

# Issues in Survival Analysis in Genetics

Dave Glidden  
University of California, San Francisco  
[dave@biostat.ucsf.edu](mailto:dave@biostat.ucsf.edu)

JSM, 8/9/2005

## Compelling Questions

- Are their mutations in a candidate gene which cause disease?  
Lin and Zeng
- What is the effect of a known mutation on disease risk?  
Hsu and Gorfine  
Gail et al.

## Lin and Zeng

- Existing cohort study (e.g., CHS study)
- Interrogate for a disease
- Sample cases and subset of controls
- Genotype based on stored samples
- Use haplotypes to relate SNPs to disease

## Gail, et al. & Hsu

- Recruit probands (cases & controls)
- Genotype probands
- Get disease information for family members
- Some variants in info. on family members
- Estimate penetrance of mutations:  
cumulative risk, relative hazard

## The Talks

- Address interesting, topical questions
- Have some similar themes
  - outcome-based sampling
  - ambiguous genetic predictors
  - confront tough semiparametrics
- Difference: family based vs. not



## Time-to-Event Haplotype

	Unknown Haplotype	Known Haplotype
Case-Cohort	Lin & Zeng (2005)	Scheike & Martinussen (2004)
Full Cohort	Lin (2004)	Cox (1972)

## Lin Approach

- Postulate general transformation model
- Estimates based on non-parametric maximum likelihood
- EM is natural here: haplotypes are latent two levels: ambiguous and missing by design
- M-step is a Cox model

## Lin Results

- Maximum likelihood based on EM with very involved E-step
- Used general transformation model worked cleverly into the E-step
- Allows departure from Hardy-Weinberg
- Full asymptotic theory
- Theory for Schieke & Martinussen (2004)

## Open Issues

- How to model covariates:  
in practice, this is the largest issue
- How would perform with larger number  
of haplotypes?
- Effect of population stratification:  
incorporating “genomic control” methods

## Complication of Family Case-Control

- Case families over sampled for other risks  
penetrance biased upwards (Begg 2002)
- Effect of mutation not same in case families  
compared with control families
- Magnitude of bias is unclear
- Requires joint model to address this bias  
copula or frailty model

## Models

	Gail	Hsu
Joint Model	General Copula	Gamma Frailty
Mutation Effect	Marginal Cox with Time- Varying HR	Family Specific Cox

## Gail Approach

- Parses likelihood:
  - $L_{cc}$ : depends on  $\beta, \Lambda_0$   
maximized over  $\beta$
  - $L_{kc}$ : depends on  $\beta, \theta, \Lambda_0, f$   
maximized over  $\beta, \theta, \Lambda_0$
  - $f$ : inferred based on case-control data
- Joint and pairwise approaches for  $L_{kc}$
- Uses an ES-type algorithm for estimation:

## Hsu's Approach

- EM approach (very natural):  
missing data: relatives genotype, frailty
- Estimates  $\theta$  using stratified Cox on proband ages with  $\beta$  treated as fixed
- Maximizes over  $\beta$  &  $\Lambda_0$  (using EM)
- E-step kinda complex
- Uses fixed value of  $f$

## Results (both papers)

- Gain in efficiency for  $\beta$  from relative's data
- Some robustness to frailty/copula
- Pairwise likelihood: good rel. efficiency
- Good performance of inference
- Theory not complete:  
neither is ML estimation

## Thoughts: Gail

- Approach seems like EM with an approximate M-step
- Pairwise likelihood is a nice idea
- Could use splines for non-prop hazards
- Possible to model:  $\Lambda_0$  &  $\Lambda_1$  for the two genotypes

## Thoughts: Hsu

- Approach is almost EM  
worth trying a real M-step for  $\theta$
- Family-specific formulation: not appealing
- Possible to model:  $\Lambda_0$  &  $\Lambda_1$  for the two genotypes
- Can this approach be used for pairwise likelihood?

# Closing Thoughts

- Interesting questions
- Important common themes
- Big thanks and congratulations to the authors and **co-authors**