Proportional Hazards Models With Continuous Marks

Yanqing Sun, Peter B. Gilbert, and Ian W. McKeague

Annals of Statistics 37(1), Feb 2009

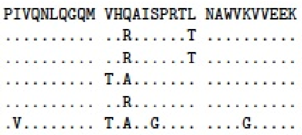
Presented by Jason Shao

University of Washington Biostatistics
Stat/Biost 572

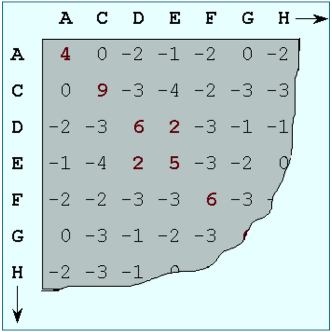
May 9, 2013

Scientific Motivation

- Randomized HIV vaccine trials
- Major difficulty: Differential vaccine efficacy (VE)
- Vaccines don't protect as well against "unfamiliar" viruses
- Can quantify genetic diversity using Hamming distance (continuous)



GAG fragment alignment



	A	C	D	E	F	G	H
A	4	0	-2	-1	-2	0	-2
C	0	9	-3	-4	-2	-3	-3
D	-2	-3	6	2	-3	-1	-1
E	-1	-4	2	5	-3	-2	0
F	-2	-2	-3	-3	6	-3	
G	0	-3	-1	-2	-3		
H	-2	-3	-1				

Part of the Hamming substitution matrix

The fundamental statistical problem

- Want to perform inference on vaccine efficacy, accounting for infecting type in the model.
- Measures on infecting type **only observed in infected subjects** - *cannot* be treated as ordinary covariates!

i	1	2	3	4	5	6	7	8	...
T	13.3	7.4	4.8	16.3	1.5	14.2	16.3	7.0	...
δ	1	1	1	0	0	0	0	1	...
V	0.72	0.80	0.51					0.06	...
X	1	0	1	0	0	1	1	0	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

- T: follow-up time
- δ: failure indicator
- V: mark variable
- X: Vaccine (1) or placebo(0). Could be a vector, including other covariates of interest (possibly time-varying).

Proportional Hazards Model

- Developed by D.R. Cox [1972], allows for semiparametric inference on treatment effects (and other covariates) on hazards:

$$\lambda(t|X_i(t)) = \lim_{\Delta t \to 0} \frac{\Pr(T \in [t, t + \Delta t)|X_i(t), T > t)}{\Delta t}$$

- We care about this because under certain assumptions, vaccine efficacy is defined as a function of hazard ratio:

$$VE = 1 - \frac{\lambda(t|X_1 = 1)}{\lambda(t|X_1 = 0)}$$

- Goal: Incorporate continuous marks into this model, so inference can be performed on $VE(v)$.

Partial Likelihood

- Under the proportional hazards assumption:

$$\lambda(t|X_i) = \lambda_0(t)\exp(\beta^T X_i)$$

- Baseline hazard λ_0 factors out in partial likelihood (PL):

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} \left(\frac{e^{\beta^T X_i}}{\sum_{j=1}^{n} Y_j(T_j) e^{\beta^T X_j}} \right)^{\delta_i}$$

    - $Y_i(t)$: indicator of ith individual being at risk at time t.
- Some intuition:
    - Each *uncensored* subject contributes a term to the PL.
    - Weighted by total risk set at subject's failure time.

Discrete competing risks [Prentice *et al.* 1978]

- Define finite risk categories (V = 1,...,K):

$$\lambda_v(t|X_i) = \lambda_{0v}(t)\exp(\beta_v^T X_i)$$

- Baseline hazard λ_{0v} still factors out in partial likelihood:

$$\mathcal{L}(\beta_v) = \prod_{i=1}^{n}\left(\frac{e^{\beta_v{}^T X_i}}{\sum_{j=1}^{n} Y_j(T_j)e^{\beta_v^T X_j}}\right)^{\delta_i \times \mathbb{I}_{V_i=v}}$$

  - $Y_i(t)$: indicator of ith individual being at risk at time t.
- Some intuition:
  - Only failures of type v contribute to PL for β_v.
  - Other failures are treated as censored observations.

Continuous competing risks [Sun *et al.* '09]

- Use a continuous (bounded) mark $V = \in (0, 1)$):

$$\lambda(v, t|X) = \lambda_0(t, v) \exp(\beta(v)^T X)$$

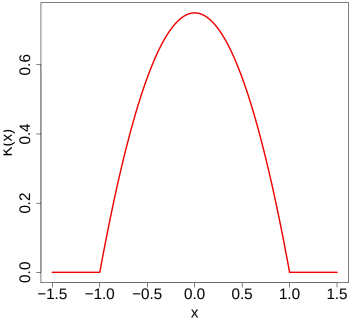
- Localized log partial likelihood contains a kernel smoothing function K:

$$\mathcal{L}(\beta(v)) = \prod_{i=1}^{n} \left(\frac{K_h(V_i - v) e^{\beta(v)^T X_i}}{\sum_{j=1}^{n} Y_j(T_j) e^{\beta(v)^T X_j}} \right)^{\delta_i}$$

  - $Y_i(t)$: indicator of ith individual being at risk at time t.
- Some intuition:
  - Can't use $\mathbb{I}_{V_i = v}$ typically unique values of v
  - Instead, kernel function $K_h(\cdot)$ "borrows" information from nearby observations, with larger weight on marks near v

Kernel functions

- Can basically use any kernel function $K(x)$ that has support $[-1, 1]$ and maximum at $x = 0$
- Sun *et al.* ['09] use Epanechnikov's kernel: $K(x) = 0.75(1 - x^2)$:



- Bandwidth: $K_h(x) = K(x/h)/h$
  - Essentially a normalizing constant adjusting for how much distance is allowed.

Counting process notation

- Another way to write the localized log partial likelihood :

$$\ell_v(\beta_v) = \sum_{i=1}^{n} \int_0^1 \int_0^\tau K_h(u-v) \times \left[\beta^T(v) X_i(t) - \log\left(\sum_{j=1}^{n} Y_j(t) e^{\beta^T(v) X_j} \right) \right] \times N_i(dt, du)$$

    - $t \in (0, \tau)$: Follow-up period
- $N_i(t, v) = I(X_i \leq t, \delta_i = 1, V_i \leq v)$ is a counting process: basically count each uncensored failure time, even if v's don't match, as long as kernel is nonzero.
- This notation is useful for derivation of asymptotic results using martingale theory.

Point estimation

- Possible targets of inference:
  - $\hat{\beta}_1(v)$: log hazard ratio. $\hat{VE}(v) = 1$ - $\exp(\widehat{\beta}_1(v))$
- Calculations of score and information functions at $\hat{\beta}(v)$ (basically follow Cox ['72]):
  - Helper functions S and J:

$$S^{(j)}(t,\beta) = n^{-1} \sum_{i=1}^{n} Y_i(t) e^{\beta_T Z_i} Z_i^{\otimes j}$$

$$J_n(t,\beta) = \frac{S^{(2)}(t,\beta)}{S^{(0)}(t,\beta)} - \left(\frac{S^{(1)}(t,\beta)}{S^{(0)}(t,\beta)} \right)^{\otimes 2}$$

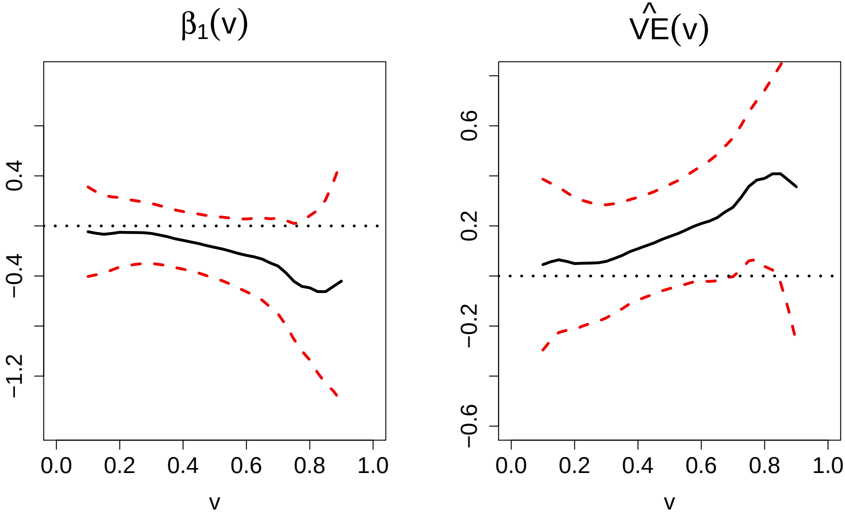
  - Information matrix for $\hat{\beta}(v)$:

$$\ddot{\ell}_\beta(v, \beta(v)) = -\sum_{i=1}^{n} \int_0^1 \int_0^\tau K_h(u-v) J_n(t,\beta) N_i(dt, du)$$

Asymptotic variances

- Asymptotic normality also follows from Cox ['72]:
  - $\sqrt{nh}(\hat{\beta}(v) - \beta(v)) \rightarrow_d N(0, \nu_0 \Sigma^{-1}(v))$
    - Here, $\nu_0 = 3/5$ (integral of squared kernel function)
    - $\hat{\Sigma}(v) = -n^{-1}\ddot{\ell}_\beta(v, \hat{\beta}(v))$
  - $\sqrt{nh}(\widehat{VE}(v) - VE(v)) \rightarrow_d N(0, \nu_0 \sigma_1^2(v) e^{2\beta_1(v)})$
    - Results easily from delta method
- These asymptotic variances can be used for pointwise confidence bands.

Pointwise confidence bands

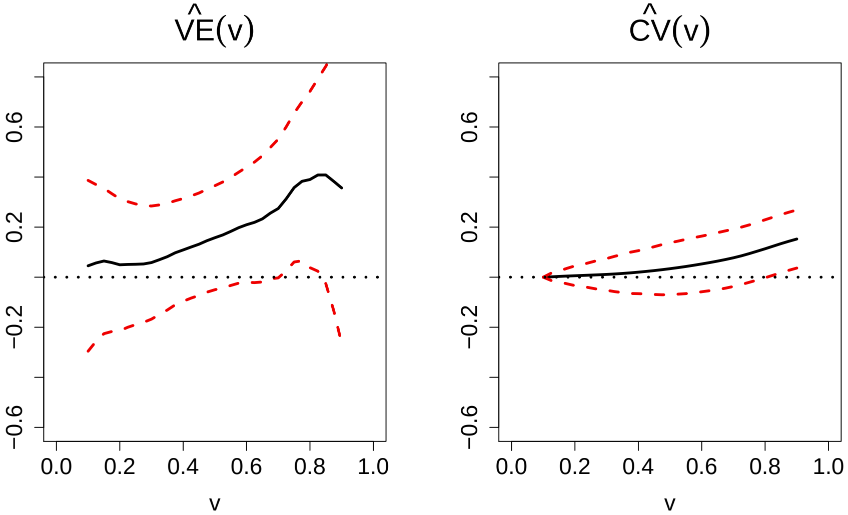


Dataset: VaxGen004: $n = 5403$; 2:1 randomization; 336 failures with observed marks

Hypothesis testing

- Goal: test for nonzero $VE(v)$ or dependence on v over a range of v.
- Another target of inference: cumulative vaccine efficacy
  - $\widehat{CV}(v) = \int_a^b \widehat{VE}(u)du$, $[a, b] \in [0, 1]$
- Asymptotics are a bit trickier:
  - $\sqrt{nh}(\widehat{CV}(v) - CV(v)) \rightarrow_d N(0, \rho^2(v))$
  - $\hat{\rho}^2(v) = \hat{\Sigma}_{\hat{A}}(v)_{(1,1)}$
  - $\Sigma_{\hat{A}}(v) = n^{-1} \sum_{i=1}^n \int_a^v \int_0^\tau \hat{A}(u) J_n(t, \hat{\beta}(u)) \hat{A}(u)^T N_i(dt, du)$
  - $\hat{A}(v) = e^{\hat{\beta}_1(v)} \hat{\Sigma}(v)^{-1}$
- Test statistics are all functions of $\widehat{CV}(v)$, scaled by $\hat{\rho}^2(v)$

Cumulative vaccine efficacy



What's next?

- Calculations from previous slide still result in pointwise CI's.
- How to evaluate simultaneous CI's?
  - Simulate stochastic processes and evaluate quantiles
  - Resampling techniques such as Gaussian multiplier method
- Test statistics also have null distributions that require these techniques.
- Simulations
  - Simulate time-to-infection data (with censoring) for simple hazard functions
  - Vary linear dependence of hazards on vaccination status and v.
  - Evalute size and power of test statistics.

To be continued...

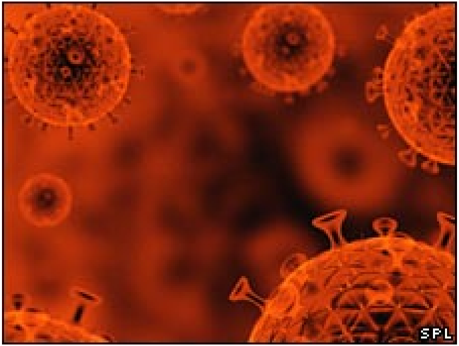


Image: Science Photo Library