7	1
risk allogenotypes	121	57	8	5

The odds of fetal loss for women with risk allogenotypes relative to the odds of fetal loss for women with no-risk allogenotypes are easily obtained from this table. In the group of young women the odds ratio is 3.77 and for old women the odds ratio is 4.38.

The hypothesis of equal odds ratios in the two age-groups, i.e. the same influence of the HLA allogenotypes in the two groups, corresponds to a log-linear model with no three-factor interaction. The likelihood ratio test of this hypothesis relative to the unrestricted model gives a p-value of 90% indicating that a common value of the odds ratio is compatible with data. The maximum likelihood estimate of the common odds ratio is 3.80 with a 95% confidence interval ranging from 2.20 to 6.84.

This analysis suggests that the risk of fetal loss for women who have the risk allogenotypes is larger than for women not having any of the risk allogenotypes. The results are, however, based on a questionable assumption of independence among all pregnancy outcomes and may therefore not be reliable. We expect some dependencies between pregnancy outcomes for the same woman. Moreover, dependencies may also be expected between pregnancy outcomes for different women in the same family, whereas pregnancy outcomes for women in different families are expected to be independent. Sophisticated models that include such dependencies may be defined by introducing appropriate conditional independence assumptions.

4 General aspects of the statistical modelling

Models defined in terms of conditional independencies, so-called *graphical models*, are conveniently represented by a graph, where the vertices in the graph represent the components of the statistical model and missing links represent the conditional independence assumptions. Some advantages of graphical models are that complex models can be described and handled using very simple elements and the obtained graph will visually represent the contents of the model.

In the following we give some basic facts to the field of Bayesian graphical modelling and how Markov chain Monte Carlo (MCMC) methods may be used to perform computations in these models.

For a general discussion on graphical models see e.g. Spiegelhalter [5], Lauritzen [6] or Edwards [7]. Concerning the use of MCMC methods in complex models we refer to Gilks *et al.* [15] and Gelman and Rubin [16].

4.1 Description of directed graphical models

Let the structural relations between the random variables be represented by a directed acyclic graph (DAG), where the vertices represent random variables and a direct influence from one variable to another is represented by a directed edge. For the purpose of interpretation, the arrows in the DAG indicate that some variables are responses to others in a recursive response structure.

Let V be a set of random variables. A directed graphical model for V is specified by assuming that the distribution of V belongs to the set of distributions which obey the directed Markov property with respect to (w.r.t.) the DAG. There are several equivalent ways of stating the directed Markov property w.r.t. a DAG. Assuming that V has density w.r.t. a product measure one of the formulations is the existence of a *recursive factorization* leading to the joint density

$$p(V) = \prod_{v \in V} p(v \mid \mathrm{pa}(v)), \tag{1}$$

where $\mathrm{pa}(v)$ denotes the parents of the vertex v. Later, we will use $\mathrm{nd}(v)$ to denote the non-descendants of v. Note that terms like parents, children and non-descendants are purely graphtheoretic entities but they are used in the same meaning as in a genealogical tree.

In words, the recursive factorization says that the joint probability of all random variables is formed by a product of the conditional probabilities for each random variable given its parents.

The directed Markov property w.r.t. the DAG imply that we have a representation of the conditional independence relations assumed for V. One of the rules for reading conditional independence relations off the DAG is the *directed local Markov property.* It says that each random variable v is conditionally independent of its non-descendants, given its parents, i.e.

$$v \perp\!\!\!\perp \mathrm{nd}\,(v) \mid \mathrm{pa}\,(v)\,,$$

where $\perp\!\!\!\perp$ is read as "is conditional independent of" and $\mid$ as "given".

A small example may illustrate the basic ideas. Let A, B, C, D be four random variables, i.e. $V = \{A, B, C, D\}$, and let the structural relations between these random variables be given by the DAG in Figure 1. The directed graph in Figure 1 is acyclic since we are not able to start in one vertex and return to this vertex just following the direction of the arrows.

By the recursive factorization we obtain, that

$$p(V) = p(A, B, C, D) = p(A)\, p(B)\, p(C \mid A, B)\, p(D \mid B, C)$$

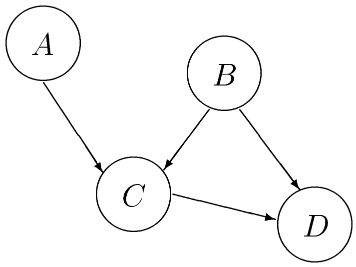

Figure 1: Simple DAG representing the structural relations between the four random variables A, B, C, D.

and the directed local Markov property says, for example, that D is conditionally independent of A given B and C, i.e. $D \perp\!\!\!\perp A \mid \{B, C\}$.

4.2 Bayesian inference using Gibbs sampling

The present work deals with Bayesian inference in complex graphical models. Our general approach is to represent the qualitative relationships between all quantities as a DAG, and then provide conditional distributions for each vertex given its parents. In other words, we specify the joint distribution of all the variables. To distinguish between the observed and unobserved variables we divide V into X, the observed variables (data), and θ, the unobserved variables (missing/latent observations and parameters). The posterior distribution of the unobserved quantities given the observed quantities is

$$p(\theta \mid X) \propto p(X \mid \theta)\, p(\theta),$$

where $p(X \mid \theta)$ is the likelihood and $p(\theta)$ is the prior distribution on θ.

The problem is that we may only be interested in the marginal posterior distribution of θ_i, say,

$$p(\theta_i \mid X) = \int p(\theta \mid X)\; d\theta_{\setminus i},$$

where $\theta_{\setminus i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p)$. Often we have to integrate over high-dimensional probability distributions to make inference about the model parameters. In general, we may not be able to calculate such complex integrals analytically. However, MCMC techniques provide powerful numerical methods for making approximations of such integrals.

BUGS [10, 11, 12] is a general purpose software intended for carrying out Bayesian inference on statistical problems for which there is no exact analytical solution and for which standard approximation techniques are not feasible. In BUGS the necessary numerical integrations are made using the simulation technique Gibbs sampling [8, 9]. BUGS (version 0.6) uses a systematic scheme Gibbs sampler for drawing dependent samples from the posterior distribution. That is, BUGS successively simulates values from the *full conditional distributions* $p\left(v \mid V \setminus v\right)$, i.e. the conditional distribution of each vertex given all the other vertices in the graph. The full conditionals can be identified from the local parent-child distributions, as

$$p\left(v \mid V \setminus v\right) \propto p\left(v \mid \mathrm{pa}\left(v\right)\right) \prod_{w:\ v \in \mathrm{pa}(w)} p\left(w \mid \mathrm{pa}\left(w\right)\right) \ .$$

In words, the full conditional distribution of the vertex v is proportional to the product of the conditional distribution of v given its parents and the conditional distributions where v is a parent vertex of another vertex w, say. Thus, the parent-child distributions are all we need to simulate values from the posterior distribution of all unobserved quantities.

One may also note from (1) that the joint distribution of all the quantities is totally specifed by the parent-child distributions. So by identifying all parent-child distributions we can both specify the model and also simulate values here from. From BUGS a stream of simulated values for each quantity of interest is obtained and these values are used to make inference about the parameters in question.

Together with BUGS a set of S-PLUS functions, called CODA [13], is provided. CODA serves as an output processor for BUGS and it may be used to calculate diagnostics for convergence of the Gibbs sampler and statistical and graphical output analyses of the simulated values obtained from BUGS. Implemented in CODA are convergence diagnostics suggested by Gelman and

Rubin [17], Geweke [18], Raftery and Lewis [19] and Heidelberger and Welch [20]. Autocorrelations for each variable in the chain and cross-correlations between variables are also available

Details of computational issues using BUGS are given in Appendix B.

5 Dependence between pregnancy outcomes for the same woman

In this section we consider a model where possible within-person dependencies are taken into account.

5.1 Model including genetic markers

The outcome of a pregnancy may depend on characteristics of the pregnant woman or more generally on the couple in question. In the following we build and explain the model step by step. The index ij refers to the jth pregnancy for couple i.

So let G_{ij} denote the outcome of the jth pregnancy for couple i, i.e.

$$G_{ij} = \begin{cases} 0 & \text{if livebirth} \\ 1 & \text{if fetal loss} \end{cases} .$$

Introduce p_{ij}, where $0 < p_{ij} < 1$, as the probability that the outcome of the jth pregnancy for couple i is a fetal loss, i.e. G_{ij} is Bernoulli distributed with probability p_{ij} denoted by

$$G_{ij} \mid p_{ij} \sim \; Bern\left(p_{ij}\right) \; .$$

The probability p_{ij} is assumed to depend on the HLA allogenotypes of the woman, the maternal age and some additional factors possessed by the couple. Let λ_i denote these additional factors for couple i and assume that λ_i is normally distributed as

$$\lambda_i \mid \mu, \tau \sim \mathcal{N}\left(\mu, \frac{1}{\tau}\right) \; ,$$

where μ denotes the mean and τ the inverse variance.

For each pregnancy the age of the woman in the couple is represented by a covariate z_{ij} defined as

$$z_{ij} = \begin{cases} 0 & \text{if age } < 35 \\ 1 & \text{if age } \geq 35 \; . \end{cases}$$

Similarly, information on the woman's HLA allogenotypes are represented by the covariate h_i, i.e.

$$h_i = \begin{cases} 0 & \text{if no-risk allogenotypes} \\ 1 & \text{if risk allogenotypes} \; . \end{cases}$$

The model specifies that p_{ij} can be described by a logistic regression model as

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \lambda_i + cz_{ij} + dh_i \; , \tag{2}$$

where c and d are regression parameters associated with the covariates z_{ij} and h_i.

Rewriting (2) it may be seen that $\exp(d)$ is the change in odds of fetal loss for couples where the woman has some risk allogenotypes relative to the odds of fetal loss for couples where the woman does not have any risk allogenotypes.

5.2 Informative illustration of the model by a DAG

The conditional independence assumptions made in this model are easily illustrated by a DAG, where, to ease interpretation, different symbols are used for different types of vertices. Hence we illustrate the DAG by a figure as e.g. introduced in Spiegelhalter *et al.* [21]:

covariate vertices represent constants given in the model. A double-edged rectangle is used to represent such a vertex and naturally covariate vertices will never have parents in the graph.

The covariates in this model are z_{ij} and h_i.

random vertices represent variables that are given a distribution. A random vertex may be *observed*, whereby it corresponds to the data, or it may be *unobserved*, whereby it represents either missing/latent data or unknown parameters. A single-edged rectangle is used to represent an observed vertex and a circle is used to represent an unobserved vertex. Random vertices may be both parents and children in the graph.

Here G_{ij} is an observed vertex and $\lambda_i, c, d, \mu, \tau$ are unobserved vertices.

deterministic vertices are logical functions of other vertices. Hence, depending on the specification, these vertices may be either constants, observed or unobserved.

Here p_{ij}, $\exp(d), 1/\tau$ are unobserved deterministic vertices.

Finally, two kinds of edges are used:

solid edges indicating a stochastic dependence of a child given its parents.

dashed edges indicate that the vertices are related through a logical function. When identifying probabilistic relationships (i.e. parents and non-descendants) deterministic links are collapsed over so that, for example, the parents of G_{ij} are $c, d, \lambda_i, h_i, z_{ij}$.

In some situations it may be helpful to simplify the illustration of the DAG. If a typical structure among some vertices is repeated (e.g. the specification of a pregnancy is the same in each case), we may illustrate this using a large box (also called a plate) around these vertices. The graph should then be interpreted as follows; the vertices within the plate are each repeated, say, n times and each group of vertices is connected following the same structure. Further, the vertices outside the plate are connected to the n repetitions as shown for one of these.

The model specified in Section 5.1 is illustrated by the DAG in Figure 2.

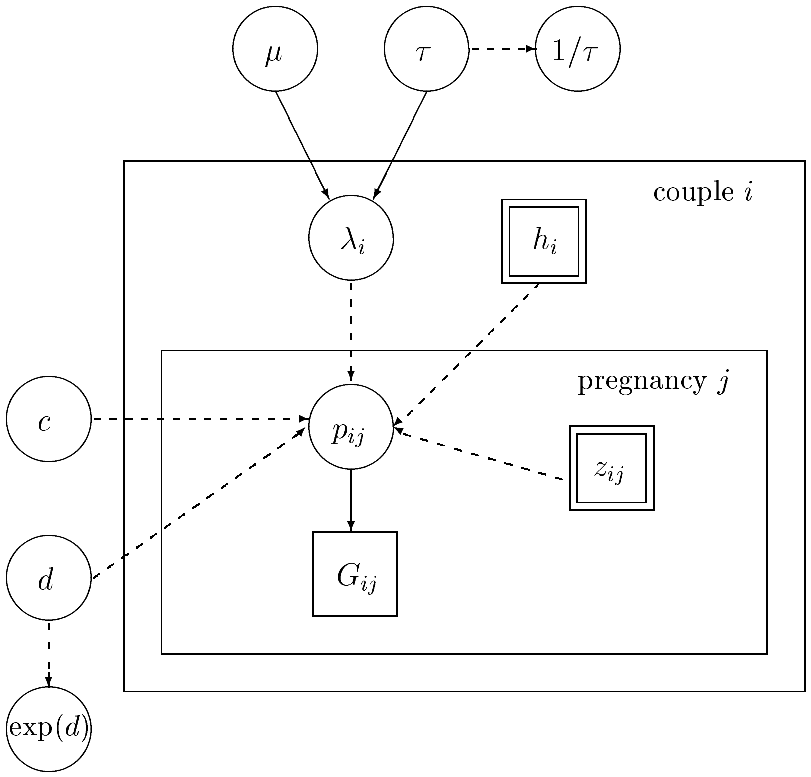

Figure 2: Illustration of the model which takes within-person dependencies into account. Here $i = 1, 2, \ldots, 145$ and for given i the index j will be at least 1 and at most 9.

5.3 Choosing prior distributions

A priori the four parameters μ, τ, c, d are assumed to be independent and their respective marginal prior distributions are chosen as follows. We know little about μ and d, motivating that normal distributions with a relatively large variance are chosen as their prior distributions. Also on τ we have little prior information except that it shall be positive number, as it represents a precision. Therefore, a gamma distribution with a large variance is chosen as the prior on τ.

Intuitively, the data do not contain much information about the regression parameter c, as only few pregnancies where the woman is 35 years old or more are included. As this is not the parameter of interest for the present study, prior information concerning this parameter is desirable. From other studies, e.g. Knudsen *et al.* [14], we have some information about the risk of abortion according to maternal age. This leads to a more specific prior distribution on c; see Appendix A for further details.

To summarize we state that

$$\begin{aligned} \mu &\sim \mathcal{N}(0,100) \\ \tau &\sim \mathcal{G}(0.01,0.01) \\ c &\sim \mathcal{N}(1.1,0.16) \\ d &\sim \mathcal{N}(0,100) \ , \end{aligned}$$

where $\mathcal{G}(a,b)$ denotes the gamma distribution with mean a/b and variance a/b^2.

5.4 Computations using BUGS and CODA

To start the Gibbs sampler some initial values of the founder vertices μ, τ, c, d have to be given and to escape the influence of these some beginning iterations of the Gibbs sampler, also called the 'burn-in', are ignored.

The general understanding is that no valid inference can be made if the runs are too short. Further, if the chains are long enough then only a few chains are needed. In this paper the general strategy will be to use few long chains from different starting points. For the analysis of this model we choose to make three separate chains, each starting from different points. Each chain is decided to consist of a 'burn-in' containing 5,000 iterations and a sample (i.e. the iterations of the Gibbs sampler that are used for making inference) of size n = 15,000. In total, three samples are obtained from $p(c, d, \mu, \tau, \lambda \mid \mathbf{h}, \mathbf{z}, \mathbf{G})$.

Convergence of the Gibbs sampler is assessed by CODA. We visually inspect the sampled values by drawing trace plots for all quantities of interest, i.e. $\exp(d), \mu, 1/\tau, c$. No systematic trend is found in the plots. Further, we calculate the convergence diagnostics provided by CODA. We note for all quantities of interest that the Gelman-Rubin shrink factor [17] has stabilized at 1. Moreover, the Geweke [18] Z-scores show no extreme values, as for all three chains the values between -1.66 and 1.17. Similarly, the Raftery-Lewis dependence factor [19] and the Heidelberger-Welch stationarity test [20] do not cause any suspicion concerning non convergence. Thus we believe that the Gibbs sampler has converged and combine the three samples to one large sample containing $N = 45{,}000$ values.

We are primarily interested in inference about $\exp(d)$ but also $\mu, 1/\tau, c$ are of some interest. From the entire sample the posterior distribution of any quantity of interest can be estimated. For example, considering the simulated values of d we may estimate a 95% posterior (probability) interval of $\exp(d)$ by the 2.5th and 97.5th percentile in the empirical distribution of $\exp(d)$. Also estimates of posterior mean and posterior standard deviation are easily obtained.

We find that $\exp(d)$ has a posterior mean at 4.95 with a 95% posterior interval ranging from 2.03 to 10.87. Results about all the parameters are found in Table 3, where we note, for example, that the 95% posterior interval for $\exp(d)$ is very wide. Moreover, the posterior distribution of c is very similar to the prior distribution on c. This indicates that data may not contain a lot of information on c.

To see whether the wide posterior intervals mainly are due to the large variation allowed in this model we add an additional informative prior. As we do not want to impose too much information on the parameter of main interest d, a more informative prior distribution on μ can be used. This prior is based on results on the risk of fetal loss for a general population known from Knudsen *et al.* [14]; for further details see Appendix A. As an informative prior on μ we choose

$$\mu \sim \mathcal{N}\left(-2.48, 0.25\right) \; .$$

Using this prior on μ new simulations are performed using the same strategy as before. We visually inspect the sampled values by drawing trace plots for all variables and calculate convergence diagnostics as provided by CODA.

Again convergence seems close and we accept that the Gibbs sampler has converged. Hence we combine the three samples to one large sample.

Using this informative prior on μ, the 95% posterior interval for $\exp(d)$ becomes 2.03 to 8.39 and the posterior mean is 4.35. Thus if one or both of the pregnant woman's HLA alleles are the risk allogenotypes then the odds of fetal loss is between 2.03 and 8.39 times larger than the odds of fetal loss when the woman does not have any of these risk allogenotypes, approximately it is 4.35 times larger. In other words, the maternal HLA allogenotypes DR1/Br, DR3 and DR10 are genetic markers for the risk of fetal loss, thereby we believe that tendency to fetal loss is heritable.

Results about all parameters are found in Table 3. Only minor differences are found between the results from the two analyses.

Table 3: The table displays results from the two analyses. Inference is in both cases based on 3 parallel runs each contaning 15,000 iterations.

	Non-informative prior on μ				
Parameter	mean	s.d.	2.5%	median	97.5%
c	0.83	0.34	0.16	0.82	1.48
$\exp(d)$	4.95	2.33	2.03	4.44	10.87
μ	−2.76	0.42	−3.66	−2.73	−2.02
$1/\tau$	1.67	0.84	0.48	1.51	3.74
	Informative prior on μ				
Parameter	mean	s.d.	2.5%	median	97.5%
c	0.82	0.34	0.16	0.82	1.48
$\exp(d)$	4.35	1.66	2.03	4.06	8.39
μ	−2.63	0.30	−3.25	−2.62	−2.08
$1/\tau$	1.51	0.70	0.49	1.40	3.16

5.5 Introducing interaction and additional covariates

In the model introduced in Section 5.1 we have assumed that the influence of the age and the influence of the HLA allogenotypes on the log-odds are additive. This assumption can be tested by introducing the interaction term $z_{ij} \cdot h_i$. The logistic regression model for p_{ij} then becomes

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \lambda_i + cz_{ij} + dh_i + kz_{ij}h_i \ ,$$

where k is the regression parameter associated with the interaction. The prior on k is chosen as $\mathcal{N}(0, 100)$, and the remaining priors are as in the latter analysis above. The 95% posterior interval concerning k becomes -2.44 to 1.01. Thus there seems not to be any visible influence of an interaction between age and HLA allogenotypes.

The model also ignores whether the pregnant woman is a sister or a sister-in-law of an IAH patient. This information can be included by introducing the covariate s_i, i.e.

$$s_i = \begin{cases} 0 & \text{if sister-in-law} \\ 1 & \text{if sister} \ . \end{cases}$$

With l as the regression parameter associated with s_i the logistic regression model becomes

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \lambda_i + cz_{ij} + dh_i + ls_i \ .$$

The prior on l is chosen as $\mathcal{N}(0, 100)$, and the remaining priors are as in the latter analysis above. The 95% posterior interval concerning l is -0.78 to 0.69. Hence the covariate s_i seems not to have any visible influence once the HLA allogenotypes of the woman are included.

Additional analyses (results not shown) did not lead to any changes when omitting still-births (i.e. fetal losses at gestational age 28 weeks or later) or including the number of inexplicable fetal losses of the IAH patient as a predictor.

6 Dependence between pregnancy outcomes for related women

We extend the model such that it takes both within-person and within-family dependencies into account.

6.1 Extended model including genetic markers

We now distinguish between two groups of families, an affected group and a general group, and assume that each family has its own frailty of fetal loss. This frailty represents the additional factors (i.e. in addition to the HLA allogenotypes of the pregnant woman and her age) which may cause a fetal loss.

For an affected family f the frailty is denoted by α_f and is assumed to be normally distributed as

$$\alpha_f \mid \mu^a, \tau^a \;\sim\; \mathcal{N}\left(\mu^a, \frac{1}{\tau^a}\right) \;.$$

Influenced by α_f each individual within the affected family has its own frailty of fetal loss. For individual i from family f the individual frailty is denoted by λ_{fi} and is assumed to be normally distributed as

$$\lambda_{fi} \mid \alpha_f, \xi \;\sim\; \mathcal{N}\left(\alpha_f, \frac{1}{\xi}\right) \;. \tag{3}$$

The individuals from an affected family can either be a brother or a sister of an IAH patient. Information on the type of family relationship is represented by the covariate s_{fi} defined as

$$s_{fi} = \begin{cases} 0 & \text{if brother} \\ 1 & \text{if sister .} \end{cases}$$

Each individual has a partner who is assumed to be picked at random from a general family. We do not have any information about possible changing partnerships, but assume that the individuals have had the same partner throughout all the pregnancies. The frailty of fetal loss for the partner-family of individual i from family f is denoted by β_{fi} and is assumed to be normally distributed as

$$\beta_{fi} \mid \mu^n, \tau^n \sim \mathcal{N}\left(\mu^n, \frac{1}{\tau^n}\right) \ .$$

This frailty of the general family influences the frailty of fetal loss possessed by the partner. The latter is denoted by γ_{fi} and is assumed to be normally distributed as

$$\gamma_{fi} \mid \beta_{fi}, \xi \sim \mathcal{N}\left(\beta_{fi}, \frac{1}{\xi}\right) \ . \tag{4}$$

Observe from (3) and (4) that the frailty possessed by an individual is influenced by the frailty of the family in the same way whether the family is affected or general.

Let couple i be the couple where individual i is the one part, such that we refer to the jth pregnancy for couple i from family f by index fij. Thereby, G_{fij}, p_{fij}, z_{fij}, h_{fi} replace G_{ij}, p_{ij}, z_{ij}, h_i from the previous analysis and c and d have the same interpretation.

The probability p_{fij} is assumed to depend on the frailty of fetal loss possessed by the pregnancy, the age of the pregnant woman and her HLA allogenotypes. The frailty possessed by the pregnancy is composed by a contribution from the foetus and a contribution from the pregnant woman. The frailty component attributed to the foetus is assumed to be the average of the frailties of its parents. Thus the paternal influence is represented through this term. It seems biologically implausible that there should be a paternal effect beyond this; although in principle one could contemplate the foetal effect to be determined from the paternal frailty alone.

If we let b represent the degree of influence of the foetus the term $b\,\frac{\lambda_{fi}+\gamma_{fi}}{2}$ emerges. Then $1-b$ represents the degree of influence of the woman's frailty and the woman's contribution to the combined frailty is expressed by $(1-b)\,(s_{fi}\lambda_{fi}+(1-s_{fi})\gamma_{fi})$.

The model specifies that p_{fij} is described by a logistic regression model as

$$\log\left(\frac{p_{fij}}{1-p_{fij}}\right) = b\,\frac{\lambda_{fi}+\gamma_{fi}}{2} + (1-b)\,(s_{fi}\lambda_{fi}+(1-s_{fi})\gamma_{fi}) \\ + cz_{fij} + dh_{fi}\;.$$

The DAG in Figure 3 illustrates the model, where the quantities of main interest are b, c, $\mu^a-\mu^n$, $\exp(d)$.

6.2 Bayesian analysis by BUGS and CODA

The prior distributions of the founder vertices are chosen by the same principals as in the previous analysis. We want especially to make inference about μ^a and d, whereby we specify their prior distributions with relatively large variance. Based on results stated in Knudsen *et al.* [14] (see Appendix A for further details) informative prior distributions are chosen for μ^n and c. We know that b is between 0 and 1 and assume that all values of b within these bounds are equally likely. Finally, priors on τ^a, τ^n and ξ are chosen

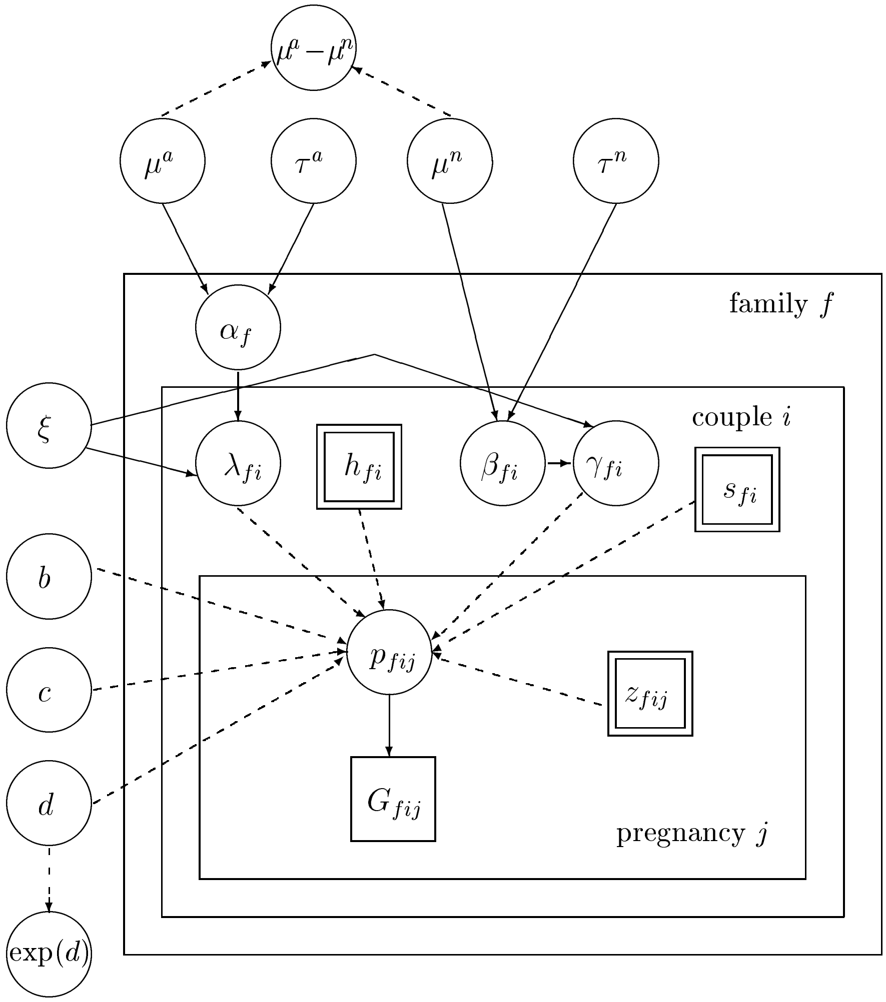

Figure 3: Illustration of the model which takes within-person and within-family dependencies into account. Here $f = 1, 2, \ldots, 85$ and for each f the index i is at least 1 and at most 8. Moreover, for given f and i the index j is at minimum 1 and at maximum 9.

non-informatively. More precisely, we let

$$\begin{aligned}
\mu^a &\sim \mathcal{N}(0, 100) \\
d &\sim \mathcal{N}(0, 100) \\
\mu^n &\sim \mathcal{N}(-2.6, 1.0) \\
c &\sim \mathcal{N}(1.1, 0.16) \\
b &\sim \mathcal{U}(0, 1) \\
\tau^a &\sim \mathcal{G}(0.01, 0.01) \\
\tau^n &\sim \mathcal{G}(0.01, 0.01) \\
\xi &\sim \mathcal{G}(0.01, 0.01) \ ,
\end{aligned}$$

where $\mathcal{U}$ denotes the uniform distribution.

Calculations are performed using the strategies outlined previously, i.e. we simulate three chains from different starting points. Each chain consists of a 'burn-in' containing 5,000 iterations followed by a sample consisting of 50,000 iterations. The large number of iterations is chosen, since this model is rather complex and it includes many unobserved vertices. Actually, ignoring the deterministic vertices, the model contains totally 1540 vertices of which 1012 are observed and 528 are unobserved. As expected (with many unobserved quantities and only relatively few observed ones) convergence difficulties appeared when we tried to handle simulation from this model. For example, most of the Geweke Z-scores fall in the extreme tails of a standard normal distribution and the Heidelberger-Welch test report many failures. The Gelman-Rubin shrink factors are found to be between 1.00 and 1.16

Some of the troubles seem to appear allowing b (the degree of influence from the foetus) to be an unknown parameter and not a constant value. Therefore, we choose to run the Gibbs sampler with different constant values of b. Consider for instance the model with $b = 0$. This model disregards an influence from the foetus but it distinguishes between the two groups of families; affected and general families. Compared with the model analyzed in Section 5, we do now distinguish between sisters and sisters-in-law of IHA patients and assume that the frailty of fetal loss differs for the two types of women (i.e. from affected or general families).

Convergence is assessed by CODA; the convergence diagnostics and trace plots basically point towards convergence. Although the results are not quite satisfying, the improvement compared to the simulations with varying b seems obvious. We note, for example, that the Gelman-Rubin shrink factors are stabilized at 1.00 concerning all quantities of interest. Moreover, few of the Geweke Z-scores fall in the extreme tails of a standard normal distribution.

From the summary statistics based on the combined samples we find that the values of $\mu^a - \mu^n$ is about 0 (posterior mean $= -0.02$ with a 95% posterior interval ranging from -0.96 to 0.83), suggesting that the additional factors from the family pattern apparently are identically distributed in the two groups.

From the summary statistics for $\exp(d)$ we find that the posterior mean is 4.65 with a 95% posterior interval ranging from 1.97 to 9.71. Further results are found in Table 4. Comparing these summary statistics with the results found in the analysis made in Section 5.4 we see a good agreement between the results. The main difference in the obtained results is that the results in the latest analysis are somewhat weaker, as e.g. the 95% posterior intervals are wider. This also matches our intuitive understanding, i.e. relatively less information is available as we introduce more dependencies into the model.

Consider instead the model with $b = 0.75$. That is, assume that frailty possessed by the pregnancy can be decomposed into 3/8 of the frailty possessed by each parent and the last 2/8 of the contribution is from the frailty of the pregnant woman. Thus if we consider a sister (from an affected family) and her husband (from a general family) the frailty possessed by the pregnancy is

$$\left(\frac{3}{8} + \frac{2}{8}\right) \cdot \lambda_{fi} + \frac{3}{8} \cdot \gamma_{fi} \ .$$

Again we run three different chains using the Gibbs sampler and check for convergence by CODA. We get some indications that convergence is close, although the results are less convincing than with $b = 0$. However, summary statistics based on the combined samples are pretty close to the summary statistics obtained with $b = 0$. Therefore, similar conclusions are drawn from this analysis. Details of the results are given in Table 4.

Table 4: The table displays results from the analyses of the model including both within-person and within-family dependencies. Moreover, b is considered as a constant. In both cases inference is based on 3 parallel runs each contaning 50,000 iterations.

			Model with $b = 0$		
Parameter	mean	s.d.	2.5%	median	97.5%
$\exp(d)$	4.65	2.02	1.97	4.24	9.71
$\mu^a - \mu^n$	−0.02	0.45	−0.96	−0.01	0.83
c	0.85	0.34	0.17	0.85	1.51
			Model with $b = 0.75$		
Parameter	mean	s.d.	2.5%	median	97.5%
$\exp(d)$	4.75	2.26	1.94	4.26	10.50
$\mu^a - \mu^n$	−0.06	1.20	−2.47	−0.05	2.36
c	0.84	0.34	0.17	0.84	1.50

7 Discussion

In this paper we have presented several models by which it was investigated whether the HLA allogenotypes DR1/Br, DR3 and DR10 are genetic markers for the risk of experiencing unexplained fetal losses. Our approach emphasizes the structural aspects of the family-patterns by specifying the problem using graphical models. Further, we have outlined how estimation of the parameters of interest can be carried out in a Bayesian framework using Gibbs sampling and it was shown how programs as BUGS and CODA can be used. Below the results obtained concerning heritability of tendency to fetal loss are discussed. Later comments are given on the statistical methods involving the combination of Bayesian graphical modelling and MCMC methods.

7.1 Heritability of tendency to fetal loss

The distribution of frailties due to additional factors in the family pattern are apparently equal in the two groups of families (i.e. affected and general). We also observe that the posterior distribution of μ in the model including only within-person dependencies and a non-informative prior on μ is very similar to the calculated prior distribution of μ^n in the model including both within-person and within-family dependencies. This means that, in our data, the affected couples are not different from the general population concerning the general frailty of fetal loss.

Yet, we see an increased tendency to fetal loss among couples where the woman has risk allogenotypes. Therefore the increased risk of fetal loss is explained by the presence of the pregnant woman's risk allogenotypes only and no additional information is found in family relations. Thereby within-family dependence seems unnecessary and the model including only within-person dependencies is sufficient to describe the problem.

We have assumed that the influence of age and HLA allogenotypes on the log-odds are additive. This assumption was tested and accepted for the model including only within-person dependencies. Also a distinction between sisters and sisters-in-law seemed unnecessary.

Therefore, we believe that the model including only within-person dependencies fit the data well. Based on this model the estimated odds ratio for experiencing a fetal loss among women with risk allogenotypes compared to the women with no-risk allogenotypes is 4.35, with a 95% posterior interval ranging from 2.03 to 8.39. Thus, we have seen that a tendency to fetal loss is heritable, since the maternal HLA allogenotypes DR1/Br, DR3 and DR10 are genetic markers for the risk of fetal loss.

It is also important to notice that we have used two different priors on μ; a non-informative and a more informative one based on results of the risk of fetal loss known from earlier research. Using the non-informative prior on μ lead to an increased uncertainty in determination of the odds ratio. In a

Bayesian analysis the posterior distribution depends on the likelihood as well as the prior distribution, but the influence of the prior becomes less dominating the more information the data contain. Nevertheless the conclusions concerning heritability of fetal loss seems robust against varying models and prior distributions.

The observed Fisher information matrix for μ, τ, c, d (i.e. minus the inverse Hessian of the log-likelihood function evaluated at the maximum likelihood estimates) can be used as a quantitative measure of the amount of information contained in the observed data; Andersen [22] discusses some early results related to these data relative to a given model.

7.2 General comments on the statistical method

MCMC techniques have become an almost miraculous tool in Bayesian analyses and also in likelihood inference the methods are starting to be introduced. The wide spread understanding is, that MCMC techniques can be used to analyze many problems which just a few years ago seemed intractable. The introduction of software packages such as BUGS and CODA has made MCMC analyses within a Bayesian framework easier.

Combining MCMC methods with graphical models enable us to analyze realistic models of, in principle, unlimited complexity and we obtain a flexibility of unprecedented dimensions. Furthermore, graphical models make it possible to represent and communicate the structure (conditional independence assumptions) in complex models in a very simple way.

However, our experience with these methods suggests that some caution is required. For instance, MCMC methods rely on convergence of some simulations, but it may be difficult to check whether convergence is obtained. We recommend that a variety of diagnostic tools should be applied and that few long chains with different starting points are used.

Another good advice is to learn as much as possible about the target density before making any MCMC analysis. For example, one should make some initial analyses of data using simpler models and approximations and hereby gain some initial information about where the distribution is located in the parameter space. Here we made the simple contingency table analysis even though we knew that the assumption of complete independence between the pregnancy outcomes was a very crude assumption. But from this crude analysis we got an idea of the expected size of the odds ratio.

Afterwards steadily increasing the complexity of the model under consideration also makes it more likely to detect model failures. Additionally, one gains a deeper understanding of the data when using this strategy.

Using Bayesian methods another complication emerges. We have to specify a full probability model treating all quantities as random variables. This means that, we shall specify prior distributions on parameters which may be more or less identifiable. Using MCMC techniques and graphical models we may build a very complex and realistic model and also perform calculations in this model. If data are sufficiently informative we will always get sensible conclusions, but if data are sparse compared to the complexity of the model then the parameters may be badly determined, sometimes even not identifiable. Some key questions are therefore: How complex can the models be to have a practical interest but such that you are still able to identify the interesting parameters satisfactorily? and To which extent do the conclusions depend on the chosen prior distributions?

To our knowledge not much attention has been given, in the literature, to a formal discussion of questions of this type. Examination of the amount of information contained in the observed data can be examined using the observed Fisher information matrix. This idea is to some extent pursued in Andersen [22].

Despite the pitfalls of an MCMC analysis, we expect that the use of these methods will expand in the future and MCMC techniques will benefit from continuing popularity.

Acknowledgment

This research was to a large extent performed while the first author was at Aalborg University. It has partially been supported by the Danish Research Councils through the PIFT-programme, the Research Councils of the County of North Jutland, Aalborg Stifts' Julelotteri, Aalborgs Frivillige Bloddonorers Fond and Novo Nordisk A/S. We are grateful to Dr. David J. Spiegelhalter and Professor Michael Væth for inspiring discussions and helpful comments on earlier versions of this work.

A Comments concerning prior distributions

In this appendix we explain our choice of the informative prior distributions. We refer to the model including only within-person dependencies (Section 5) as the simple model (even though it is not that simple at all) and the model including both within-person and within-family dependencies (Section 6) is called the complex model.

A.1 The prior on c

In Section 5.1 p_{ij} was specified by a logistic regression model. Hereby the odds of fetal loss is given by

$$odds = \frac{p_{ij}}{1 - p_{ij}} = \exp\left(\lambda_i + cz_{ij} + dh_i\right) \ .$$

Let $odds_{<}$ and $odds_{\geq}$ denote the odds of fetal loss for couples where the maternal age is less than 35 years and 35 years or more, respectively. Thus

$$c = \log\left(\frac{odds_{\geq}}{odds_{<}}\right) \ .$$

From Table IV in Knudsen *et al.* [14] we state the risk of abortion according to maternal age; see Table 5. Using this risk we find

$$c \approx \log\left(\frac{25}{75} \cdot \frac{90}{10}\right) \approx 1.1 \ .$$

Table 5: Risk of fetal loss according to the maternal age.

Age of mother	No. of pregnancies	Risk of fetal loss (%)
< 35	18178	10.2
≥ 35	1559	24.7

To find some boundaries which at least should be covered, we fix the risk of fetal loss when the maternal age is under 35, i.e. for $z_{ij} = 0$ fix $p_{ij} = 0.1$. Furthermore, for $z_{ij} = 1$ choose the boundaries $p_{ij} = 0.15$ and $p_{ij} = 0.4$. Hereby the boundaries concerning c are

$$c \approx \log\left(\frac{15}{85} \cdot \frac{90}{10}\right) \approx 0.46 \quad \text{and} \quad c \approx \log\left(\frac{40}{60} \cdot \frac{90}{10}\right) \approx 1.79 \ .$$

Choosing the mean-value of c to be 1.1 and a standard variation of 0.4 we obtain that ± 3 times the standard variation will cover the interval not to narrowly. Hence the prior on c is chosen to be

$$c \sim \mathcal{N}(1.1, 0.16) \ .$$

A.2 The prior on μ

Recall that z_{ij} and h_i are covariates in the simple model, whereby they have a constant value. In the Bayesian analysis, we consider the founder vertices μ, τ, c, d as random variables, but in the following calculations they will be fixed at constant values. The model considered in Section 5.1 is specified by the parent-child distributions

$$\lambda_i \mid \mu, \tau \sim \mathcal{N}\left(\mu, \frac{1}{\tau}\right) \quad \text{and} \quad G_{ij} \mid p_{ij} \sim \ Bern\left(p_{ij}\right)$$

and the vertex p_{ij} is specified by the logistic regression model

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \lambda_i + cz_{ij} + dh_i \ . \tag{5}$$

We wish to find the expected value of p_{ij} given μ, τ, c, d, i.e. $\mathrm{E}(p_{ij} \mid \mu, \tau, c, d)$. Therefore rewrite (5) as

$$p_{ij} = \frac{\exp(\omega)}{1 + \exp(\omega)} \ ,$$

where the random variable $\omega = \lambda_i + cz_{ij} + dh_i$ is introduced for reasons of simplicity. The local Markov property leads to

$$\omega \mid \mu, \tau, c, d \sim \mathcal{N}\left(\nu, \sigma^2\right) \ ,$$

where $\mathrm{E}(\omega \mid \mu, \tau, c, d) = \mu + cz_{ij} + dh_i = \nu$ and $\mathrm{Var}(\omega \mid \mu, \tau, c, d) = 1/\tau = \sigma^2$. Thus the expected value of p_{ij} given μ, τ, c, d can be calculated by

$$\mathrm{E}\left(p_{ij} \mid \mu, \tau, c, d\right) = \int_{-\infty}^{\infty} \frac{\exp(\omega)}{1 + \exp(\omega)} \ \varphi\left(\frac{\omega - \nu}{\sigma}\right) d\omega \ ,$$

where $\varphi(\cdot)$ represents the density function of the standard normal distribution. An approximation of this integral may be obtained by calculations similar to those in Spiegelhalter and Lauritzen (page 602) [23]. Accordingly

$$\mathrm{E}\left(p_{ij} \mid \mu, \tau, c, d\right) \approx \Phi\left(\frac{\nu\eta}{\sqrt{1 + \eta^2\sigma^2}}\right) \ ,$$

where η is chosen appropriately and $\Phi(\cdot)$ denotes the distribution function of the standard normal distribution. With $\eta = 0.607$ the approximation is of high quality [23].

Using the approximation in the reverse direction we obtain

$$\mathrm{E}(p_{ij} \mid \mu, \tau, c, d) \approx \frac{\exp\left(\frac{\mu + cz_{ij} + dh_i}{\sqrt{1 + \frac{\eta^2}{\tau}}}\right)}{1 + \exp\left(\frac{\mu + cz_{ij} + dh_i}{\sqrt{1 + \frac{\eta^2}{\tau}}}\right)} \ . \tag{6}$$

Recall that c, d, τ are fixed and z_{ij}, h_i are constant values. Hence for specific situations a value of μ can be found by means of the expected value of

p_{ij}. Fix the values of c, d, τ according to the mean of their prior distributions as stated in Section 5.3, i.e.

$$c = 1.1 \ , \ d = 0 \ , \ \tau = 1 \ .$$

Inserting the chosen values in (6) we obtain

$$\mu \approx 1.1698 \cdot \log\left(\frac{\mathrm{E}\left(p_{ij} \mid \mu, \tau, c, d\right)}{1 - \mathrm{E}\left(p_{ij} \mid \mu, \tau, c, d\right)}\right) - 1.1 \cdot z_{ij} \ .$$

From Table 5 we know, that the overall risk of having a fetal loss is about 10% when the pregnant woman is less than 35 years old. When the maternal age is 35 years or more the risk is about 25%. Therefore choose $\mathrm{E}\left(p_{ij} \mid \mu, \tau, c, d\right) = 0.1$ when $z_{ij} = 0$ and obtain

$$\mu \approx 1.1698 \cdot \log\left(\frac{10}{90}\right) = -2.570 \ .$$

Similarly choose $\mathrm{E}\left(p_{ij} \mid \mu, \tau, c, d\right) = 0.25$ when $z_{ij} = 1$, whereby

$$\mu \approx 1.1698 \cdot \log\left(\frac{25}{75}\right) - 1.1 = -2.385 \ .$$

On basis of the above considerations we specify the prior distribution on μ as

$$\mu \sim \mathcal{N}\left(-2.48, 0.25\right) \ .$$

Note that even though the variance is still quite small we have chosen not to be too strict in the choice of prior.

A.3 The prior on μ^n

From other studies, e.g. Knudsen *et al.* [14], we have some information about the risk of fetal loss in the general population. Hence to calculate a prior on μ^n we modify the model illustrated in Figure 3 such that all individuals belong to general families. That is, the vertices μ^a, τ^a are removed and μ^n, τ^n become parents of both α_f and β_{fi}; the rest of the DAG is unchanged.

Using this model and arguments similar to those in Section A.2 we may deduce that

$$\mathrm{E}\left(p_{fij} \mid \mu^n, \tau^n, \xi, b, c, d\right) \approx \frac{\exp\left(\frac{\nu}{\sqrt{1+\eta^2\sigma^2}}\right)}{1+\exp\left(\frac{\nu}{\sqrt{1+\eta^2\sigma^2}}\right)} \ ,$$

where

$$\begin{aligned} \nu &= \mu^n + cz_{fij} + dh_{fi} \\ \sigma^2 &= \frac{1}{2}\left(\frac{1}{\xi}+\frac{1}{\tau^n}\right)\left((1-b)^2+1\right) . \end{aligned}$$

Recall that z_{fij}, h_{fi} are constants and we fix the values of τ^n, ξ, b, c, d according to the mean of their prior distributions as stated in Section 6.2, i.e.

$$\tau^n = \xi = 1, b = 0.5, c = 1.1, d = 0 \ .$$

Hereby we see that

$$\nu = \mu^n + 1.1 \ z_{fij} \qquad \text{and} \qquad \sigma^2 = 1.25 \, .$$

Using the overall risks of having a fetal loss stated in Knudsen *et. al.* [14] we choose as in Section A.2

$$\begin{aligned} \mathrm{E}\left(p_{fij} \mid \mu^n, \tau^n, \xi, b, c, d\right) &= 0.1 \ \text{ when } \ z_{fij} = 0 \\ \mathrm{E}\left(p_{fij} \mid \mu^n, \tau^n, \xi, b, c, d\right) &= 0.25 \ \text{ when } \ z_{fij} = 1 \ . \end{aligned}$$

and with $\eta = 0.607$ we get the limits

$$\begin{aligned} \mu^n &\approx 1.2085 \cdot \log\left(\frac{10}{90}\right) \approx -2.655 \\ \mu^n &\approx 1.2085 \cdot \log\left(\frac{25}{75}\right) - 1.1 \approx -2.428. \end{aligned}$$

For the models where b is chosen as a constant, the calculations of the limits for μ^n are exactly the same. However, the calculated limits will depend somewhat on the value chosen for b. The results are listed in Table 6.

Table 6: Boundaries on μ^n depending on the value chosen for b.

Value of b		Interval limits for μ^n	
0		-2.896 ;	-2.548
0.5	$(b \sim \mathcal{U}(0,1))$	-2.655 ;	-2.428
0.75		-1.375 ;	-1.787

On basis of the above considerations we specify the prior distribution on μ^n in all three models to be

$$\mu^n \sim \mathcal{N}\left(-2.6, 1.0\right) \ .$$

B Calculations using BUGS

BUGS is a program which is very easily used. It provides a declarative language for specifying statistical models by stating parent-child distributions and prior distributions on the founder vertices.

It is very helpful to represent the model as a DAG as described in Section 5.2. Supported by such a picture it is straightforward to produce the BUGS code. Each vertex in the graph shall appear once and only once on the left side of the 'equations' specifying the relationship with its parents. A stochastic relationship (solid edges) is represented by $\sim$ and a deterministic relationship (dashed edges) by $<-$. Plates are coded as for-loops. A wide range of distributions and functions are provided.

Having specified the model together with the covariates and the observed data, we compile the program and from this point on the program takes care of identifying relevant terms and choosing optimal sampling distributions to simulate values of the unknown quantities. Performing a run in BUGS means that we execute an explicitly specified number of iterations of the Gibbs sampler. As already mentioned this results in a stream of simulated values (also called a sample) from the approximate posterior distribution of the unobserved quantities.

The latest (Windows) version of BUGS (WinBUGS version 1.2) allows models to be specified using a graphical representation of the model (also called Doodles). Moreover, WinBUGS implement slice sampling and stochastic relaxation algorithms as alternatives to Gibbs sampling. These can lead to improved convergence in certain models.

To learn more about BUGS/WinBUGS we refer to, for example, Gilks *et al.* [10], Best *et al.* [11], Spiegelhalter [5] and Spiegelhalter *et al.* [24, 25]. Software and user manual are available from

`http://www.mrc-bsu.cam.ac.uk/bugs`

.

B.1 BUGS code for the model including within-person dependencies

The code below illustrates the model specification for the model including within-person dependencies and with an informative prior on μ. Note the normal distribution is parameterized using the precision ($\tau = 1/\text{variance}$), otherwise the code seems self-explanatory and no further comment are given.

```
model simple;


const
   Nind  = 145, # total number of HLA-typed women
   Npreg = 361; # total number of confirmed pregnancies
var
   mu, tau, c, d, or, variance, lambda[Nind], h[Nind],
   ind[Npreg], p[Npreg], z[Npreg], g[Npreg];
```

```
data in ''simple.dat'';
inits in ''simple.in'';

{
mu ~dnorm(-2.48,4);        # prior distributions
tau ~dgamma(0.01,0.01);
c ~dnorm(1.1,6.25);
d ~dnorm(0.0,0.01);

for (i in 1:Nind) {        # loop over couples
    lambda[i] ~dnorm(mu,tau);
    }
for (j in 1:Npreg) {       # loop over pregnancies
    logit(p[j]) <- lambda[ind[j]] + c*z[j] +d*h[ind[j]];
    g[j] ~dbern(p[j]);
    }
or <- exp(d);
variance <- 1/tau;
}
```

Running this BUGS code on a PC (Pentium Pro, 200 MHz, 64 Mb Ram) for $20,000$ iterations took approximately 10 minutes.

References

[1] O.B. Christiansen, H.H. Andersen, M. Højbjerre, T.A. Kruse, S.L. Lauritzen, and N. Grunnet. Maternal HLA class II allogenotypes are markers for the predisposition to fetal losses in families of women with unexplained recurrent fetal loss. *European Journal of Immunogenetics*, 22:323–334, 1995.

[2] O.B. Christiansen, O. Mathiesen, J.G. Lauritsen, and N. Grunnet. Idiopathic recurrent spontaneous abortion. Evidence for a familial predisposition. *Acta Obstetricia Gynecologica Scandinavica*, 69(7–8):597–601, 1990.

[3] O.B. Christiansen, K.L. Rasmussen, C. Jersild, and N. Grunnet. HLA Class II alleles confer susceptibility to recurrent fetal losses in Danish women. *Tissue Antigens*, 44(4):225–233, 1994.

[4] O.B. Christiansen, O. Mathiesen, M. Husth, J.G. Lauritsen, C. Jersild, and N. Grunnet. Prognostic significance of maternal DR histocompatibility types in Danish women with recurrent miscarriages. *Human Reproduction*, 8(11):1843–1847, 1993.

[5] D.J. Spiegelhalter. Bayesian graphical modelling: a case-study in monitoring health outcomes. *Journal of the Royal Statistical Society, Series A*, 47(1):115–133, 1998.

[6] S.L. Lauritzen. *Graphical Models.* Oxford University Press, 1996.

[7] D. Edwards. *Introduction to Graphical Modelling.* Springer-Verlag, 1995.

[8] G. Casella and E.I. George. Explaining the Gibbs sampler. *The American Statistican*, 46(3):167–174, 1992.

[9] W.R. Gilks, D.C. Clayton, D.J. Spiegelhalter, N.G. Best, A.J. McNeil, L.D. Sharples, and A.J. Kirby. Modelling complexity; Applications of Gibbs sampling in medicine. *Journal of the Royal Statistical Society, Series B*, 55(1):39–52, 1993.

[10] W.R. Gilks, A. Thomas, and D.J. Spiegelhalter. A language and program for complex Bayesian modelling. *The Statistician*, 43(1):169–178, 1994.

[11] N.G. Best, D.J. Spiegelhalter, A. Thomas, and C.E.G. Brayne. Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society, Series A*, 159(2):323–342, 1996.

[12] D.J. Spiegelhalter, A. Thomas, and N.G. Best. Computation on Bayesian graphical models. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and Adrian F.M. Smith, editors, *Bayesian Statistics 5*, pages 407–425, Oxford, UK, 1996. Clarendon Press.

[13] N.G. Best, M.K. Cowles, and K. Vines. CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, manual version 0.30. *Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, UK*, 1995.

[14] U.B. Knudsen, V. Hansen, S. Juul, and N.J. Secher. Prognosis of a new pregnancy following previous spontaneous abortions. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 39:31–36, 1991.

[15] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapmann & Hall, 1996.

[16] A. Gelman and D.B. Rubin. Markov chain Monte Carlo methods in biostatistics. *Statistical Methods in Medical Research*, 5:339–355, 1996.

[17] A. Gelman and D.B. Rubin. A single series from the Gibbs sampler provides a false sense of security. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 4*, pages 625–631, Oxford, UK, 1992. Clarendon Press.

[18] J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 4*, pages 169–193, Oxford, UK, 1992. Clarendon Press.

[19] A.E. Raftery and S.M. Lewis. How many iterations in the Gibbs sampler? In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 4*, pages 763–773, Oxford, UK, 1992. Clarendon Press.

[20] P. Heidelberger and P.D. Welch. Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144, 1983.

[21] D.J. Spiegelhalter, N.G. Best, W.R. Gilks, and H. Inskip. Hepatitis B: a case study in MCMC methods. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 21–43. Chapman & Hall, 1996.

[22] H.H. Andersen. *Monte Carlo Likelihood in Complex Graphical Models.* PhD thesis, Department of Mathematics, Aalborg University, Denmark, 1997.

[23] D.J. Spiegelhalter and S.L. Lauritzen. Sequential updating of conditional probabilities on directed graphical structures. *Networks*, 20:579–605, 1990.

[24] D.J. Spiegelhalter, A. Thomas, N.G. Best, and W. Gilks. BUGS 0.5 Bayesian inference using Gibbs sampling, manual (version ii). *Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge*, 1996.

[25] D.J. Spiegelhalter, A. Thomas, N.G. Best, and W. Gilks. BUGS 0.6 Bayesian inference using Gibbs sampling (addendum to manual). *Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge*, 1997.