

MOFET: A SAS macro for testing general maternal-fetal interactions

Peter Kraft
16 August 2004

DESCRIPTION

MOFET fits flexible log-linear models for gene-disease association to nuclear family data. It uses SAS PROC NLMIXED to estimate the relative risk of disease due to a child's genotype, her mother's genotype, and/or the interaction between the child's and the mothers genotypes. SAS PROC NLMIXED also returns global model-fit statistics such as the likelihood ratio chi-square and AIC. MOFET can use all available child genotypes. The phenotypes of disease-free children or those with unknown disease status do not contribute to the likelihood. Genotype information can be missing on either of the parents (but not both).

SYNTAX

```
%macro mofet (data=in,  
              d=case,  
              g=geno,  
              momg=momg,  
              dadg=dadg,  
              mt=mt,  
              nfam=nfam,  
              maxnfam=7,  
              parms=offmain mfint = 0,  
              model=offmain*Off + mfint*MomD*Off,  
              nlmopt= );
```

data	Name of the input data set (see "data set format" below)
d	Prefix for child disease indicator variable(s)
g	Prefix for child genotype variable(s) (diallelic genotypes coded 0/1/2; see "model formula" below)
dadg	Paternal genotype (diallelic genotypes coded 0/1/2)
momg	Maternal genotype (diallelic genotypes coded 0/1/2; see "model formula" below)
mt	Mating type, coded as in Table 1
nfam	Number of genotyped children in the family
maxnfam	Maximum nfam over the data set
parms	Parameter initial values passed to PROC NLMIXED PARMs statement
model	Model for log genetic relative risk (see "model formula" below)
nlmopt	Options to be passed to PROC NLMIXED

Table 1. Mating type code, diallelic marker with alleles "D" and "d"

Genotypes		Numeric Code for Genotypes		Mating Type {D/D,d/d}
<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	
d/d	d/d	0	0	0
D/d	d/d	1	0	2
d/d	D/d	0	1	2
D/D	d/d	2	0	4
d/d	D/D	0	2	4
D/d	D/d	1	1	5
D/d	D/D	1	2	6
D/D	D/d	2	1	6
D/D	D/D	2	2	8
	<i>father missing</i>			10
	<i>mother missing</i>			12

DATA SET FORMAT

Each record represents one family. Variables *mt*, *nfam*, are required. If *mt* = 10 (12), *momg* (*dadg*) is also required. Genotypes for multiple children should be recorded in variables *geno1*, *geno2*, ... *genomaxnfam*. Genotypes *genoi* for *i* > *nfam* are ignored. Disease status should be recorded in *case1*, ... *casemaxnfam*. Children with disease should be coded "1"; children without disease or of unknown disease status should be coded "0". *geno* and *case* variables cannot be missing.

Sample data:

<i>momg</i>	<i>dadg</i>	<i>mt</i>	<i>nfam</i>	<i>geno1</i>	<i>geno2</i>	<i>geno3</i>	<i>case1</i>	<i>case2</i>	<i>case3</i>
0	1	2	1	0	.	.	1	.	.
1	1	5	3	0	1	1	0	1	0
.	1	12	2	1	2	.	1	1	.
...									

MODEL FORMAT

MOFET enables flexible modeling of child and maternal genotypes by calculating several genotype scores internally. These genotype scores can be used to form log-linear genetic relative risks. For example, the scores *OffD* and *IntD* represent dominant codings for child and maternal-fetal genotype interaction effects, respectively. The macro-variable assignments

```
parms=offmain mfint = 0,
model= offmain*OffD + mfint*IntD
```

fit a model where heterozygote and homozygote-carrier children of homozygote-non-carrier mothers have a risk of disease $RR_G = \exp(\text{offmain})$ times that of homozygote-non-carrier children of homozygote-non-carrier mothers. Heterozygote and homozygote-

carrier children of carrier mothers have a relative risk of $RR_G RR_I$, where $RR_I = \exp(mfint)$. All model parameters should be initialized to 0 via the parms macro-variable. (NOTE: The log-mating type frequencies are included in the model by default. Setting `parms= , model=0` will estimate these parameters under the null model of no genotype effects.)

Available genotype codings are listed in Table 2.

STATISTICAL DETAILS

For details on the retrospective likelihood that MOFET fits, see [Kraft, et al. 2004].

PERFORMANCE

Table 3 summarizes several small simulation studies under several scenarios. In the first, only the children's genotypes were associated with disease (the relative risk for children carrying one mutant allele was $R_1 = 3$; for those carrying two alleles $R_2 = 6$). In the second, only maternal genotypes contributed additional risk (children with mothers with one mutant allele had a relative risk of $S_1 = 2$; for those with mothers with two copies, $S_2 = 3$). For the third model, children with exactly one copy of the mutant allele were at decreased risk, but only if their mother also had exactly one copy of the mutant allele (i.e. $\alpha_{11} = .5$, all other parameters $\equiv 1$). The final model included offspring and maternal genotype main effects as well as a maternal-fetal interaction effect.

MOFET performed well, and did not exhibit any of the bias seen in a related pseudolikelihood analysis [Cordell 2004] (this is because the likelihood behind MOFET explicitly accounts for correlation among sibs due to their shared mother).

REFERENCES

- Cordell HJ. 2004. Properties of case/pseudocontrol analysis for genetic association studies: Effects of recombination, ascertainment, and multiple affected offspring. *Genet Epidemiol* 26(3):186-205.
- Kraft P, Palmer C, Woodward J, Turunen J, Minassian S, Paunio T, Lonnqvist J, Peltonen L, Sinsheimer J. 2004. RHD Maternal-Fetal Genotype Incompatibility and Schizophrenia: Extending the MFG Test to Include Multiple Siblings and Birth Order. *Eur J Hum Genet* 12:192-198.

Table 2. Genotype codings. "Off" for child's genotype; "Mom" for mother's genotype, "Int" for interaction

Offspring's	Mother's	Off	OffD	OffR	Off1	Off2	Mom	MomD	MomR	Mom1	Mom2	Int	IntD	IntR	Int11	Int12	Int21	Int22
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0
1	1	1	1	0	1	0	1	1	0	1	0	1	1	0	1	0	0	0
2	1	2	1	1	0	1	1	1	0	1	0	2	1	0	0	0	1	0
0	2	0	0	0	0	0	2	1	1	0	1	0	0	0	0	0	0	0
1	2	1	1	0	1	0	2	1	1	0	1	2	1	1	0	1	0	0
2	2	2	1	1	0	1	2	1	1	0	1	4	1	1	0	0	0	1

Table 3. Average bias and 95% coverage using CEPG pseudolikelihood approach to multiple affected siblings and CEPG likelihood explicitly accounting for shared parents (the MOFET likelihood)^a

Number sibs	Parameter	Scenario 1				Scenario 2			
		Truth	Pseudo ^b Bias	CEPG [4] Bias Coverage		Truth	Pseudo ^b Bias	CEPG [4] Bias Coverage	
2	$\ln(R_1)$	$\ln(3)$.11	.00	.97	$\ln(1)$.00	-.01	.95
	$\ln(R_2)$	$\ln(6)$	n/a	.01	.97	$\ln(1)$	n/a	-.01	.95
	$\ln(S_1)$	$\ln(1)$.00	.00	.95	$\ln(2)$.70	.01	.95
	$\ln(S_2)$	$\ln(1)$	n/a	.00	.94	$\ln(3)$	n/a	.03	.94
	$\ln(\alpha_{11})$	$\ln(1)$	-.01	.00	.94	$\ln(1)$.01	.00	.93
3	$\ln(R_1)$	$\ln(3)$.00	.01	.94	$\ln(1)$.00	-.01	.95
	$\ln(R_2)$	$\ln(6)$	n/a	.01	.96	$\ln(1)$	n/a	-.01	.95
	$\ln(S_1)$	$\ln(1)$.00	.01	.94	$\ln(2)$	1.38	.00	.96
	$\ln(S_2)$	$\ln(1)$	n/a	.00	.95	$\ln(3)$	n/a	.01	.95
	$\ln(\alpha_{11})$	$\ln(1)$.00	.00	.93	$\ln(1)$.01	.00	.95
		Scenario 3				Scenario 4			
		Truth	Pseudo ^b Bias	CEPG [4] Bias Coverage		Truth	Pseudo ^{b,c} Bias	CEPG [4] Bias Coverage	
2	$\ln(R_1)$	$\ln(1)$	-.01	.00	.97	$\ln(1.5)$	-.10	.01	.96
	$\ln(R_2)$	$\ln(1)$	n/a	.01	.95	$\ln(2)$	n/a	.00	.95
	$\ln(S_1)$	$\ln(1)$	-.01	-.01	.95	$\ln(1.8)$.79	.00	.95
	$\ln(S_2)$	$\ln(1)$	n/a	.00	.94	$\ln(2)$	n/a	.01	.96
	$\ln(\alpha_{11})$	$\ln(.5)$.09	.00	.95	$\ln(.5)$.25	-.01	.96
3	$\ln(R_1)$	$\ln(1)$.00	.00	.99	$\ln(1.5)$	-.39	.00	.95
	$\ln(R_2)$	$\ln(1)$	n/a	-.05	.97	$\ln(2)$	n/a	-.01	.94
	$\ln(S_1)$	$\ln(1)$.00	.00	.95	$\ln(1.8)$	1.53	.01	.96
	$\ln(S_2)$	$\ln(1)$	n/a	.01	.94	$\ln(2)$	n/a	.01	.97
	$\ln(\alpha_{11})$	$\ln(.5)$.19	.00	.98	$\ln(.5)$.59	-.01	.96

^a based on 500 replicate simulations of studies with 500 (333) nuclear families with exactly two (three) affected siblings. Baseline probability of disease = 5%. Causal-locus frequency 20%. Relative risk parameters are described in the text.

^b Taken from Cordell [2004]. Bias in parameters R_2 and S_2 was not reported.

^c The comparable model from Cordell [2004] (Scenario 5) included a parent-of-origin effect ($I_m = 2.5$).