Analysis of family case-control studies in Stata

David C Muller

Cancer Epidemiology Centre
Cancer Council Victoria

Centre for MEGA Epidemiology
University of Melbourne

17 September 2011

Outline

Cancer Council Victoria

Family case-control studies

Family case-control studies are frequently employed to study both genetic and non-genetic factors.

Advantages:

- Ease of control recruitment
- Greater power when the exposure is rare
- Natural setting for assessing gene-environment interaction

Challenges:

- Within family correlation (observed and unobserved)
- Correcting for ascertainment

CLR for family studies

Family studies are typically analysed as a matched case-control design using conditional logistic regression (CLR). CLR assesses the probability of an observed distribution of cases in a family, conditional on the total number of cases in the family.

- Accounts for the ascertainment of families
- Conditions out the family specific intercept

Limitation of CLR for family studies

What about unobserved within family genetic variation?

Consider the probability distribution of disease status Y_{ij} as a function of covariate X_{ij}, as well as a family specific effect a_i, and an individual level genetic effect g_{ij} for the jth person in the ith family:

$$\text{logit pr}(Y_{ij} = 1 | a_i, g_{ij}, X_{ij}) = \mu + a_i + \beta X_{ij} + g_{ij}$$

Limitation of CLR for family studies

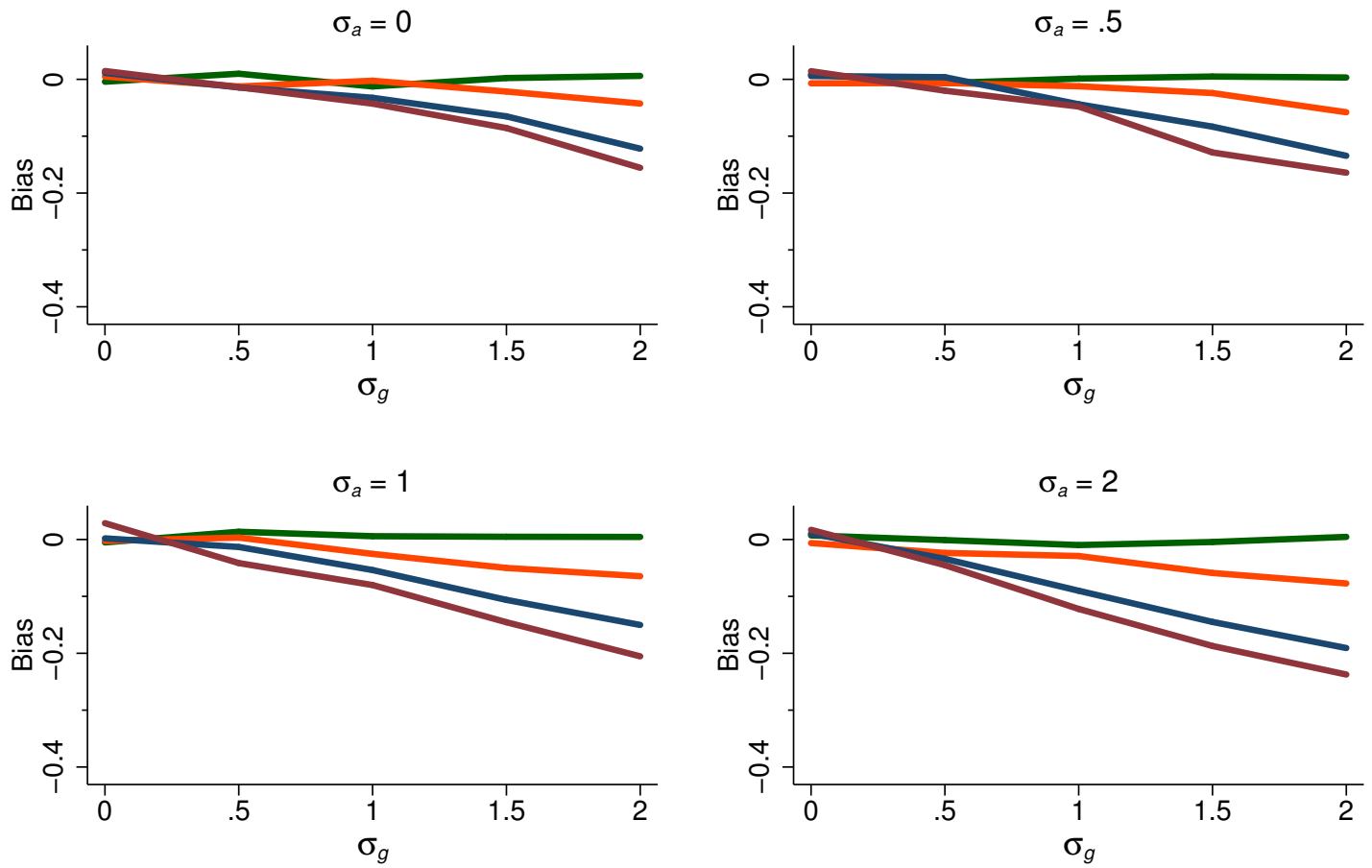
What about unobserved within family genetic variation?

Consider the probability distribution of disease status Y_{ij} as a function of covariate X_{ij}, as well as a family specific effect a_i, and an individual level genetic effect g_{ij} for the jth person in the ith family:

$$\text{logit pr}(Y_{ij} = 1 | a_i, g_{ij}, X_{ij}) = \cancel{\mu + a_i} + \beta X_{ij} + g_{ij}$$

CLR is biased for $\beta \neq 0$

True model: $\text{logit}(p_{ij}) = \mu + \beta x_{ij} + a_i + g_{ij}$



CRELR likelihood (Pfeiffer et al., 2001)

Without loss of generality, assume:

1. We are interested in the likelihood for one family consisting of 2 people
2. There is one case in the family, and the the data are sorted so that the case is Y_1.

Define $d(\beta) = \prod_{j=1}^{2}\{1 + \exp(\mu + \beta X_j + a + g_j)\}^{-1}$. The likelihood for this family is:

$$L = \frac{\int \exp(\beta X_1 + a + g_1) d(\beta) \quad dF(a, g)}{\sum_{k=1}^{2} \int \exp(\beta X_k + a + g_k) d(\beta) \quad dF(a, g)}$$

Random effects model

We assume a multivariate normal model for the distribution $F(a, g)$, and model the covariance of the g's according to the degree of kinship between members of the family.

	mother	father	child	grand-child
mother	σ_g^2	0	$0.5\sigma_g^2$	$0.25\sigma_g^2$
father	0	σ_g^2	$0.5\sigma_g^2$	$0.25\sigma_g^2$
child	$0.5\sigma_g^2$	$0.5\sigma_g^2$	σ_g^2	$0.5\sigma_g^2$
grand-child	$0.25\sigma_g^2$	$0.25\sigma_g^2$	$0.5\sigma_g^2$	σ_g^2

Maximising the likelihood

Use maximum simulated likelihood:

1. For each family draw M i.i.d. samples a and $\mathbf{g}$
2. Evaluate the likelihood at each of the M samples
3. Return the mean of the simulated likelihood values

There is not much information on μ or σ_a^2, so maximise the likelihood over a grid:

- $\mu \in \{-8, -7, \ldots, -2, -1\}$
- $\sigma_a^2 \in \{0, 0.5, 1, 1.5, \ldots, 19, 19.5, 20\}$

reclogit

Syntax

```
reclogit depvar indepvars [if] [in] ,
    group(varname) corrvars(varlist)
    mcsamples(#) [options]
```

family	rel_type	c1	c2	c3	c4
1	Sibling	1	.5	.	.
1	Sibling	.5	1	.	.
2	Mother	1	.25	.5	.5
2	Mat. Cousin	.25	1	.25	.25
2	Sibling	.5	.25	1	.5
2	Sibling	.5	.25	.5	1

`reclogit` progress

I wish I could tell you that you can download `reclogit` from the ssc archive, but you can't.

Testing is still in alpha stages, but barring any disasters `reclogit` should be ready for public consumption early next year.

Acknowledgements

- Ruth Pfeiffer (NCI, NIH)

Pfeiffer RM *et al.*, Inference for covariates that accounts for ascertainment and random genetic effects in family studies. Biometrika (2001), 88, 4, *pp*. 933–948

- Gianluca Severi
- Laura Baglietto
- Cancer Council Victoria