

# Genetic analysis of melanoma onset by using estimating equations and Bayesian random effects models

*K-A Do*<sup>1</sup>    *P Kuhnert*<sup>2</sup>    *S-J Lee*<sup>1</sup>  
*J F Aitken*<sup>3</sup>    *A Green*<sup>4</sup>    *N G Martin*<sup>4</sup>

<sup>1</sup> Department of Biostatistics, the University of Texas, M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.

<sup>2</sup> School of Mathematics, Queensland University of Technology, Brisbane, Queensland 4001, Australia.

<sup>3</sup> Queensland Cancer Fund, Spring Hill, Queensland 4004, Australia.

<sup>4</sup> Epidemiology and Population Health Unit, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Queensland 4029, Australia.

## Summary

There are complex relative contributions of genetic and shared environmental factors to an increased risk in melanoma. Data from the Queensland Familial Melanoma Project comprising 15,907 persons from the 1,912 families of 2,118 melanoma cases were analyzed to estimate the additive genetic, common and unique environmental contributions to variation in the age at onset of melanoma. Two complementary approaches for analyzing correlated time-to-onset family data were considered: the generalized estimating equations (GEE) method in which one can estimate relationship-specific dependence simultaneously with regression coefficients that describe the average population response to changing covariates; and a subject-specific Bayesian mixed model in which heterogeneity in regression parameters is explicitly modeled and the different components of variation may be estimated directly. The proportional hazards and Weibull models were utilized, as both produce natural frameworks for estimating relative risks while adjusting for simultaneous effects of other covariates. A simple Markov Chain Monte Carlo method for covariate imputation of missing data was used and the actual implementation of the Bayesian model was based on Gibbs sampling using the free ware package BUGS. In addition, we also used a Bayesian model to investigate the relative contribution of genetic and environmental effects on the expression of naevi and freckles, which are known risk factors for melanoma.

**Key words and phrases:** Multivariate survival analysis; estimating equations; family; genetic modeling; Gibbs sampling; melanoma.

# 1 INTRODUCTION

In genetic research of chronic diseases, family studies have been the major focus where age-at-onset outcomes are frequently gathered. Such data possess complex features: namely censorship due to the long-term development of diseases and correlation due to shared genome and environment. The first step in analyzing such data is to assess familial aggregation by characterizing the underlying major genes. The problem of age-at-onset is important for linkage analysis using disease status data from both affected and unaffected individuals within families. Incorporating age-at-onset data into linkage analysis can potentially increase the statistical power for detecting linkage. There is a need for statistical models featuring correlated lifetimes which can allow for testing the hypothesis of homogeneity, for estimating regression parameters and variance components (additive genetic, common and unique environment) in the presence of a possible correlation.

To account for the age-at-onset requirements, survival analysis concepts are often utilized and developed for family studies (Abel and Bonney, 1990; Siegmund and McKnight, 1998; Siegmund et al., 1998b; Siegmund et al., 1998a; Yashin and Iachine, 1995). Abel and Bonney developed a model accounting for variable age-at-onset where the hazard function is expressed in terms of a major gene effect and residual family dependence using the regressive approach (Bonney, 1986). The so-called *frailty* model was introduced by several authors (Clayton, 1978) and Vaupel *et al* (Vaupel et al., 1979) and has been extended by many others for studying disease aggregation and gene segregation (Clayton and Cuzick, 1985; Hougaard, 1986; Nielsen et al., 1992; Andersen et al., 1993; Gauderman and Thomas, 1994; Li and Thompson, 1997; Li and Wijnsman, 1998; Siegmund and McKnight, 1998).

Another approach to accommodate the correlated age-at-onset outcomes rigorously is to use the estimating equations for assessing familial aggregation of age-at-onset (Hsu and Prentice, 1996; Hsu and Zhao, 1996). It has two desirable features: (i) robustness - no higher-order distributional assumptions are required beyond pairwise ones; and (ii) computational efficiency.

Recently there has been focus on the development of Bayesian methods for the analysis of censored data in family studies, using Markov Chain Monte Carlo (MCMC) concepts. The Gibbs sampler (Geman and Geman, 1984) is the most popular algorithm used in MCMC applications to correlated data. MCMC methods have been used for linkage analysis (Lange and Sobel, 1991; Kong et al., 1992; Heath, 1997), for the estimation of parameters in the mixed model with and without covariates (Guo and Thompson, 1991; Thomas, 1992), for estimation of gene-smoking interaction and covariate imputation (Gauderman et al., 1997), for performing combined linkage and segregation analysis (Guo and Thompson, 1992; Faucett et al., 1993), and for mixed models of large complex pedigrees (Guo and Thompson, 1994). Some recent computer packages that implement Gibbs sampling for analysis of pedigree data include BUGS (Gilks et al., 1994; Spiegelhalter et al., 1996b; Spiegelhalter et al., 1996a), Genetic Analysis Package (GAP, 1996), and MIXD (Thompson, 1994; Olshen and Wijnsman, 1996). Recent work by the first author (Do et al., 2000) based the analysis of menopausal age in twins on a Bayesian representation of a mixed effects model, where genetic and environmental contributions were treated as random effects, while allowing for adjustments for observed covariates. In this paper, we applied the estimating equations and the Bayesian methods for the analysis of melanoma data from extended families. We investigated the relative contribution of genetic and environmental effects on the age at onset of melanoma, as well as on the expression of naevi and freckles, which are known risk factors of melanoma. Naevi are both precursors and markers for melanoma. In patients with a genetic susceptibility to the disease, there is strong evidence that melanoma arises in pre-existing naevi. Freckles are not precursor lesions for melanoma but is a well-recognized risk factor for melanoma, especially in fair-skinned people. A recent article, (Bataille et al., 2000), investigated the relative contribution of genetic and environmental factors in the expression of naevi and freckles in adults through the study of twins (127 monozygotic and 323 dizygotic twin pairs). They used a maximum likelihood method to model the variance-covariance matrices and contingency tables. Our

study will be the first in melanoma research to extend the methodology that can incorporate covariate effects simultaneously with age at onset and binary endpoints in a Bayesian framework and to apply to a much larger data set of 1,912 families.

## 2 MATERIAL AND METHODS

### 2.1 The data on the age at onset of melanoma and potential risk factors

We used data from the Queensland Familial Melanoma Project. Family ascertainment and data collection have been described in detail (Aitken et al., 1996). Assessing standard melanoma risk factors include counts of naevi on the arms and back, demographic and medical details, lifetime residence and family history of melanoma and other cancers. Briefly, we ascertained all 12,016 first incident cases of cutaneous melanoma (invasive and in situ) diagnosed in Queensland residents between 1982 and 1990 and reported to the Queensland Cancer Registry, or found by comparing cancer registrations for 1984 and 1987 with records of pathology laboratories throughout Queensland. It is estimated that registry records are approximately 95% complete for the study period. Doctor's permission was obtained to approach 10,407 cases of whom 7,784 (75%) returned a brief family history questionnaire, stating whether any of their first-degree relatives (parents, siblings, children) had had a diagnosis of melanoma. A total of 2,920 probands was sampled from these respondents, including all who had claimed a positive family history ( $n = 1,529$ ) and an approximate 20% random sample of the remainder ( $n = 1,391$ ). Probands were sent a detailed family history questionnaire, asking for the names and addresses of all first-degree relatives, relatives' vital status, dates of birth, and ages, and whether any of these relatives had had a melanoma diagnosed by a doctor. To avoid bias in determining the mode of inheritance, second and lower degree relatives were enrolled in the study according to a sequential sampling scheme (Cannings and Thompson, 1977). First degree relatives of all relatives with confirmed melanoma were ascertained through the detailed family history questionnaire, described above, which was mailed to all confirmed positive relatives. In total, 15,989 relatives belonging to 1,912 separate families were reported by 2,118 (73%) probands or other positive relatives. A total of 1,044 relatives for whom date of birth was unknown was excluded, leaving 14,945 relatives for analysis. There were 188 families independently ascertained through two or more probands. To avoid ascertainment bias, these families were included in the dataset separately for each proband in the family.

Medical confirmation and dates of diagnosis were sought for the relatives reported by probands or other relatives to have had melanoma. After eliminating 18.7% of subjects who refused access to their medical records, or those with lost records, or those with false positive reports (basal or squamous cell carcinoma, solar keratoses, or benign naevi), medical confirmation of melanoma as the diagnosis was obtained for 48.2% of the original number of relatives. Only the medically verified cases among relatives were classified as true events; all other relatives were treated as unaffected (censored at last dated of contact).

Risk factor questionnaires were subsequently mailed to all living relatives aged between 18 and 75 years ascertained through the sequential sampling procedure. Other relatives provided proxy reports. The combined number of proxy-reports and self-reports was 9,746 relatives for whom standard risk factor information was available.

For the Bayesian analysis, we focussed on families that included at least one parent and at least one child, where each member in the family should have information on age at diagnosis or age at last follow up and with maximum one missing covariate.

The demographic covariates and hypothesized melanoma risk factors included gender, birth year, place of birth, ability to tan (very brown, moderate tan, slight tan, no tan), propensity to burn (never burn and always tan, sometimes burn and usually tan, usually burn and sometimes tan, always burn and never tan), number of sunburns (0,1,2-5,  $\geq 6$ ), skin color (olive/dark, medium, fair/pale), hair color (black, light/dark brown, fair/blonde,

light/dark red), eye color (brown, green/hazel, blue/grey), total freckling in summer (0, 1–100, > 100), number of naevi (none, few, moderate number, very many), and numerous measures of cumulative lifetime exposures to sun and ultraviolet rays.

## 2.2 Preliminary exploratory analysis

As a preliminary analysis, we ignored correlations within families and applied a combination of parametric and non-parametric survival analysis techniques as exploratory tools to identify possible risk factors for melanoma. Once these fixed effects were identified, we considered incorporating these into a subsequent generalized estimating equation model or a Bayesian model with random effects that could account for within-family correlations. The aim was to quantify the genetic and familial associations in the presence of observed covariate effects.

Manipulation of the entire melanoma dataset resulted in a subset of 9669 observations with a range of explanatory variables that described phenotypic characteristics for each individual along with some demographic details such as birth year and gender. The response variable was the time to diagnosis (or age at the last follow-up), with the proportion of censored cases being approximately 76%.

The median age at onset of melanoma was 43. The correlation estimates for different relationship pairs with both affected members were: 0.67, 0.55, and 0.39 for sib-sib, parent-child, and second/lower order pairs respectively.

The first stage of modelling involved fitting univariate proportional hazard models to assess each variable's individual effect on the time of onset of melanoma. The SAS (Allison, 1995) package was used to fit proportional hazards models of the following form

$$h(t, x) = \Psi(x; \beta)h_0(t) \tag{1}$$

where  $\Psi$  represents a log-linear function  $e^{\beta^T x}$  of the explanatory variables  $x$  and corresponding coefficients  $\beta$ , and  $h_0(t)$  represents the baseline hazard at time  $t$ .

Significant variables associated with the age at onset of melanoma consisted of eye, hair and skin colour; freckling; number of moles; skin type; ability to burn; ability to tan; previous skin cancers; ultraviolet exposure between the ages of 5 and 12 years; cumulative sun exposure up to the age of 19 years; and birth year.

The second stage fitted multivariate proportional hazard models to those explanatory variables that were significant at the univariate stage. Table 1 displays the results from the final model, which only included significant variables (p-value < 0.05). It is worth noting that a similar result could be obtained using an automated stepwise procedure.

Results from this analysis highlight some interesting but quite obvious risk factors. For example,

*An increase of one year in birth year induces 17% increase in risk of earlier melanoma onset.*

*People with no freckles nor naevi have the lowest risk of melanoma onset.*

*The risk of earlier melanoma onset is increased by up to 36% for blue eyed people and even further (46%) for green eyed people, when compared to individuals with brown eyes.*

*“Red Heads” have an increased risk of earlier melanoma onset (46%) when compared to individuals with black hair. There was no significant increase noted however for individuals with fair or light red hair.*

*A person's ability to burn easily increases the risk of earlier melanoma onset, in some cases by up to 100% compared to those that never burn.*

However striking this last statement is, issues of confounding, must also be considered. The most obvious illustration of this is the confounding that occurred between mole count and freckling. This is seen through close inspection of the parameter estimates which changed in magnitude when mole count was added to the model after adjusting for freckling. (Results not displayed here.)

To reduce the dimension of the problem further and avoid some of these confounding issues, a survival tree was constructed using RPART (Recursive Partitioning and Regression Trees) (Therneau and Atkinson, 1997).

From this model we can see that there are a few scenarios that indicate high risk for earlier onset of melanoma. These scenarios may be described as,

- individuals born after 1966 with many moles (RR=11.3)
- individuals born between 1947 and 1966 with many moles (RR=4.5)
- individuals born between 1933 and 1953 with few moles, but many freckles (RR=2.4)
- individuals born between 1923 and 1933 with many moles (RR=1.41)

It is obvious from these results that birth year has a substantial impact on the age-at onset of melanoma. Once this is taken into account, mole count and freckling only provide a small contribution to the risk.

### 2.3 Family history of melanoma

As described earlier, in the Data section, family history was collected regarding first-degree (siblings and parents) and second-degree relatives. From this pedigree structure, other higher-order types of relative pairs could also be formed. Some of these relatives were diseased with melanoma, resulting in pairs of relatives who both may be diseased (++) , both not diseased (-), and one diseased while the other was not diseased (+-). Table 2 lists the concordant and discordant pairs of specific relationships: sib-sib, parent-child, and second-degree/lower order relative pairs. The second-degree relative pairs include grandparent-grandchild, and aunt-niece, while the lower order relative pairs include the in-law pairs. From table 2, the percentages of both diseased pairs are 0.6%, 0.9% and 0.4% among sib-sib, parent-child, and second/lower order pairs, respectively. Crude estimates of correlations coefficients can be calculated from these percentages, without accounting for ages at onset among these relatives. However, the risk of developing melanoma may depend on the subject's age. Hence, adjusting for the age at onset is essential in quantifying the correlation of age at onset between pairs of relatives. In a subsequent section, we describe how this can be done rigorously via the generalized estimating equations approach. On average, melanomas were diagnosed slightly earlier in relatives (47.5 years) than in probands (50.2 years). Among relatives, melanomas were diagnosed at younger ages in later generations. To account for the different ages at censoring in each generation due to termination of the study or death from causes other than melanoma, we examined the disease-free survival distribution for each generation using the standard failure-time analysis technique. The median age at diagnosis of melanoma was 64 among parents of probands, 50 among siblings of probands, and 33 among children of probands. The disease-free survival functions differed significantly between the children generation from earlier generations (log-rank test with  $p < 0.01$ ); but there was no significant differences in age-at-onset of melanoma between the siblings and parents of probands (log rank test with  $p > 0.9$ ).

### 2.4 Preliminary segregation analysis

A preliminary segregation analysis was conducted (Aitken et al., 1998) to investigate whether the familial clustering of cutaneous melanoma is consistent with Mendelian inheritance of a major autosomal gene. Analyses were performed with the S.A.G.E. (SAG, 1992) statistical package, using the maximum likelihood REGTL program for a binary trait with a variable age of onset. The hypothesis of codominant Mendelian inheritance

gave a significantly better fit to the data than either dominant or recessive Mendelian inheritance. Overall, both Mendelian inheritance of a single major gene, and purely environmental transmission were rejected. However, there was strong evidence of familial dependence in melanoma occurrence.

The inclusion of risk factors in the models may reveal whether all or a combination of these explains the familial dependence that was demonstrated by the segregation analysis. If a major gene exists that operates independently on these covariates, including them in the models may reduce residual variation and increase the power of the analysis to detect a major gene effect.

## 2.5 Estimating equations approach

Let  $y = (\delta_{ki}, t_{ki}, \mathbf{Z}_{ki})$  denote the data collected for the  $i^{th}$  member in the  $k^{th}$  family ( $k = 1, \dots, K$  and  $i = 1, 2$ ) where  $\delta_{ki} = 0$  if the observation is censored,  $t_{ki}$  is either the recorded age at diagnosis of melanoma or the age at the most recent follow-up for unaffected women, and  $\mathbf{Z}_{ki}$  is a vector of measured covariates. We assume that censoring time, age at diagnosis of melanoma and the covariates are independently distributed. These assumptions can be relaxed in more general models, subject to identification constraints. The hazard rate for melanoma is the instantaneous probability that melanoma is diagnosed immediately after time  $t$ , given that the woman is unaffected at time  $t$ . The hazard rate under the Cox proportional hazards model (Cox, 1972) is given by

$$\lambda(t_{ki}) = \lambda_0(t_{ki}) \exp(\beta' \mathbf{Z}_{ki}),$$

where  $\lambda_0()$  is the baseline hazard function, and  $\beta$  is a vector of regression coefficients.

For a specific pair of relatives, we follow Clayton (Clayton, 1978) in modeling the bivariate survivor function

$$F(t_{k1}, t_{k2}) = (F_1(t_{k1})^{-\theta} + F_2(t_{k2})^{-\theta} - 1)^{-1/\theta},$$

where  $F_1$  and  $F_2$  are univariate survivor functions,  $\theta$  is a scalar parameter that measures the degree of dependence between the relatives' times at onset, independence being implied by  $\theta = 0$ , and positive association by  $\theta > 0$ . The Clayton model allows negative dependencies and has the property that failure times are absolutely continuous for  $\theta > -0.5$ . In addition, the cross-ratio (or odds-ratio) function as studied by Oakes (Oakes, 1989) is

$$c(t_{k1}, t_{k2}) = \lambda(t_{k1}|T_{k2} = t_{k2}) / \lambda(t_{k1}|T_{k2} \geq t_{k2}) = 1 + \theta.$$

This is equivalent to assuming that the odds-ratio is invariant over the grid region that supports the data. Heuristically, the parameter  $1 + \theta$  is an odds-ratio that depends on the degree of dependence between the onset ages of the two relatives. If genetic factors do influence the age at onset of melanoma, we would expect to see a higher concordance in the age of onset in first degree relatives who on average, share half their genes in common in comparison to second degree relatives. Under the current model, this translates as  $\theta_{first-degree} > \theta_{second-degree}$ .

We may use a standard method to estimate within pair correlations for 2X2 tables from odds ratios. Estimates of relation-pair correlations  $\rho_{sib-sib}$ ,  $\rho_{parent-child}$ , and  $\rho_{second-order}$  are recovered from using the relationships, e.g.,  $r_{sib-sib} = \min(1, \ln(1 + \theta_{sib-sib}))$ . Testing for the presence of genetic factors underlying the age at diagnosis of melanoma is equivalent to testing  $H_0 : \rho_{first-order} = \rho_{second-order}$ . We may test this hypothesis using a  $z$ -transform (Kendall and Stuart, 1979) (p. 315) of the point estimates of the correlation coefficients. Let  $n_1$  and  $n_2$  denote the number of first order and second order relatives, let  $z_1$  and  $z_2$  denote the transformed statistics of  $r_1$  and  $r_2$ , the correlation estimates for first-order and second-order relative pairs respectively. Specifically, we reject  $H_0$  when  $E/D > Z_{1-\alpha}$ , where

$$E = E(z_1 - z_2) = \frac{1}{2} \log \left[ \left( \frac{1+r_1}{1-r_1} \right) \left( \frac{1-r_2}{1+r_2} \right) \right],$$

$$D^2 = V(z_1 - z_2) = \frac{1}{(n_1 - 3)} + \frac{1}{(n_2 - 3)},$$

and  $Z_{1-\alpha}$  is the standard normal deviate corresponding to the one-sided  $\alpha$  significance level.

This approach has the advantage of providing a test for the presence of genetic effects through a single parameter ( $\theta$ ). However, it is limited in its ability to attribute the phenotypic variance to specific effects (e.g., additive gene action).

Mathematical details of the GEE model and the iterative procedure to estimate the regression coefficients  $\beta$  and specific degrees of dependence  $\theta$  for the different types of relative pairs are available from the first author.

## 2.6 MCMC analysis using BUGS

### *The Bayesian paradigm and Gibbs sampling*

Markov Chain Monte Carlo (MCMC) is an alternative Bayesian approach that provides estimates of likelihoods and associated parameter values when exact computation is infeasible (Metropolis et al., 1953; Hastings, 1970). MCMC methods can be used to draw samples from the underlying joint distribution of major genotypes and polygenic values, conditional on the observed data. From these samples, desired parameters and likelihoods can be estimated without the need to resort to exact computation. MCMC methods have been used for linkage analysis (Lange and Sobel, 1991; Kong et al., 1992), for estimation of parameters in the mixed model with and without covariates (Guo and Thompson, 1991; Thomas, 1992), for estimation of gene-smoking interaction and covariate imputation (Gauderman et al., 1997), for performing combined linkage and segregation analysis (Guo and Thompson, 1992; Faucett et al., 1993), and for mixed models of large complex pedigrees (Guo and Thompson, 1994).

The Gibbs sampler (Geman and Geman, 1984) is the most popular algorithm used in MCMC applications to correlated data. Gibbs sampling was introduced to the main statistical community by Gelfand and Smith (Gelfand and Smith, 1990), and has since been applied in even a wider array of problems. The Gibbs sampler is easy to implement because it only depends on the local neighborhood structure. In the context of pedigree analysis (Olshen and Wijsman, 1996), the basic procedure is a sequential updating of missing and latent data including the underlying and unobserved major genotypes, polygenic effects, and environmental effects. Values for the missing or latent data are sampled from the local conditional distribution, a function of the observed individual data, the current sampled values of other missing/latent data for this particular individual such as polygenic and environmental effects, and the values for the sampled genetic effects in the immediate neighbors of an individual.

Some of the most recent and popular packages that implement Gibbs sampling for analysis of pedigree data include BUGS (Gilks et al., 1994; Spiegelhalter et al., 1996b; Spiegelhalter et al., 1996a), Genetic Analysis Package (GAP, 1996), and MIXD (Thompson, 1994; Olshen and Wijsman, 1996). We have used BUGS mainly because of its flexibility in programming hierarchical models besides being a freeware product.

### *The Model*

The aim here is to model the correlation structure within the family structure to satisfy the fundamental additive genetic model (Crow and Kimura, 1970; Falconer, 1990; Kempthorne, 1960) as follows. Consider a nuclear family structure consisting of 4 members: father, mother, and two children denoted by  $F, M, S_1$ , and  $S_2$  respectively. Using similar notation as in Burton *et al* (1999), a conventional mixed linear model consisting of fixed and random effects may be written in the form

$$Q_{ij} = \beta' \mathbf{z} + A_{ij} + C_{ij} + C_{S_{ij}} + E_{ij} \quad (2)$$

where  $Q_{ij}$  is the observed value of a normally distributed continuous trait for the  $j^{th}$  individual in the  $i^{th}$  nuclear family;  $\mathbf{z}_{ij}$  is a vector of observed covariates representing fixed effects, and  $\beta$  is a corresponding vector of unknown fixed regression coefficients;

$A_{ij}$ ,  $C_{ij}$ , and  $C_{S_{ij}}$  denote random effects that represent additive polygenic, common family environment, and common sibling environment effects respectively. The variation in an individual response is represented by a composite covariance matrix,  $V_T$ , and is the sum of an additive genetic covariance matrix  $V_A$ , a common family environment matrix  $V_C$ , a shared sibling environment matrix  $V_{C_s}$ , and residual environmental effects. The different variance components are

$$V_A = \begin{matrix} & F & M & S_1 & S_2 \\ \begin{matrix} F \\ M \\ S_1 \\ S_2 \end{matrix} & \begin{pmatrix} \sigma_A^2 & 0 & \frac{1}{2}\sigma_A^2 & \frac{1}{2}\sigma_A^2 \\ 0 & \sigma_A^2 & \frac{1}{2}\sigma_A^2 & \frac{1}{2}\sigma_A^2 \\ \frac{1}{2}\sigma_A^2 & \frac{1}{2}\sigma_A^2 & \sigma_A^2 & \frac{1}{2}\sigma_A^2 \\ \frac{1}{2}\sigma_A^2 & \frac{1}{2}\sigma_A^2 & \frac{1}{2}\sigma_A^2 & \sigma_A^2 \end{pmatrix} \end{matrix}$$

$$V_C = \begin{matrix} & F & M & S_1 & S_2 \\ \begin{matrix} F \\ M \\ S_1 \\ S_2 \end{matrix} & \begin{pmatrix} \sigma_C^2 & \sigma_C^2 & \sigma_C^2 & \sigma_C^2 \\ \sigma_C^2 & \sigma_C^2 & \sigma_C^2 & \sigma_C^2 \\ \sigma_C^2 & \sigma_C^2 & \sigma_C^2 & \sigma_C^2 \\ \sigma_C^2 & \sigma_C^2 & \sigma_C^2 & \sigma_C^2 \end{pmatrix} \end{matrix}$$

$$V_{C_s} = \begin{matrix} & F & M & S_1 & S_2 \\ \begin{matrix} F \\ M \\ S_1 \\ S_2 \end{matrix} & \begin{pmatrix} \sigma_{C_s}^2 & 0 & 0 & 0 \\ 0 & \sigma_{C_s}^2 & 0 & 0 \\ 0 & 0 & \sigma_{C_s}^2 & \sigma_{C_s}^2 \\ 0 & 0 & \sigma_{C_s}^2 & \sigma_{C_s}^2 \end{pmatrix} \end{matrix}$$

The overall total covariance matrix is

$$V_T = \begin{matrix} & F & M & S_1 & S_2 \\ \begin{matrix} F \\ M \\ S_1 \\ S_2 \end{matrix} & \begin{pmatrix} \sigma_A^2 + \sigma_C^2 + \sigma_{C_s}^2 + \sigma_{EP}^2 & \sigma_C^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 \\ \sigma_C^2 & \sigma_A^2 + \sigma_C^2 + \sigma_{EP}^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 \\ \frac{1}{2}\sigma_A^2 + \sigma_C^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 & \sigma_A^2 + \sigma_C^2 + \sigma_{EC}^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 + \sigma_{C_s}^2 \\ \frac{1}{2}\sigma_A^2 + \sigma_C^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 + \sigma_{C_s}^2 & \sigma_A^2 + \sigma_C^2 + \sigma_{EC}^2 + \sigma_{C_s}^2 \end{pmatrix} \end{matrix}$$

The *components of variance*  $\sigma_A^2, \sigma_C^2, \sigma_{C_s}^2$  need not be positive as long as  $V_T$  is positive definite. A negative value for  $\sigma_{C_s}^2$  simply implies that the realized covariance between siblings is *less* than the realized covariance between a parent and a child. A negative value for  $\sigma_C^2$  may suggest dominance. To optimize convergence in BUGS, Model (1) may be reparameterized as:

$$Q_{ij} = \begin{cases} \alpha + \beta' \mathbf{z} + F_i + G_i + R_{ij}^P & \text{for fathers} \\ \alpha + \beta' \mathbf{z} + F_i - G_i + R_{ij}^P & \text{for mothers} \\ \alpha + \beta' \mathbf{z} + F_i + H_i + R_{ij}^C & \text{for children} \end{cases} \quad (3)$$

where  $F_i, G_i, H_i$  are independent additive random effects or latent variables;  $R_{ij}^P$  and  $R_{ij}^C$  are the residual error terms for parents and children respectively. If we model  $F_i \sim N(0, \frac{1}{2}\sigma_A^2 + \sigma_C^2)$ ,  $G_i \sim N(0, \frac{1}{2}\sigma_A^2)$ ,  $H_i \sim N(0, \sigma_C^2)$ , then the basic genetics covariance model (additive genetic, common environment and unique environment) for this particular four-member family structure is satisfied.

In survival models, unobserved or unmeasured explanatory variables, some of which may be genetic, are often referred to as frailties. The frailties take values restricted to the positive line and may be assumed to act multiplicatively on the hazard. Extending the above model to correlated family data with time-to-onset endpoint, a multiplicative individual heterogeneity or frailty term representing the latent genetic and common environment variables may be modeled as random effects simultaneously with the effects associated with observed covariates. Consider right censored time to onset of melanoma data  $\{(T_{ij}, \delta_{ij}, \mathbf{z}_{ij}); 1 \leq j \leq n\}$  from  $n$  relative pairs; here  $T_{ij}$  denotes the true age at onset of the  $j^{\text{th}}$  family member or the censored time depending on whether  $\delta_{ij} = 1$  or 0 respectively, and  $\mathbf{z}$  denotes a  $p \times 1$  vector of covariates. A Weibull distribution may be used to model time to failure as

$$f(t_i, \mathbf{z}_i) = e^{\beta' \mathbf{z}_i} \gamma t_i^{\gamma-1} \exp(-e^{\beta' \mathbf{z}_i} t_i^\gamma), \quad (4)$$

where  $\beta$  is a vector of unknown regression coefficients, and  $\gamma$  is the shape parameter of the Weibull distribution. This leads to a baseline hazard of the form

$$\lambda_0(t_i) = \gamma t_i^{\gamma-1}.$$

Reparameterize by letting  $\mu_i = e^{\beta' \mathbf{z}_i}$ , the conditional distribution of  $t_i$  given  $\mu_i$  is then Weibull( $\gamma, \mu_i$ ). We formulated a mixed model to represent the conditional distribution of  $t_{ij}$  given covariate effects, random additive genetic and common environment effects as

$$t_{ij} | \mu_{ij} \sim Weibull(\gamma, \mu_{ij}) \quad i = 1, \dots, n; j = 1, 2$$

where

$$\log \mu_{ij} = \begin{cases} \alpha + \beta' \mathbf{z} + F_i + G_i + R_{ij}^P & \text{for fathers} \\ \alpha + \beta' \mathbf{z} + F_i - G_i + R_{ij}^P & \text{for mothers} \\ \alpha + \beta' \mathbf{z} + F_i + H_i + R_{ij}^C & \text{for children} \end{cases} \quad (5)$$

The regression coefficients and the precision of the random effects ( $\tau_G, \tau_F, \tau_H$ ) were given “non-informative” Normal and Gamma priors respectively. The shape parameter,  $\gamma$ , of the time to onset of melanoma distribution was also given a non-informative Gamma prior which was slowly decreasing on the positive real line.

We implemented the Gibbs sampler using the BUGS program (Gilks et al., 1994). Imputation of missing data was handled naturally in the Gibbs sampling framework by treating missing values as additional unknown quantities and randomly sampling values from their full conditional distributions. We chose simple prior distributions for imputation, since the number of missing values for covariates was relatively small (complete for birth year, 16% for naevi and 21% missing for freckles) and there was no indication of non-random missingness in our data. Therefore imputation for missing naevi and freckles covariates were based on Bernoulli prior distributions with respective parameter values estimated from the complete observations. We performed an initial 10,000 burn-in iterations followed by an additional 20,000. Parameter estimates were the mean and standard deviation (SD) of all post convergence Gibbs samples with a thinning interval between 20 and 50; credible intervals were computed as the lower and upper  $\alpha/2$  percentiles from the last 20,000 iterations. Convergence to the posterior distribution was confirmed by using the different criteria provided by the add-on CODA package including those of Gelman & Rubin (Gelman and Rubin, 1992), Geweke (Geweke, 1992), Raftery & Lewis (Raftery and Lewis, 1992a; Raftery and Lewis, 1992b).

## 3 RESULTS

### 3.1 GEE approach

An inspection of residual plots following preliminary model-fitting provided no evidence for the failure of proportional hazards assumption and did not detect influential observations. We then proceeded to apply the GEE approach that could estimate regression coefficients while incorporating a dependence structure between relative pairs. The results are displayed in Table 3 which suggest that later birth year, having at least a moderate number of naevi and freckles were all simultaneously associated with later age at onset of melanoma. Table 3 also presents the estimated odds-ratios for quantifying the correlation between paired relatives of specific relationships. The odds ratio for sib-sib pairs is 2.973 which is significantly different from 1 ( $P < 0.01$ ). The odds ratio for parent-child pairs is 1.650 which is slightly greater than 1, indicating a mild dependency between these pairs, although they are not quite statistically significant. The odds ratio for second-degree and higher relative pairs is 1.155, indicating no dependence at all between these pairs. This pattern of familial aggregation, in view of genetics (Falconer, 1990), indicates that there is a dominance variance in addition to the genetic additive variance.

### 3.2 Bayesian approach

We re-analyzed the age-at-onset of melanoma by using Gibbs sampling to impute missing covariates and to estimate subject-specific covariate effects, random additive genetic, common family environment, and shared sibling environment effects on the log scale. The estimated shape parameter  $\gamma$  was 14.3 with a 95% credible interval of (14.2, 14.6). The results are summarized in Table 4 (Model A). The residual plots did not indicate a gross departure from the underlying Weibull model and revealed no influential observation. We checked the sensitivity of the analyses to initial parameter values by re-running the Gibbs sampler five more times using different starting values. The resulting estimates did not differ by more than 5% from the values reported here. The mean estimate for  $(\sigma_A^2)$  was 0.452 with 95% CI = (0.348, 0.566), for  $(\sigma_C^2)$  was -0.223 with 95% CI = (-0.282, -0.169), and for  $(\sigma_{C_s}^2)$  was 0.467 with 95% CI = (0.393, 0.545). The results here indicate that additive genetics and shared sibling environment seem to impact equally on the variation of the age at onset of melanoma. A negative value for the common family effect suggests siblings are more correlated than parent-offspring pairs. This implies that other models should be investigated such as the dominance model or a pure environmental model.

In addition, we investigated the relative contribution of genetic and environmental effects on the expression of naevi (Model B) and freckles (Model C), which are known risk factors for melanoma. The expressions of naevi and freckles were coded as binary variables (none or few moles versus moderate or many moles; and no freckles versus one or more freckles). A hierarchical Bayesian binomial model was fitted to estimate the random variance components. The results in Table 4 indicate that common family environment effect contributed the most to the expression of naevi ( $\sigma_C^2 = 0.704$ ) relative to the contributions of additive genetic effect ( $\sigma_A^2 = 0.142$ ) and of shared sibling effect ( $\sigma_{C_s}^2 = 0.142$ ), both of which were non negligible. In contrast, the variation in the expression of freckles were largely explained by additive genetic and shared family effects ( $\sigma_A^2 = 2.050$ ,  $\sigma_C^2 = 2.600$ ), compared to a relatively small shared sibling effect ( $\sigma_{C_s}^2 = 0.115$ ).

## 4 DISCUSSION

We applied two methods - generalized estimating equation and Bayesian analysis - to the genetic analysis of age at onset of melanoma based on a nuclear family structure. Under both approaches, the results suggest that additive genetic factors played an important role in the age at onset of melanoma and that shared sibling environmental factors were not negligible. We focussed attention on these approaches because they are more appropriate

for modeling correlated age at onset data and they allow the inclusion of covariates in the analyses. Under both approaches, there were suggestions that earlier melanoma onset was influenced by later birth year, having a moderate number of naevi, and being freckly. The principal difference between the two approaches is in the interpretation of the regression coefficients. The GEE method uses a marginal approach resulting in regression coefficients that describe the average population response to changing covariates, whereas the Bayesian approach produces subject-specific coefficients. A secondary distinction is in the nature of the within-pair dependence. The GEE model only describes a common covariance among specific relative pairs, whereas the Bayesian approach can explicitly describe the source of this covariance. A third advantage of the Bayesian method is its flexibility in incorporating prior information, if available, for the covariates or latent effects by modifying their prior distributions. Further, the Bayesian method would permit a more accurate decomposition of the genetic variance into additive and dominant components and thus provide the means for a direct assessment of the no-dominance assumption. Finally, it is also interesting to record the amount of CPU time required for each method: 20 seconds for the GEE approach and approximately 3 hours (for binary traits) and 12 hours (for age at onset outcome) to run BUGS on a single-user Intel Pentium III 600 MHz personal computer with Linux Mandrake 7.1 operating system. In our opinion, the amount of human and financial resources dedicated to collecting, maintaining, and updating the melanoma family database is extremely high. Therefore, the extra CPU time requirement by the MCMC method is well worth the additional genetic information and flexibility that it provides.

**Acknowledgments:** This work was partially supported by the Australian NHMRC grants ID 930223 and ID 961061; and an NIH grant HG02275.

## References

- (1992). *S.A.G.E. Statistical Analysis for Genetic Epidemiology, Release 2.1*.
- (1996). *Genetic Analysis Package, Release 1.0*. Epicenter Software, Pasadena, California. Computer program.
- Abel, L. and Bonney, G. E. (1990). A time-dependent logistic hazard function for modelling variable age of onset in the analysis of familial diseases. *Genetic Epidemiology*, **12**, 391–407.
- Aitken, J. F., Bailey-Wilson, J., Green, A. C., MacLennan, R., and Martin, N. G. (1998). Segregation analysis of cutaneous melanoma in Queensland. *Genet Epidemiol*, **15**, 391–401.
- Aitken, J. F., Green, A. C., MacLennan, R., Youl, P., and Martin, N. G. (1996). The Queensland familial melanoma project: study design and characteristics of participants. *Melanoma Research*, **6**, 155–165.
- Allison, P. D. (1995). *Survival Analysis Using the SAS System - A Practical Guide*. SAS Institute Inc., Cary, NC.
- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- Bataille, V., Snieder, H., MacGregor, A. J., Sasieni, P., and Spector, T. D. (2000). Genetics of risk factors for melanoma: an adult twin study of nevi and freckles. *J National Cancer Inst*, **92**, 457–463.
- Bernardo, J. M., Berber, J. O., Dawid, A. P., and Smith, A. F. M., editors (1992). *Bayesian Statistics 4*. Clarendon Press, Oxford, UK.
- Bonney, G. E. (1986). Regressive logistic model for familial disease and other binary traits. *Biometrics*, **42**, 611–625.
- Cannings, C. and Thompson, E. A. (1977). Ascertainment in the sequential sampling of pedigrees. *Clin Genet*, **12**, 208–212.

- Clayton, D. G. (1978). A model for association in bivariate life-tables and its application in epidemiological studies of chronic disease incidence. *Biometrika*, **65**, 141–151.
- Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *J R Statist Soc A*, **148**, 82–117.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *J R Statist Soc B*, **34**, 187–202.
- Crow, J. F. and Kimura, M. (1970). *Introduction to Population Genetics Theory*. Harper and Row, New York.
- Do, K.-A., Broom, B. M., Kuhnert, P., Duffy, D. L., Todorov, A. A., Treloar, S. A., and Martin, N. G. (2000). Genetic analysis of the age at menopause by using estimating equations and Bayesian random effects models. *Statistics in Medicine*, **19**, 1217–1235.
- Falconer, D. S. (1990). *Introduction to Quantitative Genetics*. Longman Group Ltd., New York, 3rd edition.
- Faucett, C., Gauderman, W. J., Thomas, D., Ziogas, A., and Sobel, E. (1993). Combined segregation and linkage analysis of late-onset Alzheimer's disease in Duke families using Gibbs sampling. *Genet Epidemiol*, **10**, 489–494.
- Gauderman, W. J., Morrison, J. L., Carpenter, C. L., and Thomas, D. C. (1997). Analysis of gene-smoking interaction in lung cancer. *Genet Epidemiol*, **14**, 199–214.
- Gauderman, W. J. and Thomas, D. C. (1994). Censored survival models for genetic epidemiology: a Gibbs sampling approach. *Genet Epidemiol*, **11**, 171–188.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J Amer Statist Assoc*, **85**, 398–409.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Stat Sci*, **7**, 457–511.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intelligence*, **6**, 721–741.
- Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*. In (Bernardo et al., 1992).
- Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *Statistica*, **43**, 169–178.
- Guo, S. W. and Thompson, E. A. (1991). Monte Carlo EM for the estimation of multiple random effects models on large pedigrees. *IMA J Meth Appl Med Biol*, **8**, 171–189.
- Guo, S. W. and Thompson, E. A. (1992). A Monte Carlo method for combined segregation and linkage analysis. *Am J Hum Genet*, **51**, 1111–1126.
- Guo, S. W. and Thompson, E. A. (1994). Monte Carlo estimation of mixed models for large complex pedigrees. *Biometrics*, **50**, 417–432.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.
- Heath, A. C. (1997). Markov Chain Monte Carlo segregation and linkage analysis for oligogenic models. *American Journal of Human Genetics*, **61**, 748–760.
- Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, **73**, 671–678.
- Hsu, L. and Prentice, R. L. (1996). On assessing the strength of dependency between failure time variates. *Biometrika*, **83**, 491–506.
- Hsu, L. and Zhao, L. P. (1996). Assessing familial aggregation of age at onset, by using estimating equations, with application to breast cancer. *Am J Hum Genet*, **58**, 1057–1071.

- Kempthorne, O. (1960). *Biometrical Genetics*. Pergammon Press, New York.
- Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics*, volume 2. Oxford University Press, Oxford, 4th edition.
- Kong, K., Frigge, M., Cox, N., and Wong, W. H. (1992). Linkage analysis with adjustments for covariates: A method combining peeling with Gibbs sampling. *Cytogenet Cell Genet*, **59**, 208–210.
- Lange, K. and Sobel, E. (1991). A random walk method for computing genetic location scores. *Am J Hum Genet*, **49**, 1320–1334.
- Li, H. and Thompson, E. A. (1997). Semiparametric estimation of major gene and random environmental effects for age of onset. *Biometrics*, **53**, 282–293.
- Li, H. and Wijsman, E. A. (1998). Semiparametric estimation of major gene effects for age of onset. *Genet Epidemiol*, **15**, 279–298.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J Chem Phys*, **21**, 1087–1092.
- Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorenson, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand J Statist*, **19**, 25–43.
- Oakes, D. (1989). Bivariate survival models induced by frailties. *J Amer Statist Assoc*, **84**, 487–493.
- Olshen, A. B. and Wijsman, E. M. (1996). Pedigree analysis package vs. MIXD: Fitting the mixed model on a large pedigree. *Genet Epidemiol*, **13**, 91–106.
- Raftery, A. L. and Lewis, S. (1992a). Comment: One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Stat Sci*, **7**, 493–497.
- Raftery, A. L. and Lewis, S. (1992b). *How many interactions in the Gibbs sampler?*, pages 763–774. In (Bernardo et al., 1992).
- Siegmund, K. D. and McKnight, B. (1998). Modelling hazard functions in families. *Genetic Epidemiology*, **15**, 147–171.
- Siegmund, K. D., Province, M. A., Williams, R., Keller, J., and Todorov, A. A. (1998a). Modeling disease incidence rates in families. *Epidemiology*, **9**, 557–562.
- Siegmund, K. D., Todorov, A. A., and Province, M. A. (1998b). A frailty approach for modelling diseases with variable age of onset in families: The NHLBI family heart study. *Statistics in Medicine*, **17**, In press.
- Spiegelhalter, D. J., Thomas, A., and Best, N. G. (1996a). *Computation on Bayesian Graphical Models*, pages 407–425. Oxford University Press, Oxford.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. (1996b). *BUGS 0.6 - Bayesian Inference Using Gibbs Sampling Manual*.
- Therneau, T. M. and Atkinson, E. (1997). An introduction to recursive partitioning using the RPART routines. Technical report, Mayo Foundation, Rochester.
- Thomas, D. (1992). Fitting genetic data using Gibbs sampling: An application to nevus counts in 38 Utah kindreds. *Cytogenet Cell Genet*, **59**, 228–230.
- Thompson, E. A. (1994). Monte Carlo programs for pedigree analysis. Technical Report 267, Dept. of Statistics, University of Washington.
- Vaupel, J. M., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**, 439–454.
- Yashin, A. I. and Iachine, I. A. (1995). Genetic analysis of durations: Correlated frailty model applied to survival of Danish twins. *Genet Epidemiol*, **12**, 529–538.

Table 1: Results from fitting a multivariate proportional hazards model to the melanoma data based on univariate results. The results reported in this table are the parameter estimates  $\beta$ , their standard errors  $se(\beta)$ , the relative risk  $e^\beta$  and the p-value for each estimate.

Variable	$\beta$	$se(\beta)$	$e^\beta$	p-value
Birth Year	0.16	0.00	1.17	< 0.05
Eye Colour (Baseline: <b>Brown</b> )				
- Blue/Grey	0.31	0.07	1.36	< 0.05
- Green/Hazel	0.38	0.07	1.46	< 0.05
Hair Colour (Baseline: <b>Black</b> )				
- Light Red/Ginger	0.17	0.15	1.19	0.27
- Dark Red/Auburn	0.38	0.15	1.46	< 0.05
- Fair/Blonde	0.06	0.12	1.06	0.62
- Light Brown	0.14	0.12	1.15	0.22
- Dark Brown	0.02	0.12	1.02	0.87
Skin Type (Baseline: <b>never burn</b> )				
- always burn	0.69	0.16	1.98	< 0.05
- usually burn	0.45	0.15	1.57	< 0.05
- sometime burn	0.31	0.15	1.36	< 0.05
Freckling (Baseline: <b>none</b> )				
- 1 to 100	0.17	0.06	1.18	< 0.05
- > 100	0.09	0.08	1.10	0.23
Mole Count (Baseline: <b>none</b> )				
- few	0.29	0.07	1.34	< 0.05
- moderate	0.79	0.08	2.20	< 0.05
- many	1.12	0.10	3.08	< 0.05
Number of Sunburns (Baseline: <b>none</b> )				
- one	-0.07	0.11	0.93	0.49
- 2 to 5	-0.06	0.09	0.94	0.50
- > 6	0.17	0.09	1.19	0.07
Cumulative Sun Exposure (< 5 yrs)	0.04	0.01	1.04	< 0.05
UV Exposure (5-12 yrs)	0.0003	0.00	1	< 0.05

Table 2: **Concordant and Discordant Pairs of Relatives in 1912 Families from the Queensland Familial Melanoma Project.** Probands are not included for the calculation of concordancy.

	Sib-Sib	Parent-Child	Second/Others	Total
++	49	15	41	105
+-	763	536	1200	2499
-	7011	1078	9817	17906
Total	7823	1629	11058	20510

Table 3: **GEE approach:** Estimated Regression Coefficients in the Proportional Hazard Model and Estimated Odds Ratios for Quantifying Familial Aggregation in Age at Onset of Melanoma in Queensland Families (\*\* indicates significance).

Covariate	RR = $e^\beta$	Coefficient $\beta$	Robust se( $\beta$ )	Z-statistic
<i>A. Mean effects</i>				
<i>Year of birth</i>	1.142	0.132	0.051	2.588 **
<i>Naevi (Baseline = No or few moles)</i>	1.765	0.568	0.073	7.781 **
<i>Freckling (Baseline = No freckles)</i>	1.160	0.148	0.049	3.020 **
<i>B. Patterns of familial aggregation</i>				
Relationship		$1 + \theta$	se( $\theta$ )	Z-Statistic
<i>Sib-sib</i>		2.973	0.6217	3.17 **
<i>Parent-child</i>		1.650	0.434	1.50 **
<i>Second/Others</i>		1.155	0.3270	0.47

Table 4: **Gibbs Sampling Approach:** Estimated Regression Coefficients and Estimated Variance Components in a Melanoma Study of Queensland Families (\*\* indicates significance). *Naevi* is a binary variable with Baseline 0 = No or few moles; *Freckling* is coded as a binary variable with Baseline 0 = No freckles.

Covariate	RR = $e^\beta$	Coefficient $\beta$	Robust se( $\beta$ )	95% CI of $\beta$
<i>Weibull Model: A. Mean effects - Response variable is Age-at-onset</i>				
<i>Year of birth</i>	1.378	0.321	0.0027	(0.316,0.326) **
<i>Naevi</i>	1.126	0.119	0.0021	(0.058,0.185) **
<i>Freckling</i>	1.017	0.017	0.1400	(-0.055,0.085)
<i>Weibull Model: B: Variance components - Response variable is Age-at-onset</i>				
Latent effect	Mean from 5000 iterations		se( $\sigma^2$ )	95% CI of $\sigma^2$
$\sigma_A^2$		0.452	0.054	(0.348,0.566) **
$\sigma_C^2$		-0.223	0.027	(-0.282,-0.169)
$\sigma_{C_s}^2$		0.467	0.040	(0.393,0.545) **
$\gamma$		14.3	0.104	(14.2, 14.6)
<i>Binomial Model: Variance components - Response variable is Naevi</i>				
Latent effect	Mean from 5000 iterations		se( $\sigma^2$ )	95% CI of $\sigma^2$
$\sigma_A^2$		0.142	0.149	(0.002,0.498) **
$\sigma_C^2$		0.704	0.156	(0.403,1.010) **
$\sigma_{C_s}^2$		0.195	0.157	(0.0025,0.553) **
<i>Binomial Model: Variance components - Response variable is Freckling</i>				
Latent effect	Mean from 5000 iterations		se( $\sigma^2$ )	95% CI of $\sigma^2$
$\sigma_A^2$		2.050	0.779	(0.835,3.570) **
$\sigma_C^2$		2.600	0.418	(1.780,3.460) **
$\sigma_C^2$		0.115	0.088	(0.011,0.312) **