

Incorporating Heterogeneity in Detecting Non-Inherited Maternal Antigen (NIMA) Coding Alleles as a Disease Risk Factor

Jingyuan Yang

Department of Statistics
The Ohio State University

May 29th 2007

Outline

Outline

Background

- ▶ Microchimerism: Persistence of maternal cells in the offspring.

Background

- ▶ Microchimerism: Persistence of maternal cells in the offspring.
- ▶ Non-Inherited Maternal Antigen (NIMA): Antigens in the offspring, which are not encoded by the inherited alleles but by the mother's alleles, and passed to the offspring through microchimerism.

Background

- ▶ Microchimerism: Persistence of maternal cells in the offspring.
- ▶ Non-Inherited Maternal Antigen (NIMA): Antigens in the offspring, which are not encoded by the inherited alleles but by the mother's alleles, and passed to the offspring through microchimerism.
- ▶ NIMA happens when there is a maternal-fetal genotype incompatibility, i.e., the mother carries the antigen alleles but the offspring doesn't.

Background

- ▶ Microchimerism: Persistence of maternal cells in the offspring.
- ▶ Non-Inherited Maternal Antigen (NIMA): Antigens in the offspring, which are not encoded by the inherited alleles but by the mother's alleles, and passed to the offspring through microchimerism.
- ▶ NIMA happens when there is a maternal-fetal genotype incompatibility, i.e., the mother carries the antigen alleles but the offspring doesn't.
- ▶ It has been suggested that NIMA may be involved in the pathogenesis of certain autoimmune diseases, such as scleroderma, thyroiditis and rheumatoid arthritis (RA).

NIMA and RA

- ▶ RA is highly associated with HLA-DRB1 *0101, *0102, *0401, *0404, *0405, *0408, *10, *1001, *1042 and *1406 alleles that encode a shared epitope (SE).

NIMA and RA

- ▶ RA is highly associated with HLA-DRB1 *0101, *0102, *0401, *0404, *0405, *0408, *10, *1001, *1042 and *1406 alleles that encode a shared epitope (SE).
- ▶ However, as many as 30% of the patients do not carry SE coding alleles.

NIMA and RA

- ▶ RA is highly associated with HLA-DRB1 *0101, *0102, *0401, *0404, *0405, *0408, *10, *1001, *1042 and *1406 alleles that encode a shared epitope (SE).
- ▶ However, as many as 30% of the patients do not carry SE coding alleles.
- ▶ One explanation of the latter finding is that NIMA may be involved in RA pathogenesis through microchimerism, but other studies have not found a significant NIMA effect.

NIMA and RA

- ▶ RA is highly associated with HLA-DRB1 *0101, *0102, *0401, *0404, *0405, *0408, *10, *1001, *1042 and *1406 alleles that encode a shared epitope (SE).
- ▶ However, as many as 30% of the patients do not carry SE coding alleles.
- ▶ One explanation of the latter finding is that NIMA may be involved in RA pathogenesis through microchimerism, but other studies have not found a significant NIMA effect.
- ▶ A possible reason of the failure is that the original MFG test hasn't taken into account the heterogeneity in people's susceptibility to microchimerism.

Outline

Maternal-Fetal Genotype Incompatibility (MFG) Test

The MFG test models the joint distribution of parental and affected offspring genotypes given the status of affected offsprings in nuclear families as $\prod_{i=1}^N \Pr(\mathbf{G}_c^{(i)}, \mathbf{G}_p^{(i)} | \mathbf{D}^{(i)})$

- ▶ $\mathbf{G}_c^{(i)} = (G_{c1}^{(i)}, G_{c2}^{(i)}, \dots, G_{ck_i}^{(i)})$ denote the genotypes of k_i affected offsprings
- ▶ $\mathbf{G}_p^{(i)} = (G_m^{(i)}, G_f^{(i)})$ denote the parents genotypes
- ▶ $\mathbf{D}^{(i)}$ denotes the event that k_i offsprings are affected.

Maternal-Fetal Genotype Incompatibility (MFG) Test

The MFG test models the joint distribution of parental and affected offspring genotypes given the status of affected offsprings in nuclear families as $\prod_{i=1}^N \Pr(\mathbf{G}_c^{(i)}, \mathbf{G}_p^{(i)} | \mathbf{D}^{(i)})$

- ▶ $\mathbf{G}_c^{(i)} = (G_{c1}^{(i)}, G_{c2}^{(i)}, \dots, G_{ck_i}^{(i)})$ denote the genotypes of k_i affected offsprings
- ▶ $\mathbf{G}_p^{(i)} = (G_m^{(i)}, G_f^{(i)})$ denote the parents genotypes
- ▶ $\mathbf{D}^{(i)}$ denotes the event that k_i offsprings are affected.

Under the assumption that siblings' phenotypes are independent given parental genotypes, likelihood is parameterized in terms of penetrance functions and population mating type frequencies.

$$\Pr(\mathbf{G}_c^{(i)}, \mathbf{G}_p^{(i)} | \mathbf{D}^{(i)}) = \prod_{j=1}^{k_i} \frac{P(D_j^{(i)}=1 | G_{c_j}, \mathbf{G}_p^{(i)}) P(G_{c_j} | \mathbf{G}_p^{(i)}) P(\mathbf{G}_p^{(i)})}{P(D_j^{(i)}=1)}$$

MFG Test - Continued

The penetrance is modeled as:

$$P(D|G_c, \mathbf{G}_p) = p \times \mu^{I[M]} \times \rho_1^{I[Z=1]} \times \rho_2^{I[Z=2]}$$

where $I[.]$ is the indicator function. M indicates the offspring doesn't carry the disease allele but the mother has, and $Z = i, i = 1, 2$, indicates the offspring carries i disease alleles.

MFG Test - Continued

The penetrance is modeled as:

$$P(D|G_c, \mathbf{G}_p) = p \times \mu^{I[M]} \times \rho_1^{I[Z=1]} \times \rho_2^{I[Z=2]}$$

where $I[.]$ is the indicator function. M indicates the offspring doesn't carry the disease allele but the mother has, and $Z = i$, $i = 1, 2$, indicates the offspring carries i disease alleles.

- ▶ The parameter p denotes the population baseline disease incidence rate.
- ▶ The parameter μ is the relative risk due to NIMA.
- ▶ ρ_1 and ρ_2 are the relative risks when the offspring carries one or two SE coding allele, respectively, relative to zero copy.

MFG Test - Continued

The penetrance is modeled as:

$$P(D|G_c, \mathbf{G}_p) = p \times \mu^{I[M]} \times \rho_1^{I[Z=1]} \times \rho_2^{I[Z=2]}$$

where $I[\cdot]$ is the indicator function. M indicates the offspring doesn't carry the disease allele but the mother has, and $Z = i$, $i = 1, 2$, indicates the offspring carries i disease alleles.

- ▶ The parameter p denotes the population baseline disease incidence rate.
- ▶ The parameter μ is the relative risk due to NIMA.
- ▶ ρ_1 and ρ_2 are the relative risks when the offspring carries one or two SE coding allele, respectively, relative to zero copy.

Then by testing $H : \mu = 1$ vs. $K : \mu \neq 1$, one can test the effect of NIMA.

MFG - Continued

- ▶ The MFG test doesn't take into account the population heterogeneity of the susceptibility to microchimerism.
- ▶ Only the affected offsprings' phenotypes are used in the test.
- ▶ Potential misinterpretation of μ :

μ is interpreted as the relative risk due to NIMA, which is relative to zero inherited SE encoding allele, so

$$\mu = \frac{P(D=1|G_m=1, G_c=0)}{P(D=1|G_c=0)} = \frac{P(D=1|G_m=1, G_c=0)}{P(D=1|G_m=1, G_c=0)P(G_m=1|G_c=0) + P(D=1|G_m=0, G_c=0)P(G_m=0|G_c=0)}$$

Outline

Model

We model both affected and unaffected offsprings in the likelihood:

$$\begin{aligned} & L(\alpha, \beta | \mathbf{D}, \mathbf{G}_m, \mathbf{G}_c) \\ &= \prod_{i=1}^N P(G_m^{(i)}, \mathbf{G}_c^{(i)}, \mathbf{D}^{(i)} | \alpha, \beta) \\ &= \prod_{i=1}^N P(\mathbf{D}^{(i)} | G_m^{(i)}, \mathbf{G}_c^{(i)}, \alpha, \beta) P(\mathbf{G}_c^{(i)} | G_m^{(i)}) P(G_m^{(i)}) \\ &= \prod_{i=1}^N P(\mathbf{D}^{(i)} | G_m^{(i)}, \mathbf{G}_c^{(i)}, \alpha, \beta) \prod_{i=1}^N P(\mathbf{G}_c^{(i)} | G_m^{(i)}) P(G_m^{(i)}) \\ &\propto \prod_{i=1}^N \prod_{j=1}^{n_i} P(D_j^{(i)} | M^{(i)} = s, Z_j^{(i)} = t, \alpha, \beta) \end{aligned}$$

Model - Continued

There are 7 possible cases of $M=s$ and $Z=t$
 $(s, t) \in \{(0, 0), (0, 1), (1, 1), (2, 1), (1, 2), (2, 2), (1, 0)\}$.

Let $I_{ij}^k = I[(M^{(i)}, Z_j^{(i)}) = \text{Case } k]$, $k=1, \dots, 7$, $i=1, \dots, N$, $j=1, \dots, n_i$.

Model the likelihood as:

$L \propto \prod_i \prod_j (P_{ij})^{I[D_{ij}=1]} (1 - P_{ij})^{I[D_{ij}=0]}$, where

$$P_{ij} = \text{logit}^{-1} \left(\sum_{k=1}^6 (\beta_k I_{ij}^k) \right) + \left(\frac{(\alpha) \exp(\beta_7)}{1 + \exp(\beta_7)} + \frac{(1-\alpha) \exp(\beta_8)}{1 + \exp(\beta_8)} \right) I_{ij}^7.$$

Interpretation of Parameters

- ▶ $\beta \in \mathbf{R}^8$
 - ▶ For $k=1$ to 6 , β_k can be interpreted as the log odds of disease when (M, Z) is Case k .
 - ▶ For $k=7,8$, β_7 is the log odds of disease when (M, Z) is Case 7 and microchemirism happens, β_8 is the log odds of disease when (M, Z) is Case 7 and microchemirism doesn't happen.
- ▶ $\alpha \in (0, 1)$
 - ▶ It is the heterogeneity parameter, which is the probability of microchemirism, given the maternal-fetal genotype incompatibility.

Estimate the Parameters

- ▶ Bayesian Estimators
- ▶ Maximized Likelihood Estimators

Outline

Simulation

Generate simulated data:

- ▶ Simulate all genotypes under the assumptions of H-W equilibrium and random mating although they are not necessary in the analysis.
- ▶ Allele frequencies mimic Asian and African population.
- ▶ Simulate the phenotypes (disease status) by using the analysis model.

Fit the Simulated Data

Only tried Bayesian method so far:

- ▶ Non-informative priors of all parameters were used.
- ▶ Distributions of posterior samples of all parameters converge after 5000 updates.
- ▶ When the family size N is fairly large (I tried 500 and 5000), the estimation is not satisfactory, especially for α and β_8 .
- ▶ When an informative prior is assigned to α , the estimation of α is dramatically improved, but β_8 remains bad.