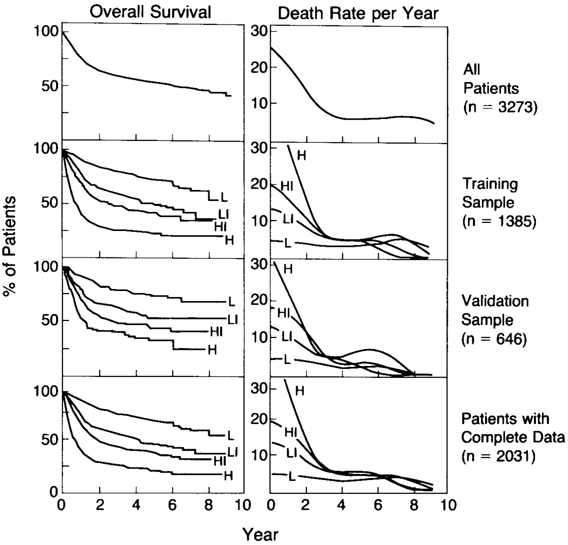
Robust Survival Prediction via Linear Transformation Models

Keith Betts
Dave Harrington
dph@jimmy.harvard.edu

Analysis Group, Harvard University

7 August 2014

Graphic from 1993 *NEJM* paper, prognosis in non-Hodgkin's lymphoma



Covariates in model:

- Age
- Stage
- Tumor size
- Extent of disease
- Performance status

‘Prognostic models’ now widely used

Models that predict the risk of disease or disease progression have a long history in medical literature.

- Gail model for breast cancer risk (*JNCI*, 1989)
- N. Cook’s work in cardiovascular disease (*JAMA, NEJM*)
- Mayo Clinic model in primary biliary cirrhosis (*Hep*, 1989)
- Models for Cardiovascular disease from Framingham Heart Study (D’Agostino, *Circ*, 1998, 2004, 2008)
- Coronary Heart Disease Policy Model, (Weinstein et. al, *Am J Pub Health*, 1987)
- Non-Hodgkin’s lymphoma, (Shipp, et al. *NEJM*, 1993)

Sample of important methodologic literature, prediction with event time data

- Measures of explained variation with event time data, Korn and Simon, 1990,
- Prediction error, Graf et. al, 1999
- Haegerty, et al., 2000
- Gerds and Schumacher, 2006
- Cai, Tian, Solomon, Uno, Wei (2007a, 2007b, 2008)

Our general approach

- Use flexible class of linear transformation models for censored data
- Evaluate time-dependent mean-squared error of prediction, and its standard error
- Avoid bias in apparent error rate
- Obtain unbiased estimates of error rates when model is misspecified

Notation and Models

- T = event time, C = potential censoring time, Z = p-dimensional vector of covariates
- Observed time $\tilde{T} = \min(T, C)$, $\delta = (T \leq C)$
- $S(t|Z) = \Pr(T > t|Z)$
- Semi-parametric Linear Transformation Model

  $$h(T) = -\beta^T Z + \epsilon,$$

  $h(\cdot)$ is a unknown monotone strictly increasing function, ϵ has 'known' distribution.
- Equivalent to

  $$g^{-1}(S(t|Z)) = h(t) + \beta^T Z,$$

  with $g^{-1} = 1 - F_\epsilon$

Goal

- Predict survival probability,

  $$\hat{S}(t|Z^0) = g(\hat{h}(t, \hat{\beta}) + \hat{\beta}^T Z^0)$$

  for an 'out of sample' individual.
- Estimate mean squared error of prediction (MSEP) as a function of time,

  $$\overline{\mathrm{MSEP}}(t, \hat{S}, G) = E_{T,Z}\{I(T > t) - \hat{S}(t|Z)\}^2,$$

  even when working model is wrong.
- This is expected Brier score, originally used in weather prediction

Assumptions

- $(T \perp C)|Z$
- Z is bounded
- $G(t|Z) = \Pr(C > t|Z)$ can be consistently estimated
- An assortment of regularity conditions

Formulation for *MSEP* similar to Graf, Gerds.

Proofs rely on work by H. Uno, T Cai, L Tian and LJ Wei on asymptotics of mis-specified models

Estimating equations and main results

- Estimating Equations

$$
\begin{aligned}
U_1(h(t), \beta) &= \sum_{i=1}^{n} \left[I(\tilde{T}_i \geq t) - g(h(t) + \beta^T Z_i) \hat{G}(t|Z) \right] \\
U_2(\hat{h}(t, \beta), \beta) &= \sum_{i=1}^{n} \int_{\tau_a}^{\tau_b} Z_i \left[I(\tilde{T}_i \geq t) - g(\hat{h}(t, \beta) + \beta^T Z_i) \hat{G}(t|Z) \right] dt
\end{aligned}
$$

- Main results: Even when S is mis-specified
  - Unique solutions $\hat{h}(t, \beta)$ and $h_*(t, \beta)$ for $U_1(h(t), \beta)$ and its expectation for a fixed β
  - Unique solutions $\hat{\beta}$ and β_*, for $U_2(\hat{h}(t), \beta)$ and $E[U_2(h_*(t), \beta)]$
  - Consistency, $\hat{\beta} \xrightarrow{p} \beta_*$
  - Uniform consistency, $\sup_t |\hat{h}(t, \hat{\beta}) - h_*(t, \beta_*)| \xrightarrow{p} 0$

Results . . .

There exists a survivor function $\bar{S}$ such that

- $\sqrt{n}\{\hat{S}(t|Z^0) - \bar{S}(t|Z^0)\}$ converges to a Gaussian process, and

$$\overline{\text{MSEP}}(t, \bar{S}, G) = E_Z\{S(t|Z) - \bar{S}(t|Z)\}^2 + E_Z\{S(t|Z)(1 - S(t|Z)\}$$

- Limiting distribution has complicated covariance structure but does not depend on censoring distribution, even when S has been mis-specified.

Estimation of *MSEP*

- *MSEP* estimate

$$\widehat{\text{MSEP}}(t, \hat{S}, \hat{G}) = n^{-1} \sum_{i=1}^{n} \{I(\tilde{T}_i \geq t) - \hat{S}(t|Z_i)\}^2 w(t, \hat{G}, Z_i),$$

  where

$$w(t, \hat{G}, Z_i) = \frac{I(\tilde{T}_i \leq t)\delta_i}{\hat{G}(\tilde{T}_i - |Z_i)} + \frac{I(\tilde{T}_i > t)}{\hat{G}(t|Z_i)}.$$

- Uniform consistency

$$\sup_t |\widehat{\text{MSEP}}(t, \hat{S}, \hat{G}) - \overline{\text{MSEP}}(t, \bar{S}, G)| \overset{p}{\to} 0.$$

- Inference on $\sqrt{n}\{\widehat{\text{MSEP}}(t, \hat{S}, \hat{G}) - \overline{\text{MSEP}}(t, \bar{S}, G)\}$ using perturbation resampling.

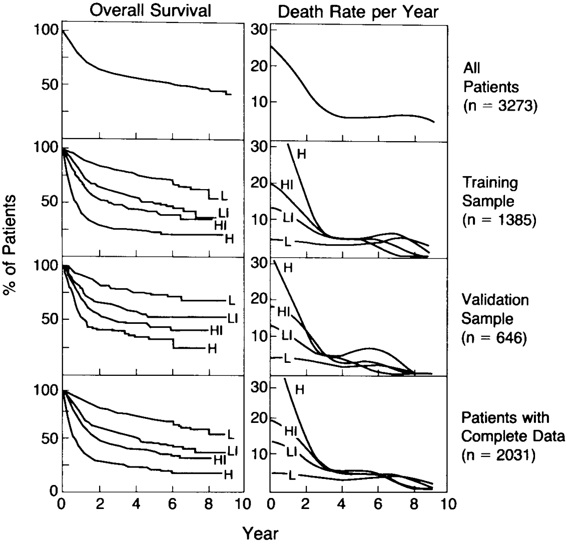
Cross validation to estimate *MSEP*

- Ideally: Independent training and test datasets
- Apparent error will be overly optimistic
- K-fold cross validation

$$\widehat{\mathrm{MSEP}}^{\mathrm{CV}} = \frac{1}{K}\sum_{k=1}^{K}\widehat{\mathrm{MSEP}}(\hat{h}^{(-k)}(t,\hat{\beta}^{(-k)}),\hat{\beta}^{(-k)}),$$

  where $\hat{h}^{(-k)}(t,\hat{\beta}^{(-k)})$, and $\hat{\beta}^{(-k)}$ are estimated using the data in the $K-1$ datasets not including set k.
- Simulations show procedure works reasonably well – more interesting to look at examples.

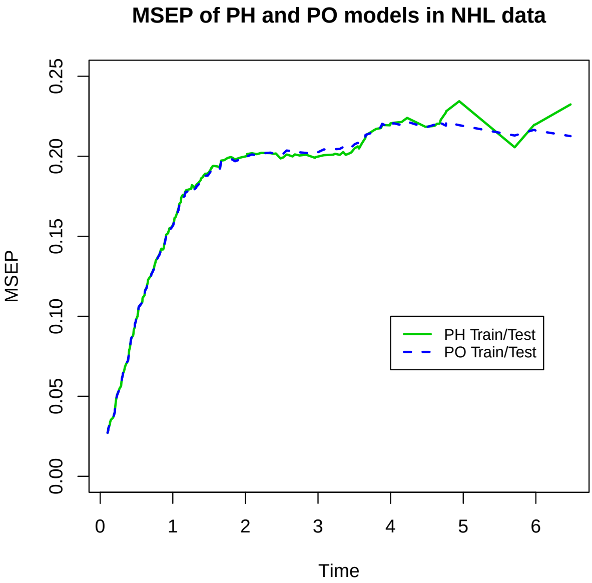
Lymphoma Data



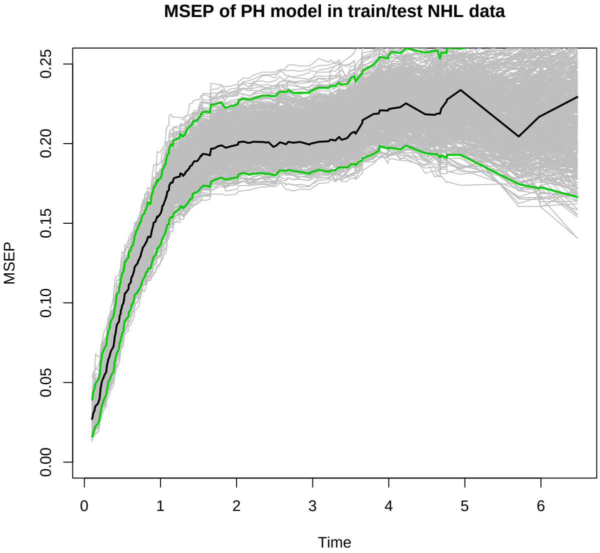
Covariates in model:

- Age
- Stage
- Tumor size
- Extent of disease
- Performance status

Lymphoma data: *MSEP*, original binary covariates



Lymphoma data: mean squared error of prediction with confidence intervals, validation dataset

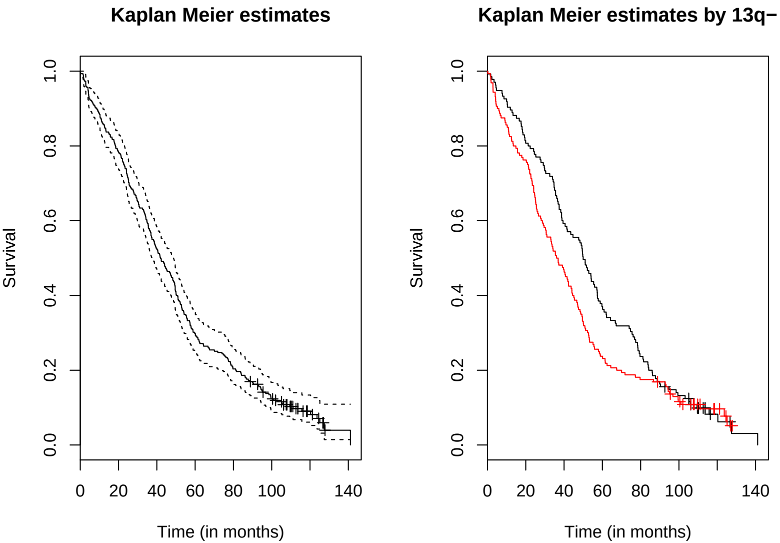


Multiple Myeloma

Myeloma study from Eastern Cooperative Oncology Group (E9486)

- Randomized trial of three treatments in multiple myeloma, no survival differences observed among 653 patients
- 295 participant specimens randomly chosen for analysis of a deletion on long arm of chromosome 13 (13q-). 270 deaths
- Originally reported in *JCO* (1999), *Blood* (2001), *Biometrics* (2002)

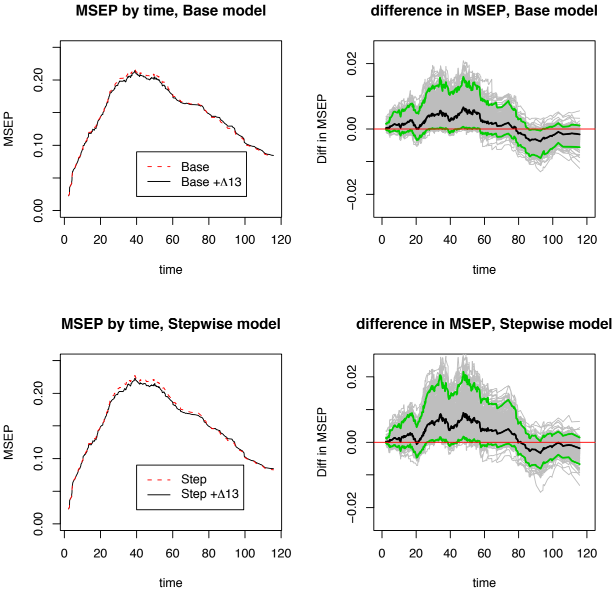
Myeloma: Value of an additional marker?



Working PH model coefficient estimates

	Base	Base $+\Delta 13$	Stepwise	Stepwise $+\Delta 13$
$\Delta 13$	-	-0.57 (.21)	-	-0.68 (.21)
Albumin	0.30 (.27)	0.25 (.27)	-	-
β_2 micro	-0.48 (.25)	-0.48 (.26)	-0.55 (.26)	-0.53 (.25)
Creatine	-0.65 (.28)	-0.80 (.29)	-0.76 (.27)	-0.88 (.28)
Hemoglobin	-0.31 (.30)	-0.29 (.31)	-	-
IgA	0.10 (.22)	0.10 (.24)	-	-
IgG	-0.42 (.32)	-0.46 (.32)	-	-
Light chain (κ)	-0.67 (.39)	-0.76 (.38)	-	-
% plasma cells	-0.78 (.23)	-0.69 (.24)	-	-
PCLI	0.49 (.24)	0.41 (.23)	-	-
IL-6	-0.42 (.21)	-0.44 (.21)	-0.51 (.20)	-0.52 (.20)
C-reactive	-0.88 (.23)	-0.87 (.25)	-0.91 (.22)	-0.89 (.23)
Durie-Salmon	-0.07 (.25)	-0.16 (.25)	-	-

MSEP for Myeloma models



Some simulations

True vs. apparent vs. cross-validated *MSEP* for the Linear Transformation Model (LTM) and the Cox model evaluated at 1st quartile and median for simulated data.

		$q = .25$			$q = .5$		
		Truth	App.	CV	Truth	App.	CV
A	MSEP_{LTM}	.131	.134	.132	.132	.134	.130
	MSEP_{Cox}	-	.133	.131	-	.134	.131
B	MSEP_{LTM}	.161	.169	.165	.210	.215	.211
	MSEP_{Cox}	-	.168	.164	-	.214	.211
C	MSEP_{LTM}	.136	.142	.139	.145	.152	.147
	MSEP_{Cox}	-	.155	.152	-	.159	.154

A: PH data, correctly fit with LTM
B: PH data, correct LTM, but neglected covariate
C: PH data, PO model fit for LTM

Limitations

- Estimating equations are not efficient
- Must estimate censoring distribution (correctly!)
- *MSEP* not easily interpreted
- Falls short of predicting event times