Genetic variance components analysis for twin and family studies: 1. Gibbs sampling versus other methods 2. Applications to melanoma and menopause data

Kim-Anh DO
Department of Biostatistics
The University of Texas MD Anderson Cancer Center
Email: kim@mdanderson.org

Main References

Do *et al* (2000) *Stat in Med*, pp 1217-1235.

Burton *et al* (1999), *Genetic Epidemiology*: 118-140.

Important focus of genetic research: common complex diseases such as asthma, cancer.

Studies based on large numbers of twins or simple pedigrees ascertained from population-based sampling frames are becoming commonplace.

Many genetic and environmental factors causing these conditions are unknown.

Even after all known determinants have been taken into account, there is often a strong residual covariance between relatives.

Basic family structures:

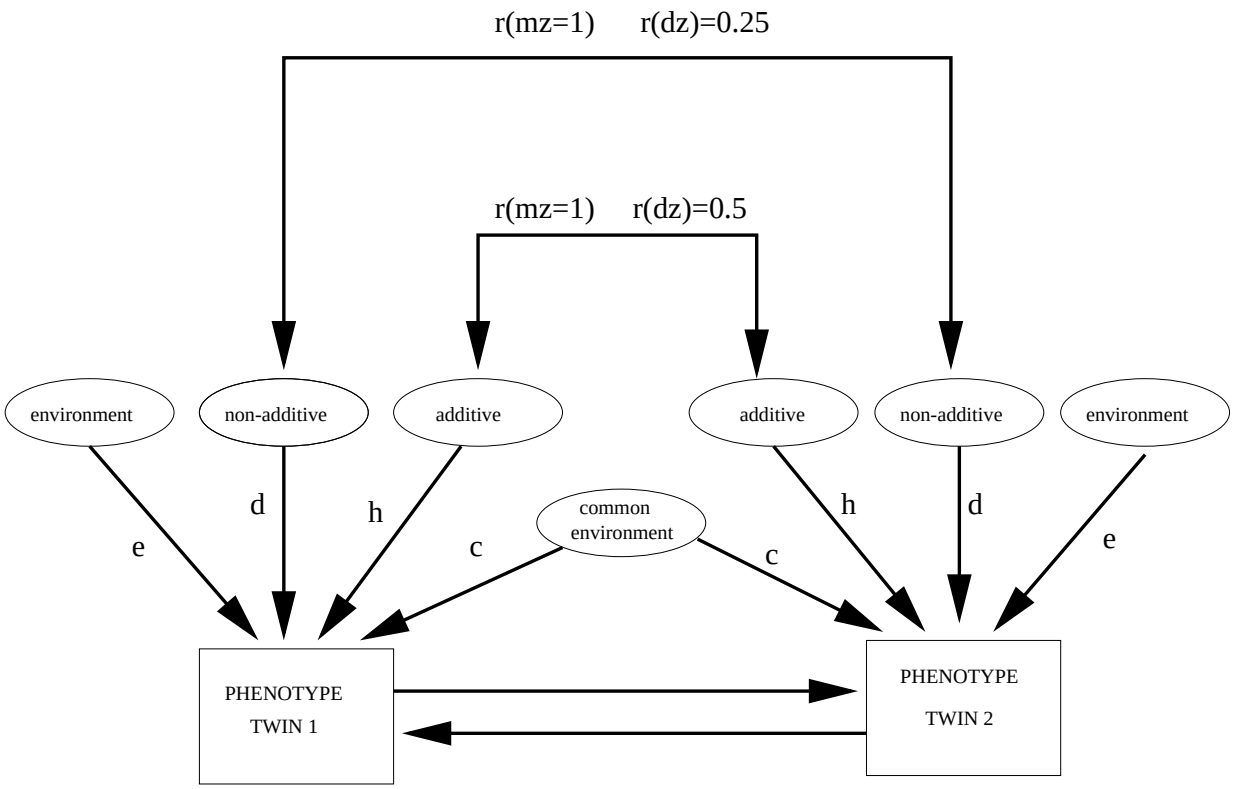
- Twins: Monozygotic (MZ) versus dizygotic (DZ) twins,
- Nuclear families: Father, mother and offsprings,
- Extended families: First-order, second-order and higher order relationships.

Needs correct modeling, whether scientific interest focuses on:

- Fixed effects, as in an association analysis,
- Covariances or correlations.

Analysis is complicated by the variety of types of trait commonly encountered in genetic research:

- Normally distributed continuous traits,
- Binary and ordinal traits,
- Time-to-event endpoints.


r(mz=1) r(dz)=0.25
r(mz=1) r(dz)=0.5
environment
non-additive
additive
additive
non-additive
environment
e
d
h
c
common environment
c
h
d
e
PHENOTYPE TWIN 1
PHENOTYPE TWIN 2

To model covariance structure associated with multivariate Normal traits in nuclear families:

- Variance component models (Fisher 1918; Hopper, 1993): ML, REML, GLS, WLS;
- Model fitting (Mather & Jinks 82; Lange et al 76; Neale & Cardon 92, Burton 95, Gauderman et al 95): ANOVA, EM algorithm, path analysis, multi-level modeling, Bayesian methods;
- GEE methods (Liang & Zeger 86, Qaqish & Liang 92): to handle fixed effects and covariance structure;
- Regressive models (Bonney, 1984): generates a Markov structure reflecting the serial dependence of family members.

A conventional mixed linear model

$$Q_{ij} = \beta' \mathbf{z} + A_{ij} + C_{ij} + C_{S_{ij}} + E_{ij} \qquad (1)$$

Q_{ij} is the observed value of a normally distributed continuous trait for the j^{th} individual in the i^{th} nuclear family;

z_{ij} is a vector of observed covariates representing fixed effects, and β is a corresponding vector of unknown fixed regression coefficients;

A_{ij}, C_{ij}, and $C_{S_{ij}}$ are random effects representing 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_{Cs}, and residual environmental effects.

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

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

$$V_{Cs} = \begin{array}{c} \\ F \\ M \\ S_1 \\ S_2 \end{array} \begin{array}{c} \begin{array}{cccc} F & M & S_1 & S_2 \end{array} \\ \begin{pmatrix} \sigma_{Cs}^2 & 0 & 0 & 0 \\ 0 & \sigma_{Cs}^2 & 0 & 0 \\ 0 & 0 & \sigma_{Cs}^2 & \sigma_C^2 s \\ 0 & 0 & \sigma_{Cs}^2 & \sigma_C^2 s \end{pmatrix} \end{array}$$

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

Menopause Study (DO et al(1998) *Human Biology*)

- Age at menopause (AaM) in > 6000 twin individuals. AaM was defined as age at last menstrual period, determined retrospectively after a woman had stopped menstruating for 12 months.
- Potential covariates:
  - Year of birth, zygosity (DZ or MZ), determined via self-report (95
  - Reproductive: Age at menarche, Age at first full-term pregnancy, parity;
  - Lifestyle: cumulative quantity of alcohol and smoking;
  - Socioeconomic: Education, income, occupation, social class.

Objectives:

- Covariate effects on AaM, familial aggregation of AaM,
- Quantify the variance components.

Melanoma Study (Aitken et al 1996, DO et al 2000, submitted $JRSSC$)

- Queensland Familial Melanoma Project: ascertained 12,016 first incident cases of cutaneous melanoma giagnosed 1982-1990 and reported to Qld Cancer Registry.
- Analysed 1.912 separate families with 15,989 relatives.
- Potential covariates (demographic and melanoma risk factors:
  - Gender, birth year, place of birth.
  - Ability to tan (very brown, moderate tan, slight tan, no tan)
  - Propensity to burn, number of sunburns,
  - Skin color (olive/dark, medium, fair, pale),
  - Hair color (black, brown, blonde, red),
  - Eye color (brown, green/hazel, blue/grey),
  - Total freckling in summer (0, 1-100, > 100),
  - Naevi (none, few, moderate number, very many),
  - Numerous measures of cumulative lifetime exposures.

Objectives: Investigate the relative contributions of genetic and environmental effects on the age at onset of melanoma, and on the expression of naevi and freckles.

Preliminary Analysis: Cox PH and RPART (Therneau & Atkinson 97)

Cox PH results:

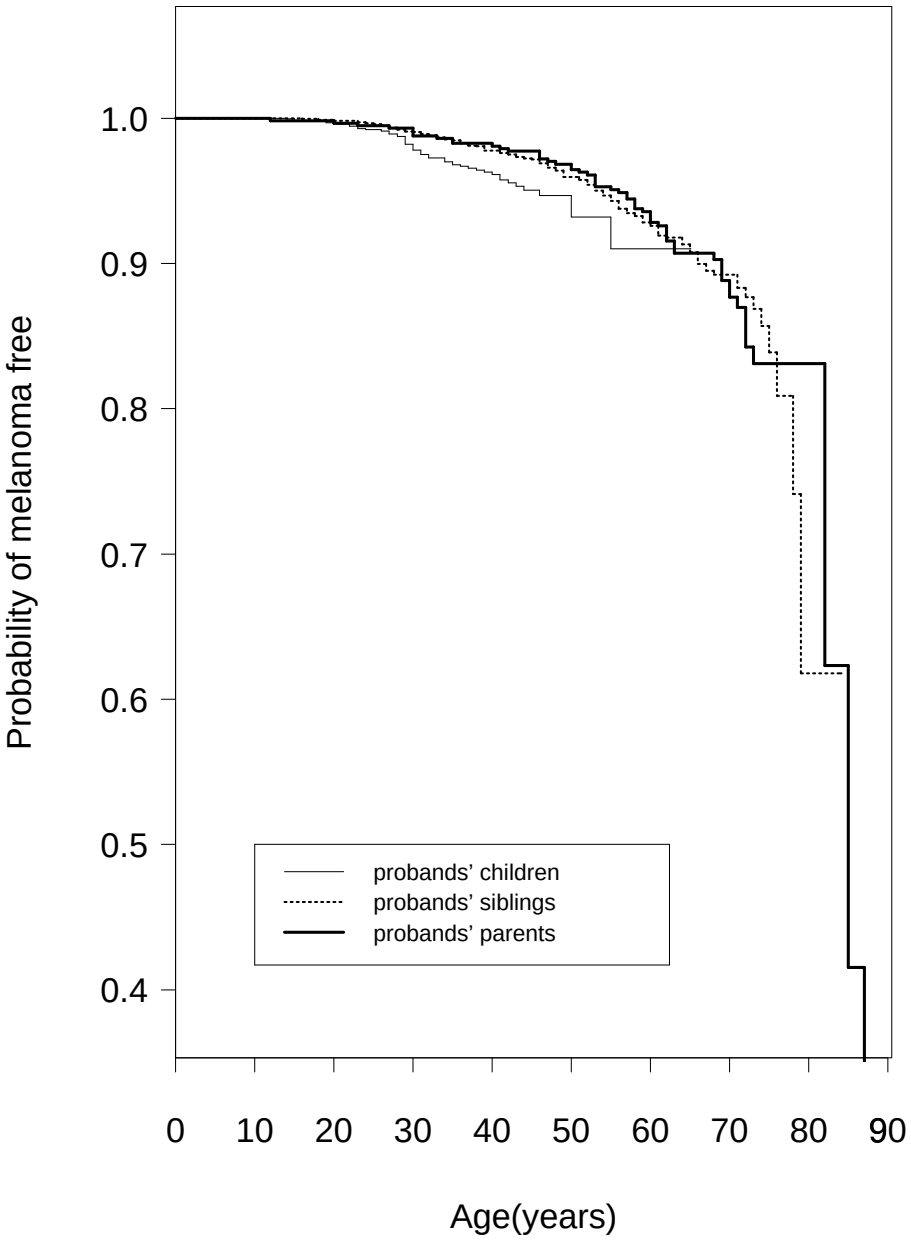
> *An increase of one year in birth year induces 17% increase in risk of earlier melanoma onset.*
>
> *People with no freckles nor naevi have the lowest risk of melanoma onset.*
>
> *The risk of earlier melanoma onset is increased by up to 36% for blue eyed people and even further (46%) for green eyed 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.*

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 β, their standard errors se(β), the relative risk e^{β} and the p-value for each estimate.

Variable	β	se(β)	e^{β}	p-value
Birth Year	0.16	0.00	1.17	< 0.05
Eye Colour (Baseline: **Brown**)				
- Blue/Grey	0.31	0.07	1.36	< 0.05
- Green/Hazel	0.38	0.07	1.46	< 0.05
Hair Colour (Baseline: **Black**)				
- 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: **never burn**)				
- 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: **none**)				
- 1 to 100	0.17	0.06	1.18	< 0.05
- > 100	0.09	0.08	1.10	0.23
Mole Count (Baseline: **none**)				
- 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: **none**)				
- 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


Probability of melanoma free
1.0
0.9
0.8
0.7
0.6
0.5
0.4
probands' children
probands' siblings
probands' parents
0
10
20
30
40
50
60
70
80
90
Age(years)

Preliminary analysis for Menopause Data

- Mx was used to identify appropriate genetic model: ACE or ADE?

  ACE since $r_{MZ} = 0.49, r_{DZ} = 0.33$

- Cox regression was used to screen covariates for inclusion into the model: Birthyear and smoking were highly significantly associated with later AaM, parity, menarche, university education were just significant.

To accommodate age-at-onset outcomes:

1. Estimating equations: Hsu ($PhD\ thesis$ 94), Hsu & Zhao ($Amer\ J\ Hum\ Genet$ 96), Hsu & Prentice ($Biometrika$ 96).
2. The frailty (random effects) model: Clayton ($Biometrika$ 78), Clayton & Cuzick $JRSSA$ 85), Self & Prentice ($Proc\ SIAM$ 86), Oakes ($JASA$ 89), Nielsen $et\ al$ ($Scnd\ J\ Stat$ 92), Gauderman & Thomas ($Genetic\ Epi$ 94), Meyer $et\ al$ ($Amer\ J\ Hum\ Genet$ 91), McGilchrist & Yau ($Aust\ J\ Stat$ 96), Bandeen-Roche & Liang ($Biometrika$ 96), Yashin,Vaupel & Iachine, Yashin & Iachine ($Mechanisms\ of\ Aging\ and\ Development$ 95), Petersen, Andersen & Gill ($Stat\ Neerlandica$ 96), Hougaard ($Lifetime\ Data\ Analysis$ 95).

Specific issues to consider:

- Plausibility of the model,
- Ease of numerical implementation,
- How well can it answer our fundamental questions.

Frailty approach

- Introduce unobserved frailty (liability) for each family.
- Conditioning on frailty, each AaO outcome follows Cox PH model independently.
- Frailty follows some parametric dist, often chosen for mathematical convenience.

PROS

- Frailty model can be used to characterize major gene in segregation analysis of correlated AaO outcomes (Gauderman & Thomas 94).
- Can be used for modeling genetic and environmental effects under restricted assumptions (Meyer $et\ al$ 91).

CONS

- Computation requires double iterative procedure $\rightarrow$ may not be feasible for large pedigrees.
- Restricted assumptions to model frailty.

Estimating Equations approach

- For correlated outcome: extension of Liang & Zeger 86, Prentice 88, Zhao & Prentice 90.
- In modeling correlated phenotypes:
  - Model the dependence of a phenotype on covariates via a marginal regression model.
  - Model the nature of correlation of phenotypes between family members via correlation coefficients.

PROS

- Assume only existence of means and correlations.
- Stat inference is robust, only need pairwise distribution assumptions.
- Manageable computing requirements.

CONS

- No known way to decompose total variation into genetic and environmental components.

Estimating equations

$$\sum_k D_k^{'} \, W_k^{-1} \, F_k = 0.$$

D : Matrix of derivatives of mean and correlation regression models wrt unknown parameters,

W : weights, can be arbitrarily chosen without affecting estimation consistency,

F : deviations of observed phenotypes and correlation functions from their expectations under the assumed mean and correlation regression models.

Cox PH model for twins

$T =$ failure time,

$\delta = 1$ (menopausal), 0 (censored),

$X =$ observed age at menopause or censored age,

$Z^{'} = (Z_1, \ldots, Z_p) =$ vector of covariates,

$K =$ Number of independent pairs of twins or families.

Data for the i^{th} individual $(i = 1, 2)$ from the k^{th} pair is $(X_{ki}, \delta_{ki}, Z_{ki})$.

Model the hazard rate function:

$$\lambda_{ki}(t) = \lambda_0(t) exp(\beta^{'} Z_{ki})$$

where λ_0 is the baseline hazard function describing hazard rates in the absence of covariates.

Obtain consistent and efficient estimators of β by maximizing the *partial likelihood*

$$L = \prod_{i \in events} \frac{exp(\beta^{'} Z_i)}{\sum_{j \in R(t)} exp(\beta^{'} Z_j)}$$

EE for parameters in Cox model

Probability ki^{th} individual fails at time X_{ki} given covariates is:

$$Pr(\delta_{ki} = 1 | X_{ki}, Z_{ki}) = \lambda_{ki}(X_{ki}).$$

$$(I) \quad EE: \quad U_1 = \sum_{families} Z_k{}' \{\delta_k - \hat{\Lambda}_k(X_k, \beta)\} = 0.$$

where $\delta_k{}' = (\delta_{k1}, \delta_{k2})$, $\hat{\Lambda}_k =$ est cum hazard for k^{th} family.

Solve p equations for p covariates.

Var$(\hat{\beta})$ can accommodate correlation in the family.

Equation (I) is robust to departure of misspecification of joint distribution among the individuals within a family.

Modeling familial aggregation of AaO

T_1, T_2 random survival times.

Marginal survival function: $S_i(t_i) = P(T_i \geq t_i)$

Joint survival function: $S(t_1, t_2) = P(T_1 \geq t_1, T_2 \geq t_2)$

General odds ratio:

$$\theta^*(t) = \frac{f(t_1|T_2 = t_2)/S(t_1|T_2 = t_2)}{f(t_1|T_2 > t_2)/S(t_1|T_2 > t_2)} = \frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 > t_2)}$$

Clayton model: Constant odds ratio for each relation type.

$$Pr(T_1 \geq t_1, T_2 \geq t_2) = \{Pr(T_1 \geq t_1)^{-\theta} + Pr(T_2 \geq t_2)^{-\theta} - 1\}^{-1/(\theta)}$$

$1 + \theta \in (0, \infty)$ is the odds ratio.

$1 + \theta = 1$ at grid point $(t_i, t_j) \iff$ Independence between two AaO outcomes at that grid point.

$1 + \theta \to 1$ gives maximal positive dependence.

Simplicity but no general estimation procedure for estimating odds ratio that varies over time domain.

EE for parameters in the odds-ratio function

$$(II) \quad EE: \quad U_2 = \sum_{families} \{a_{k12}(X_{k1}, X_{k2}) - \alpha_{k12}(X_{k1}, X_{k2}; \theta)\}$$

where $a_{k12} = [\delta_{k1} - \Lambda_{k1}(X_{k1})]\,[\delta_{k2} - \Lambda_{k2}(X_{k2})] =$ sample covariance between twin 1 and twin 2 in k^{th} pair = product of differences between disease status and expected mean.

α_{k12} is the expected covariance function conditional on (X_{k1}, X_{k2}) based on Clayton's model.

Estimation procedure

Solve EE1 & EE2 simultaneously for (β, θ).

NO EXPLICIT SOLUTION $\Rightarrow$ Newton-Raphson iterative procedure.

$$\begin{pmatrix} \beta_1 \\ \theta_1 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \theta_0 \end{pmatrix} + \Sigma_1^{-1} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$$

Hsu (1994 PhD thesis) showed consistency & asymptotic normality of estimates with robust covariance matrix that can be empirically estimated from data.

$$\Sigma_1^{-1} \begin{pmatrix} \Sigma^{(11)} & \Sigma^{(12)} \\ \Sigma^{(21)} & \Sigma^{(22)} \end{pmatrix} \Sigma_1^{-1},$$

where $\Sigma^{(11)}$ is the information on the marginal parameters β from the mean function, $\Sigma^{(22)}$ is the information on the dependence parameter θ from the covariance function, and $\Sigma^{(21)}$ is the information of the marginal parameter β from the covariance function. See papers by Hsu and co-authors for explicit mathematical formulation.

A Bayesian analysis using BUGS

The Bayesian paradigm

Let y be the observed data, and θ be everything not observed including parameters and latent variables. The problem, in general terms, is to obtain the expected value of a function of interest $s(.)$ under the posterior density $p(\theta|x)$

$$E[s(\theta)] = \frac{\int_\Theta s(\theta)p(\theta)p(x|\theta)d\theta}{\int_\Theta p(\theta)p(x|\theta)d\theta},$$

which cannot generally be found analytically. One method to carry out the integration on the RHS is to perform simulation of exact Bayesian posterior distributions using Markov chain Monte Carlo techniques such as Gibbs sampling.

MCMC ALGORITHM

1. *Step 1:* Setting initial values for unobserved quantities (parematers and latent variables),
2. *Step2:* For each parameter or latent variable θ_j, sample from its "full conditional distribution" given the current values of all other quantities in the model,
3. *Step 3:* Examine sampled values of parameters and latent variables to monitor convergence and to provide summary measures.

The Model

A Weibull distribution may be used to model time to failure as

$$f(t_i, z_i) = e^{\beta' z_i} t_i^{r-1} \exp(-e^{\beta' z_i} t_i^r), \qquad (2)$$

where β is a vector of unknown regression coefficients. This leads to a baseline hazard of the form

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

Reparameterise by letting $\mu_i = e^{\beta' z_i}$, the conditional distribution of t_i given μ_i is Weibull(r, μ_i).

Twin menopause data

Aim to model the correlation structure within twin pairs to satisfy

$$var(MZ) = var(DZ) = \sigma_T^2 = \sigma_A^2 + \sigma_C^2 + \sigma_E^2$$

$$r_{MZ} = \sigma_A^2 + \sigma_C^2$$

$$r_{DZ} = \frac{1}{2}\sigma_A^2 + \sigma_C^2$$

Mixed model

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

where

$$\log\mu_{ij} \quad = \quad \begin{cases} \alpha + \beta' z + m_i & \text{for MZ twin} \\ \alpha + \beta' z + d_i + d_{ij} & \text{for DZ twin} \end{cases} \tag{3}$$

and m_i, d_i, d_{ij} are independent additive random effects where $m_i \sim N(0, \sigma_A^2 + \sigma_C^2)$,
$d_i \sim N(0, \frac{1}{2}\sigma_A^2 + \sigma_C^2)$, and $d_{ij} \sim N(0, \frac{1}{2}\sigma_A^2)$.

Family melanoma data

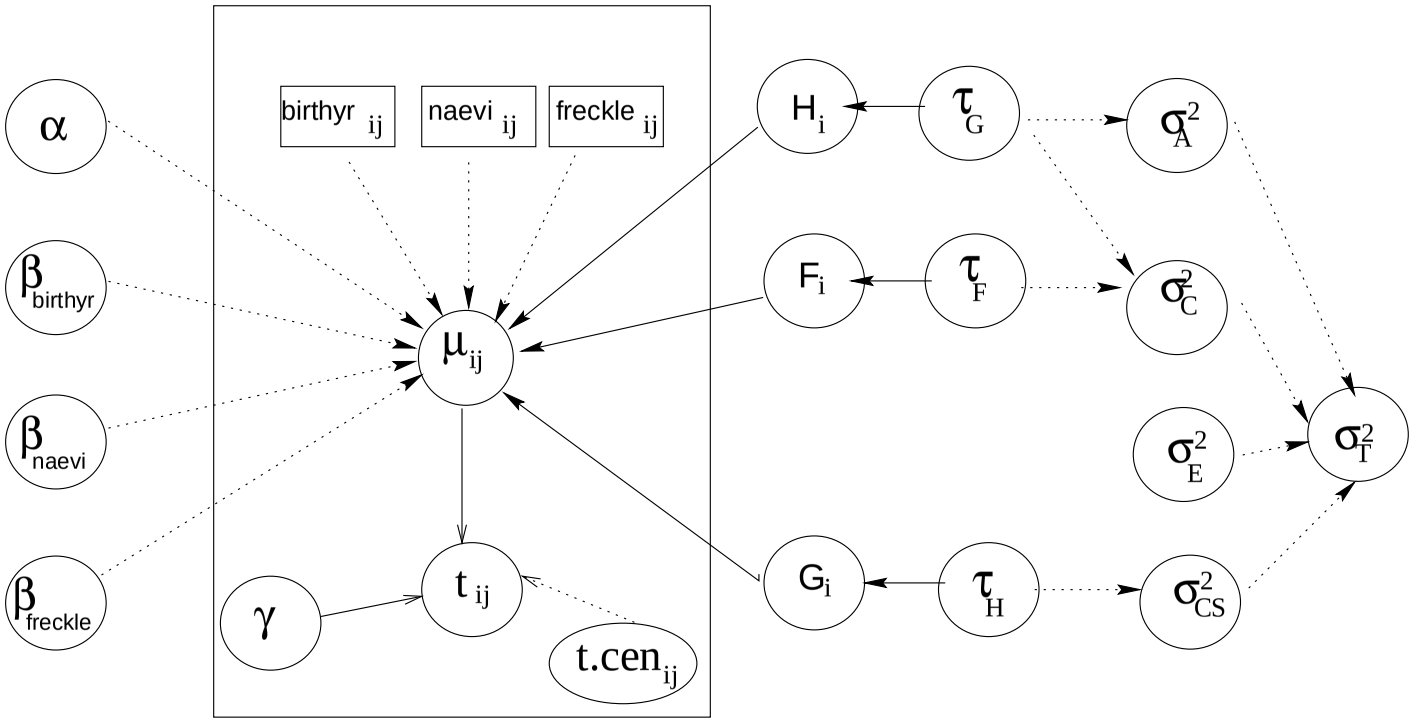
$$t_{ij}|\mu_{ij} \sim Weibull(\gamma, \mu_{ij}) \quad i = 1, \ldots, n;\ j = 1, 2$$

where

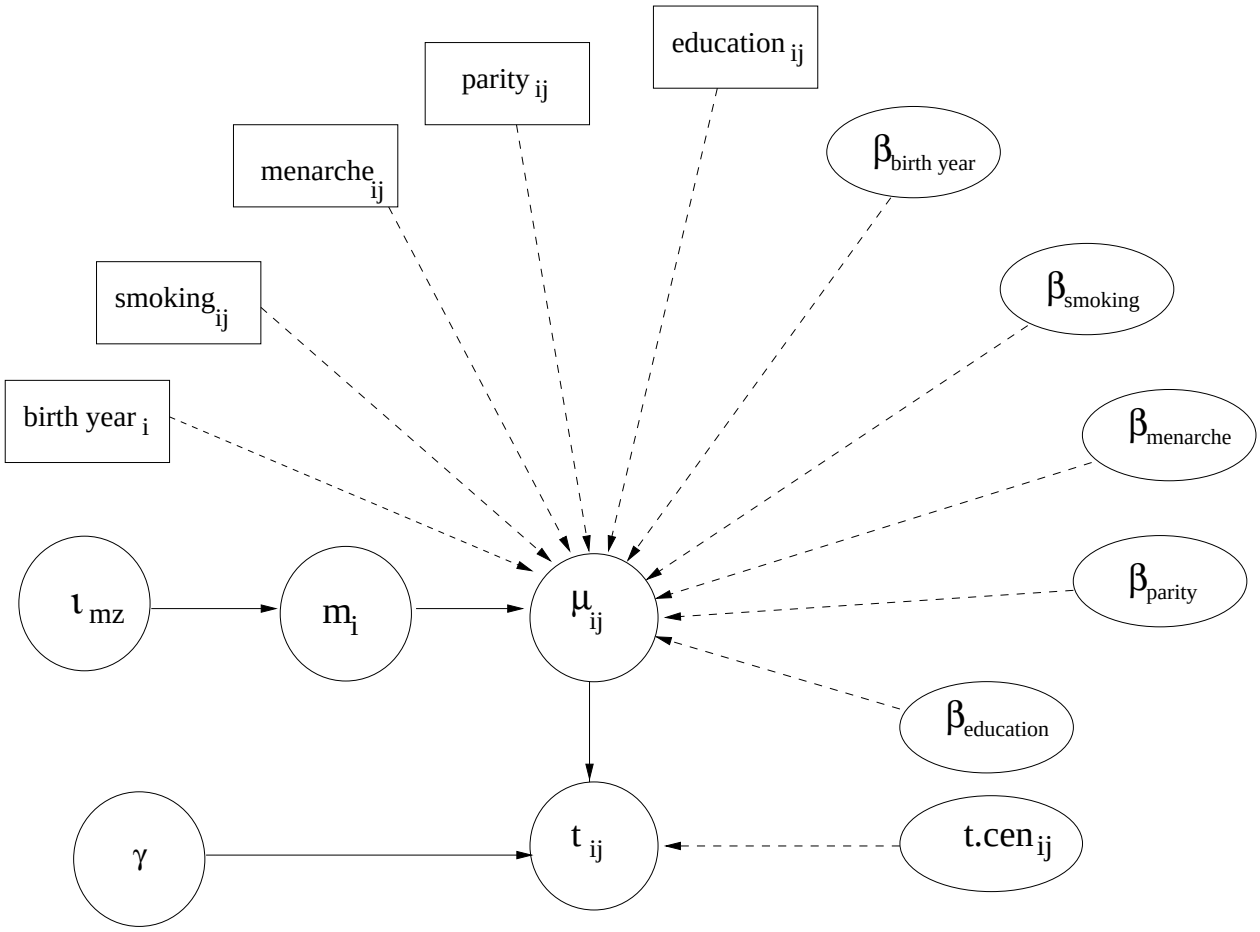
$$\log\mu_{ij} \quad = \quad \begin{cases} \alpha + \beta'\mathbf{z} + F_i + G_i + R_{ij}^P & \text{for fathers} \\ \alpha + \beta'\mathbf{z} + F_i - G_i + R_{ij}^P & \text{for mothers} \\ \alpha + \beta'\mathbf{z} + F_i + H_i + R_{ij}^C & \text{for children} \end{cases}$$

where $F_i \sim N(0, \frac{1}{2}\sigma_A^2 + \sigma_C^2)$, $G_i \sim N(0, \frac{1}{2}\sigma_A^2)$, $H_i \sim N(0, \sigma_{Cs}^2)$.

The regression coefficients and the precision of the random effects (τ_G, τ_F, τ_H) may be given "non-informative" Normal and Gamma priors respectively. The shape parameter, γ, of the time to onset of melanoma distribution may be given a non-informative Gamma prior which slowly decreases on the positive real line.



Individual j in family i



Graphical model of covariate and random family effects for an MZ twin. t_{ij} represents the observed failure time for the j^{th} twin in the i^{th} pair with $t.cen_{ij}$ being an indicator variable of censoring status. Full arrows indicate stochastic links to which a probability is attached; broken arrows denote deterministic relationships; β's are regression coefficients, τ is the precision of the prior distribution and equals the inverse of the variance; m_i is an independent additive random effect modeled as $m_i|t \sim N(0, \sigma_A^2 + \sigma_C^2)$. Rectangles represent actual data values for the covariates; γ and μ_{ij} are scale and shape parameters for the underlying Weibull distribution.

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** β	**Robust se**(β)	**Z-statistic**
	A. Mean effects			
Year of birth	1.142	0.132	0.051	2.588 **
Naevi	1.765	0.568	0.073	7. 781 **
Freckling	1.160	0.148	0.049	3.020 **
	B. Patterns of familial aggregation			
Relationship		$1 + \theta$	**se**(θ)	**Z-Statisti c**
Sib-sib		2.973	0.6217	3.17 **
Parent-child		1.650	0.434	1.50 **
Second/Others		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^{β}	**Coefficient** β	**Robust se(**β**)**	**95% CI of** *beta*
	Weibull Model: A. Mean effects - Response variable is Age-at-onset			
Year of birth	1.378	0.321	0.0027	(0.316,0.326) **
Naevi	1.126	0.119	0.0021	(0.058,0.185) **
Freckling	1.017	0.017	0.1400	(-0.055,0.085)
	Weibull Model: B: Variance components - Response variable is Age-at-onset			
Latent	**Mean from 5000 iterations**		**se(**σ^2**)**	**9 5% CI of** σ^2
σ^2_A		0.452	0.054	(0.348,0.566) **
σ^2_C		-0.223	0.027	(-0.282,-0.169)
σ^2_{Cs}		0.467	0.040	(0.393,0.545) **
γ		14.3	0.104	(14.2, 14.6)
	Binomial Model: Variance components - Response variable is Naevi			
Latent	**Mean from 5000 iterations**		**se(**σ^2**)**	**9 5% CI of** σ^2
σ^2_A		0.142	0.149	(0.002,0.498) **
σ^2_C		0.704	0.156	(0.403,1.010) **
σ^2_{Cs}		0.195	0.157	(0.0025,0.553) **
	Binomial Model: Variance components - Response variable is Freckling			
Latent	**Mean from 5000 iterations**		**se(**σ^2**)**	**9 5% CI of** σ^2
σ^2_A		2.050	0.779	(0.835,3.570) **
σ^2_C		2.600	0.418	(1.780,3.460)**
σ^2_{Cs}		0.115	0.088	(0.011,0.312)**

Table 5: Estimated Regression Coefficients in the Proportional Hazard Model and Estimated Odds Ratios for Quantifying Familial Aggregation on Menopause Data From a Longitudinal Study of Australian Twins Using the Estimating Equations Approach. ** indicates significance.

Covariate	**RR** $= e^{\beta}$	**Coefficient** β	**Robust se**(β)	**Z-statistic**
	A. Mean effects			
Year of birth	0.978	-0.022	0.0038	-5.79 **
Smoking	1.123	0.116	0.0680	1.72
Uni education	0.643	-0.442	0.1077	-4.10 **
Menarche	0.984	-0.016	0.0226	0.71
Parity	0.624	-0.471	0.0960	-4.90 **
	B. Patterns of familial aggregation			
Zygosity		$1 + \theta$	**se**(θ)	**Z-Statistic**
MZ		1.764	0.1480	5.16 **
DZ		1.355	0.1270	2.79 **

Table 6: Estimated Regression Coefficients in the Weibull Model and Estimated Variance Components on Menopause Data From a Longitudinal Study of Australian Twins using the Gibbs Sampling Approach. ** indicates significance

Covariate	**RR =** e^{β}	**Coefficient** β	**Robust se(**β**)**	**95% CI of** β
	A. Mean effects			
Year of birth	0.971	-0.029	0.0035	(-0.036,-0.023) **
Smoking	1.148	0.138	0.0788	(-0.187,0.293)
Uni education	0.672	-0.397	0.1400	(-0.676,-0.123) **
Menarche	0.976	-0.024	0.0204	(-0.063,0.015)
Parity	0.556	-0.586	0.1260	(-0.830,-0.033) **
	Variance components			
Latent	**Mean from 5000 iterations**		**se(**σ^2**)**	**95% CI of** σ^2
σ^2_A	0.730		0.329	(0.129,1.410)
σ^2_C	0.011		0.240	(-0.456,0.489)

CONCLUSIONS - MELANOMA STUDY

GEE results:

- GEE results: Later birth year, having at least a moderate number of naevi and freckles are associated with later age at onset of melanoma.
- OR $= 2.973$ (sib-sib), 1.650 (parent-child), 1.155 (higher order) indicate a dominance variance in addition to genetic additive variance.

MCMC results for random effects:

- Additive genetics and shared sibling environment seem to impact equally on the variation of the age at onset of melanoma; a negative estimate for σ_C^2 suggests that the sibling correlation is much larger than the parent-sibling correlation. Should try another model, a dominance model or even a purely environmental model.
- Common family environment effect contributes the most to the expression of naevi relative to addititve and shared sibling effects.
- The variation in the expression of freckles is largely explained by additive genetic and shared family effects.

GEE versus Bayesian approaches

- GEE is a marginal approach producing regression coefficient estimates that describe the average population response to changing covariates; MCMC produces subject-specific coefficients.
- GEE describes a common covariance among specific relative pairs; whereas the Bayesian approach can explicitly describe the source of this covariance.
- Bayesian method has the flexibility in incorporating prior info for the covariates or latent effects.

CONCLUSIONS

- Based on 267 MZ and 159 DZ post-menopausal twin pairs: $r_{MZ} = 0.49, r_{DZ} = 0.33$.
- Age of menopause increased with later birthyear, university education, and parity of two or more.
- Smoking may be associated with earlier menopause.
- Late age of menarche may be associated with earlier menopause.
- Additive genetics effect: 31% to 53% of the variance.
- Unique environmental effect: 47% to 69% of the variance.
- Bayesian analysis: give exact decomposition of variance components but is extremely computer-intensive, 30 hours CPU time (on Ultra-Sparc) to obtain convergent results incomparison to 20 seconds of CPU time for the estimating equations approach.
- Future work:
  - (i) Assess MCMC performance under model misspecification of the hazard function and/or the genetic model;
  - (ii) Include mother/twin relationships and other sibships;
  - (iii) Consider random covariate effects, rather than fixed.

Other methods

Mixed-effects frailty model - McGilchrist & Yau (96): Proposed three approaches for failure time analysis which involved modeling frailty and other covariates as mixed effects.

Imputation method - Yashin & Iachine (95): Replace censored failure times by imputed uncensored failure times using the correlated frailty model suggested. The approach exploits an important advantage of survival model with frailty: additive decomposition of frailty on genetic and environmental components induces the competing risks structure of the respective failure-time model.

Pickles' method (94): A threshhold is presumed in the liability distribution, beyond which individuals are affected. For twin data, a multifactorial threshold model can be fitted by maximum likelihood methods to contingency tables that cross-classify the menopausal status of Twin 1 with Twin 2. Twin liability correlations may be parameterized in terms of the different variance components (A, C, D, or E) using Mx. To handle censoring, group twins into a finite number of age groups, within each age group construct a contingency table of age-at-onset for Twin 1 versus Twin 2, those with censored observations are placed in a category of their own, a category on the underlying latent scale that extended from the last estimated threshold to infinity. Analysis then proceeds in the usual fashion as if there were no censored data. This method is simple only when the censoring mechanism is simple.

FUTURE WORK

- Application to glioma and childhood sarcoma data sets collected at MDACC,
- Extend methodology applicable for extended families,
- Incorporate biomarker info and extend models to have the ability of decomposing variance into major gen and polygene effects,
- Extend the methodology to include Bayesian model averaging and model search strategies applicable for finding the best fitting genetic model.