

# Multipoint Mapping for Complex Diseases

## Issues and Applications to Age at Onset

Dave Glidden, UCSF

March 19, 2002

### **Outline**

- ARP Linkage Analysis
- Heterogeneity: Age at Onset
- Multipoint Mapping Methods
- Incorporating Age at Onset
- Numerical studies
- Reflections

Slide 1

### **Complex Genetic Diseases**

- Relatively Common Disease
- Multiple genetic loci
- Environmental risk factors
- Non-genetic cases (phenocopies)
- Gene-environment interaction

Slide 2

### ARP Linkage Analysis

- Recruit affected relatives  
*usually sib pairs*
- Type large # markers  
*known locations*
- Infer # IBD Allele shared  
*0,1 or 2 at markers*
- Markers not always fully informative
- No linkage  
 $IBD = (1/4, 1/2, 1/4)$
- Several test stats for linkage

Slide 3

### Age at Onset Problem

- Locus affects *if* and *when*
- Early onset cases 'genetic'
- Age at onset penetrance
- Disease allele: covariate in Cox model
- Li and Hsu (2000)  
*grouping is inefficient*

Slide 4

### Multipoint Approach

- Can test  $H_0$ :  
*LR, mean IBD tests*
- Region with linkage,  
where is disease locus?
- Point with lowest p-value?
- Liang et al (2001)  
*Use multipoint information,  
formally estimate location*

Slide 5

### Data Set-Up

- $n$  affected sib pairs
- $R$ : region with linkage
- Type  $M$  markers at  
 $(t_1, \dots, t_M)$  cMorgans
- $S_i^*(t_j)$ : # IBD at  $t_j$   
*observed or inferred*
- Where is disease locus  $\tau$ ?

Slide 6

### A Clever Representation

$$E\{S^*(t_j) - 1\} = (1 - 2\theta_{t_j, \tau})^2 C \quad (1)$$

- $C = E\{S(\tau) - 1\}$   
*Excess IBD sharing at  $\tau$*
- $\theta_{t_j, \tau}$  recomb betw  $t_j$  &  $\tau$
- IBD sharing at  $t_j$  depends on  
*location of  $\tau$ , recomb and signal*

Slide 7

### Estimating $\tau$

- $\theta_{t, \tau}$  Haldane mapping  
$$\theta_{t, \tau} = (1 - e^{-0.02|t - \tau|})/2$$
- Model  $E\{S^*(t_j)\}$  ( $j = 1, \dots, M$ ) using (1)
- Analogous to marginal mean in longitudinal data
- Working independence GEE
- Estimate  $\tau$  and  $C$
- Test  $H_0 : C = 0 \Rightarrow S(\tau) = 1$

Slide 8

### Age at Onset

- $X$ : age at onset summary  
*e.g., avg onset for pair*

$$E\{S^*(t_j|x) - 1\} = (1 - 2\theta_{t_j, \tau})^2 C(x) \quad (2)$$

- $C(x) = E\{S(\tau|x) - 1\}$
- How to model  $C(x)$ ?
- Is  $C(x) = C$ ?
- Note: (2)  $\Rightarrow$  (1)
- Modeling  $C(\cdot)$  narrow  $\tau$ ?

Slide 9

### Age at onset

Slide 10

### Parametric Models

- $\text{logit}\{C(x)\} = \alpha + \beta x$   
 $1 \geq C(x) \geq 0$
- $\log\left\{\frac{1+C(x)}{1-C(x)}\right\} = \alpha + \beta x$ ,  
 $1 \geq C(x) \geq -1$
- Generalize to spline basis
- Categorical predictors  
 separate  $C$  values

Slide 11

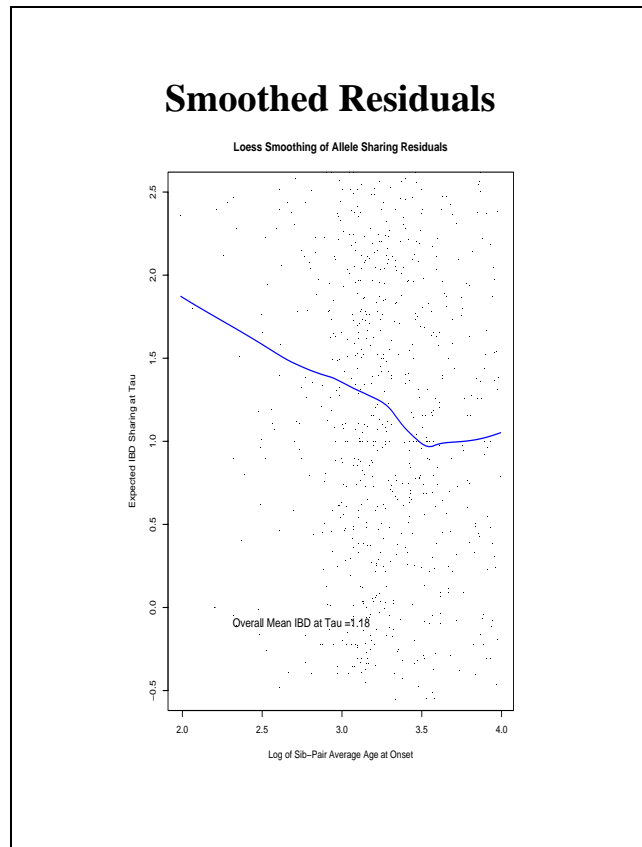
### Exploratory Approach

- Estimate  $\tau$  using (1)
- Calculate  $\hat{C}_i$

$$\hat{C}_i = \sum_{j=1}^M w(t_j) \{S_i^*(t_j) - 1\} (1 - 2\theta_{t_j, \hat{\tau}})^{-2}$$

- $w(t_j) : \sum_{j=1}^M w(t_j) = 1$
- Plot & smooth  $\hat{C}_i$  vs  $X_i$

Slide 12



Slide 13

### Simulation Studies

- ASP with recorded age at onset
- Ascertained with onset  $\leq 45$
- Weibull penetrance (Li & Hsu, 2000)  
 $r = 4.8$   $\lambda = 0.0113$   
*co-dom, gen HR = 3.5, allele fr.=0.05*
- Age onset  
 $X_1$ : log mean age onset (continuous)  
 $X_2$ : mean onset  $\leq 25$  (categorical)

Slide 14

### More on Simulations

- Markers at (0,10,20,...,80,90,100)
- $\tau = 45$
- $\hat{\tau}_1$ : fit based on (1)
- $\hat{\tau}_2$ : fit from logistic model
- 5,000 simulations
- 250, 500 sibpairs

Slide 15

### Age Onset Continuous ( $X_1$ )

$n$		mean	SE	Rel Efft
250	$\hat{\tau}_1$	45.1	5.05	-
	$\hat{\tau}_2$	44.9	3.58	1.41
500	$\hat{\tau}_1$	45.0	3.20	-
	$\hat{\tau}_2$	45.0	2.26	1.42

Slide 16

**Age Onset Categorical ( $X_2$ )**

$n$	mean	SE	Rel Efft
250			
	$\hat{\tau}_1$ 45.1	5.05	-
	$\hat{\tau}_2$ 44.9	3.62	1.39
500			
	$\hat{\tau}_1$ 45.0	3.20	-
	$\hat{\tau}_2$ 45.0	2.27	1.41

Slide 17

**Conclusions**

- Approach: model  $C(x)$
- Residuals guide exploration
- Splines extend modeling
- Model improves  $\hat{\tau}$   
*even when wrong*
- Methods test/explore heterogeneity
- General approach  
*general relatives, covariates*  
*multiple members, unaffecteds*

Slide 18

### **Future Work**

- Compare parametric models
- Explore spline fits
- Extend programs
- Application to dataset
- Compare to other methods  
*Olson, Greenwood & Bull*
- Explore asymptotics

**Slide 19**

### **Many Thanks to:**

- Kung-Yee Liang
- Hongzhe Li
- Mark Segal
- Yen-Feng Chiu

**Slide 20**

### **Primary References**

- Liang, K.-Y., Chiu, Y.-F., Beaty T.H. (2001).  
*Hum. Hered.* **51**, 64-78.
- Li H., Hsu L. (2000).  
*Ann. Hum. Genet.* **64**, 239-253.