Biomarkers, Cancer prediction, Web calculators and Bayesian MCMC

Donna Pauler Ankerst, Ph.D.

Institute for Medical Informatics, Biometry and Epidemiology, University of Munich
Department of Biostatistics, University of Michigan
Department of Urology, University of Texas Health Science Center at San Antonio

1

Why online calculators for cancer risk?

Cancer is typically influenced by several risk factors.

Statistical models assess the simultaneous effect of the risk factors on cancer risk, and report odds ratios or other measures that give the effect of each risk factor adjusted for the other risk factors.

It is not easy for a clinician to translate a list of odds ratios from a statistical model into a coherent risk for a patient.

2

Why online calculators for cancer risk?

The problem is exacerbated for serial measurements of risk factors, for predictions of time until an event, or for high-dimensional risk predictors such as genomic or proteomic markers.

Statisticians need to bring results from their models back to the „patient bedside“ for them to be usable.

Therefore, there has been a recent surge in online risk prediction and prognosis models; just recently a National Institutes of Health (NIH) funding announcement for this purpose.

3

Examples

Gail Breast cancer model http://www.halls.md/breast/risk.htm

Prostate cancer survival www.prostatecalculator.org

BRCAPRO – Breast/Genetic
http://www.isds.duke.edu/~gp/brcapro.html

Prostate Cancer Prevention Trial (PCPT)
prostate cancer risk tool
www.compass.org/edrnnci/bin/calculator/main.asp →

4

Risk factors for biopsy-detectable prostate cancer

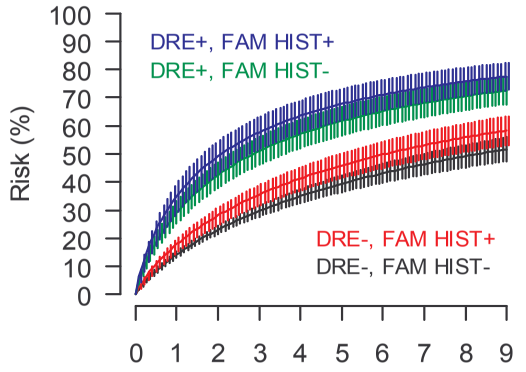
Using multivariable logistic regression models we analyzed the relationship between prostate-specific antigen (PSA), digital rectal examination (DRE), family history of prostate cancer, age, race, and history of a prior biopsy in 5519 men from the PCPT placebo arm who had a biopsy performed within one year of their PSA and DRE.

	OR	95% CI	p-value
Log(PSA)	2.34	(2.13, 2.56)	< 0.0001
Family history	1.31	(1.11, 1.55)	0.002
DRE	2.47	(2.03, 3.01)	< 0.0001
Prior biopsy	0.64	(0.53, 0.78)	< 0.0001

Thompson, Ankerst, Chi, et al., JNCI, 2006.

5

Prostate cancer risk by PSA, DRE, and Family history for men with no prior biopsy



6

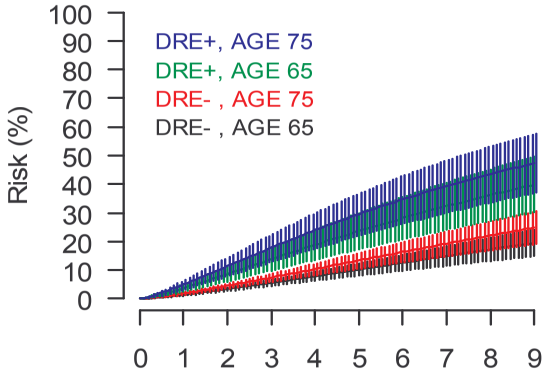
Results: Gleason ≥ 7 Disease

Based on a multivariate analysis, the following factors contributed independent diagnostic information to the risk of high-grade prostate cancer.

	OR	95% CI	p-value
Log(PSA)	3.64	(3.04, 4.37)	< 0.0001
DRE	2.72	(1.96, 3.77)	< 0.0001
Age	1.03	(1.01,1.06)	0.01
African American	2.61	(1.55, 4.41)	< 0.0001
Prior biopsy	0.70	(0.49, 0.99)	0.04

7

Risk of high-grade disease



8

Online PCPT Risk Calculator

http://www.compass.fhcrc.org/edrnnci/bin/calculator/main.asp

After accepting a disclaimer

Predicting Likelihood Of Cancer If A Prostate Biopsy Is Performed

The fields with * sign are required.

Race: *	Caucasian	
Age: *	65	
PSA Level: *	2.5	ng/ml
Family History of Prostate Cancer: *	No	
Digital Rectal Examination Result: *	Normal	
Prior Negative Prostate Biopsy: *	No	

Submit

9

Predicting Likelihood Of Cancer If A Prostate Biopsy Is Performed

The Result:

Based on the data provided, the person's estimated risk of biopsy-detectable cancer is **27** % .

The **95**% Confidence Interval for this prediction is **25**% to **28**%.
More information about confidence interval ...

The person's estimated risk of biopsy-detectable high grade prostate cancer is **4** % .

The **95**% Confidence Interval for this prediction is **3.5**% to **5.3**%.
More information about confidence interval ...

The result is based on:

Age:	65
Race:	Caucasian
PSA Level:	2.5 ng/ml
Family History of Prostate Cancer:	No
Digital Rectal Examination Result:	Normal
Prior Negative Prostate Biopsy:	No

10

The burden to validate and compare

Internet prediction tools go beyond the reporting of findings of risk associations and are subject to misuse by the public.

Therefore, despite the numerous disclaimers and elucidation of the specific population used to develop the calculator, there is ongoing validation of calculators on multiple heterogeneous external populations.

There is also the requirement to compare new prediction tools to current practice based on simple rules in terms of accuracy and operating characteristics.

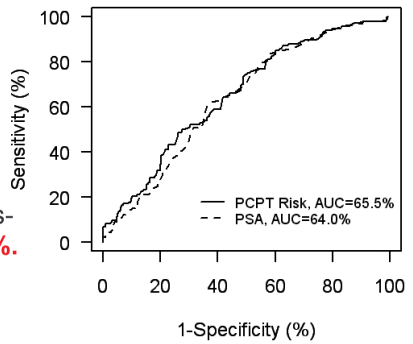
11

External validation on the San Antonio study of Biomarkers of Risk for Prostate Cancer (SABOR)

Risk no better than PSA (also in PCPT)

AUC for PCPT Risk calculator applied to **SABOR: 65.5%**

compared to internal cross-validation on **PCPT: 70.2%.**



Only 5% reduction from the population on which it was developed (SABOR younger, more ethnic than PCPT).

Predictions based on serial biomarkers

Potentially more powerful since more information is contained in an individual's marker history than in a single measurement; each individual serves as his own control.

Computationally more difficult -- based on joint statistical models for longitudinal marker measurements and either dichotomous, categorical or time-to-event outcomes fit to to a training set; the mean longitudinal trajectory is often nonlinear, there is considerable person-to-person and within-person variation, a prediction for a new person needs to condition on the new person's longitudinal history.

Bayesian Markov Chain Monte Carlo (MCMC) methods

13

2 Applications

- Screening for ovarian cancer — UK-SCREEN
- Monitoring for prostate cancer recurrence — US-NIH

Common steps

Identify large cohort
Find models/MCMC methods that accurately describe the data
Develop model-based predictions
Compare operating characteristics against current clinical practice
Validate

14

- Screening for ovarian cancer — UK-SCREEN
- Monitoring for prostate cancer recurrence — US-NIH

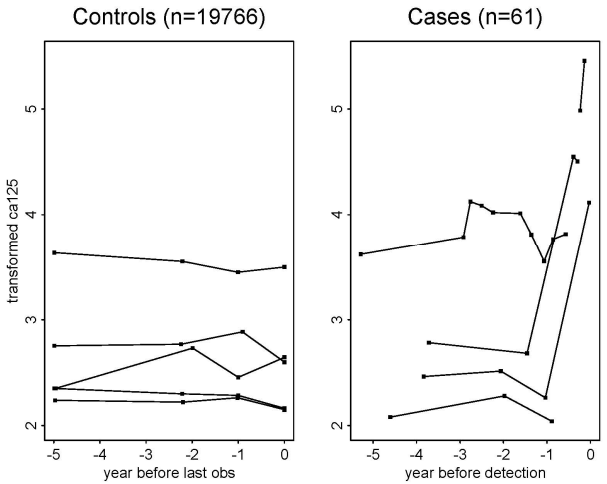
15

Screening for ovarian cancer

- 5-year survival rate if detected in
  - Stage III, IV : 20% — get symptoms, go to doctor
  - Stage I,II: 90% — woman does not know
- No ultrasound screening in the United States due to low incidence and high cost.
- CA125 is a blood test that increases for most ovarian cancers up to 2 years prior to clinical detection.
- CA125 > 35 U/mL refer to ultrasound.
- Can longitudinal CA125 screening improve early detection?

16

UK Screening Trial, Jacobs 1980 (n=22,000)



17

Longitudinal mixture models

Y_{it} = log CA125 for patient i at time t

Controls: constant individual-specific mean over time
$Y_{it} \sim N(\theta_i, \sigma^2)$

Cases: 85% have significant elevation in CA125 with cancer
15% of cancer cases do not express extra CA125
Longitudinal CA125 follows a mixture distribution

$Y_{it} \mid \{I_i = 0\} \sim N(\theta_i, \sigma^2)$ Need reversible jump
$Y_{it} \mid \{I_i = 1\} \sim N(\theta_i + \beta_i(t_{it} - \tau_i)^+, \sigma^2)$ Must borrow information

$I_i \sim Ber(\pi)$ is estimated for each cancer case
$d_i - \tau_i \sim N(2, .75^2)I(0,5)$, where d_i is time of clinical detection,
$I(\cdot)$ is indicator for the interval 0 to 5.

18

Fit to UK screening data

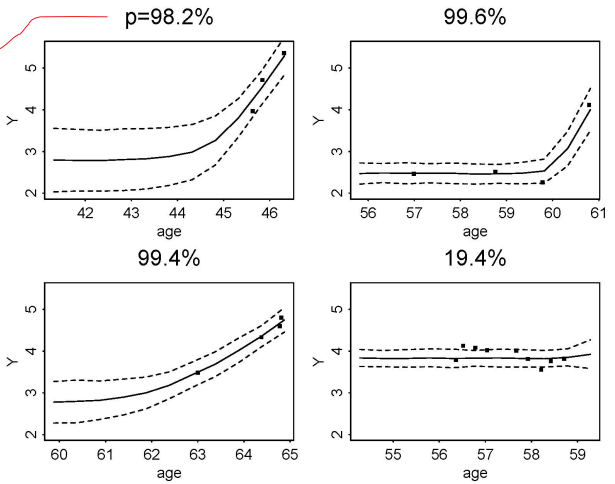
- Population distributions for all random effects.
- Prior distributions for all population parameters.
- Fit to 19766 controls and 61 cases using Markov chain Monte Carlo combination of Gibbs, Metropolis-Hastings and reversible jump steps.
- C program; 5,000 burn-in+5,000 iterations; 4 hours.
- Convergence diagnostics.
- Goodness of fit statistics.
- Sensitivity to prior.

19

Goodness of fit of CA125 model for 4 cancer cases

Posterior probability of a changepoint, $E(I_i|data)$.

Graphs begin 5 years prior to clinical detection and end at clinical detection



Calculating risk of cancer for a new woman

Y_{Train} = CA125 and ovarian cancer status UK screening trial
Y_{New} = CA125 for a new woman

$$P(\text{Cancer} \mid Y_{New}, Y_{Train}) = \frac{P(Y_{New} \mid \text{Cancer}, Y_{Train})\, P(\text{Cancer})}{P(Y_{New} \mid \text{Cancer}, Y_{Train})\, P(\text{Cancer}) + P(Y_{New} \mid \text{Not Cancer}, Y_{Train})\, P(\text{Not Cancer})}$$

$P(\text{Cancer}) \approx 1/1000$ (age-adjusted population prevalence)

$$P(Y_{New} \mid \text{Cancer}, Y_{Train}) = \int P(Y_{New} \mid \Theta, \text{Cancer})\, P(\Theta \mid Y_{Train})\, d\Theta$$

$$P(Y_{New} \mid \text{Not Cancer}, Y_{Train}) = \int P(Y_{New} \mid \Theta, \text{Not Cancer})\, P(\Theta \mid Y_{Train})\, d\Theta$$

Based only on posterior draws/delta method for standard errors

Validation

1.) External validation on an independent Swedish trial:
10 cases, 4276 controls.

2.) Simulated 7 year trial with 500 cases, 500 controls,
posterior CA125 distributions from the UK trial.

3.) Specificity=P(test negative|no cancer)
Sensitivity=P(test positive|cancer)

4.) Receiver operating characteristic (ROC) curve:
plot of Specificity versus Sensitivity as cut-off for
positive test varies (e.g. Risk > c denotes positive test).

Validation on independent Swedish trial

Test set
10 cases — Small number
4276 controls

Test positive if
CA125 > 30 U/mL — Standard Approach
Risk > 1% — Bayesian Risk

Both obtain specificity 99% — False positive rate 1%

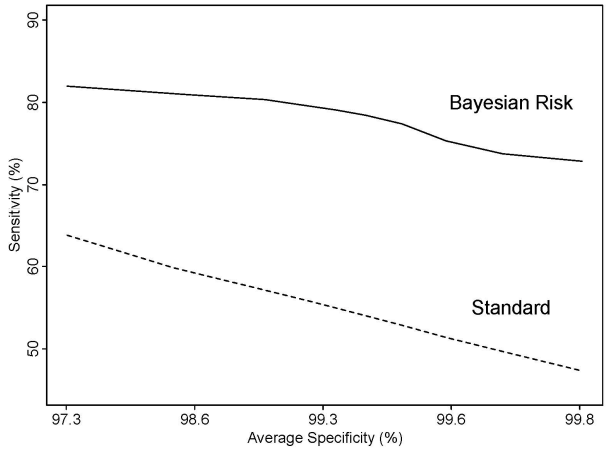
But, the sensitivity
Standard = 65% — Misses 35% cancers
Risk = 80% — Misses 20% cancers

Results: Simulated 7-year study

Population:	500 cases 500 controls	Better power
Standard:	CA125<c CA125>c	Screen in one year Refer to ultrasound
Risk:	Risk<.05% .05%<Risk<d Risk>d	Screen in 1 year Repeat CA125 in x months Refer to ultrasound

x is a linear interpolation between 3 months and 1 yr by risk
Vary c and d to obtain ROC curve

ROC curve *Skates, Pauler, Jacobs, JASA,2001*



25

The importance of early stage detection

In the simulations, define the pre-clinical duration as the time between the change-point in CA125 and clinical detection = $d_i - \tau_i$.

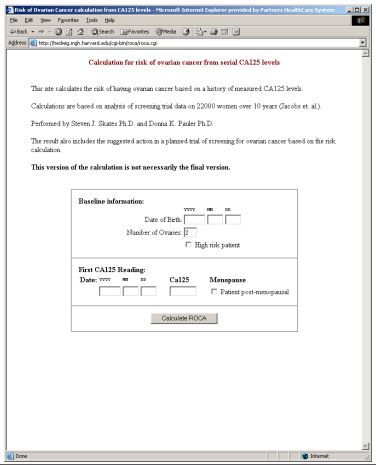
Suppose the early stage was the first 60% of the pre-clinical duration.

Then in the simulations the Bayesian risk screening program would have caught 61% of early stage cancers compared to only 39% by the standard approach.

On the basis of these results the Bayesian risk method was accepted as part of a UK screening trial initiated in 2001.

26

Ongoing until 2012: Risk of cancer calculator evaluated on one arm of UK Trial (n=200,000)



27

- Screening for ovarian cancer UK-SCREEN
- Monitoring for prostate cancer recurrence US-NIH

28

Monitoring patients after radiotherapy for first presentation of prostate cancer

Current practice

Refer to biopsy, ultrasound or immediate salvage therapy if

- current PSA ≥ c or
- PSA doubles in size from its lowest value (nadir) or
- PSA shows 2 or 3 consecutive rises over any time frame

Several shortcomings

- do not account for measurement error, „bouncing PSA“
- subjective, time frames can be ill-chosen
- uncertainty resulting in multiple PSA readings

29

Proposed approach

Joint statistical models

- Account for measurement error
- Allow simultaneous incorporation of multiple time-varying predictors
- Yield an intuitive dynamic probability of recurrence or estimated time to recurrence for new patients

Shortcoming

- Not easy to fit and a black box to physicians
- Requires training data and computation time

Therefore

- Must develop fast online prediction
- Show it performes better than current practice

30

Joint statistical models for longitudinal PSA and time to recurrence for prostate cancer patients treated by radiotherapy and possibly hormones

- Develop form of the joint model: covariate selection for longitudinal and survival components.
- Combine multiple large databases including Massachusetts General Hospital (MGH), University of Michigan Comprehensive Cancer Center (UMCC) and Radiation Therapy and Oncology Group (RTOG)
- Focus on predictions for a new patient—develop an online risk of recurrence tool

31

Building a prediction model

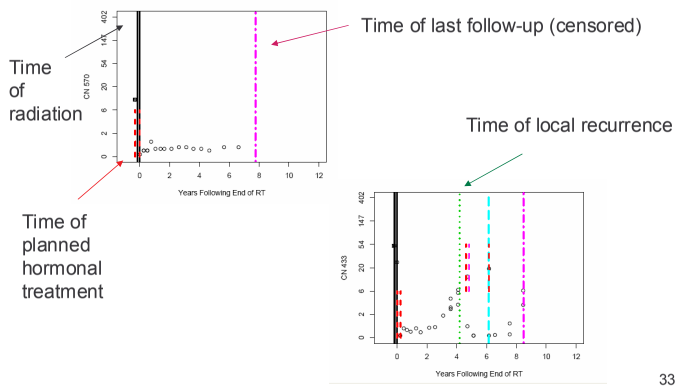
Build and validate model on data from 6 sources of patients treated by radiotherapy and hormonal therapy followed up to 12 years, large censoring.

Massachusetts General Hospital:	n=1011
University of Michigan Cancer Center:	n=1215
Radiation Therapy Oncology Group Trials:	
RTOG 9292	n=779
RTOG 9413	n=1323
RTOG 9406	n=1084
	---
	n=5142

Radiation only (varying doses), radiation plus hormones, salvage hormones, chemotherapy and surgery, advanced versus not-advanced patients.

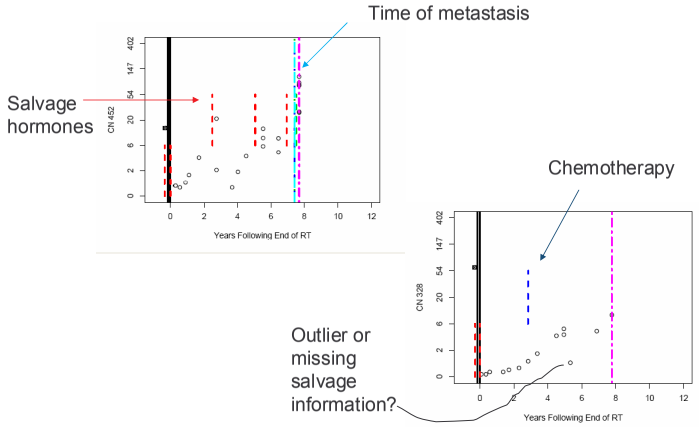
32

Clean-cut examples: PSA, treatment, events from RTOG Study 9413



33

Not-so-clean examples from RTOG Study 9413



Building a radiation-only model: MGH

Massachusetts General Hospital:	n=1011
University of Michigan Cancer Center:	n=1215
Radiation Therapy Oncology Group Trials:	
RTOG 9292	n=779
RTOG 9413	n=1323
RTOG 9406	n=1084

- Accrued from 1988 to 1995, followed through 1999
- Average follow-up 5 years (range 3 days to 11 years)
- Biopsies or ultrasounds annually
- Only 258 events; 75% censoring
- 5332 PSA measurements in total
- No hormone administered during radiotherapy or as salvage

35

A joint longitudinal failure time model

For $i = 1, ..., I$:

$y_i(t)$ the log PSA measurement at time t, s_i the time of recurrence or censoring, d_i the censoring indicator, $y_i = (y_i(t_{i1}), ..., y_i(t_{in_i}))$, θ_i subject-specific random effects, θ a vector of population parameters,

$$p(y_i, s_i, d_i, \theta_i, \theta) = p(y_i|\theta_i, \theta)p(s_i, d_i|\theta_i, \theta)p(\theta_i|\theta)p(\theta),$$

where observations corresponding to different patients i are independent.

36

Longitudinal Component:

$$
\begin{aligned}
y_i(t) &\sim N(\mu_i(t), \sigma^2), \\
\mu_i(t) &= \alpha_i + \beta_i t - \gamma_i \log(t+1), \\
(\alpha_i, \log\beta_i, \log\gamma_i)' &\sim N_3(\mu, D), \\
t_{\text{nadir}} &= \gamma/\beta - 1,
\end{aligned}
$$

for each person, dropped *i*

Survival Component:

$$
\begin{aligned}
\text{hazard}(t) &= \lambda_0(t)\exp[\eta' x(\alpha_i, \beta_i, \gamma_i, t)], \\
\lambda_0(t) &= \Sigma_k \lambda_k I_{(t_k, t_{k+1}]}(t).
\end{aligned}
$$

$x(\cdot)$ *determined by model selection.*

Prior distributions:

$\sigma^2 \sim 1/\sigma^2$

$\mu \sim N_3(0, \text{Diag}(K))$ possible for some elements to depend on covariates X

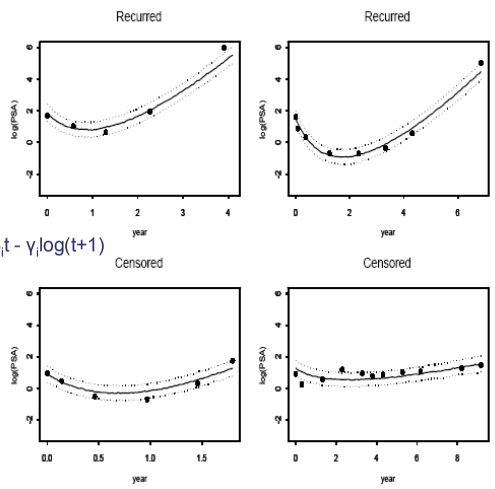
$D \sim \text{InvWish}_{13}(S_0)$ possible to depend on X

$\eta \sim$ flat improper

$\Lambda_k \sim \text{Gamma}(\epsilon, \epsilon)$

Longitudinal fits: individual means, pointwise standard deviations

$\mu_i(t) = \alpha_i + \beta_i t - \gamma_i \log(t+1)$



Model Selection for survival component done in two stages (not yet done for longitudinal component).

- Specify $x(\alpha_i, \beta_i, \gamma_i, t)$ for the Cox subcomponent.
- Do a preliminary fit to 5332 longitudinal PSA measures only. To get BLUP-estimators for random effects, which are then plugged into a Cox model.
- Use SAS Proc PHREG to select covariates for survival which minimize Schwarz's Bayesian Criterion:

$$
\begin{aligned}
\text{SBC} &= -2 \times [\text{maximized log partial likelihood} \\
&\quad - \text{no. of parameters} \times \log \text{sample size}].
\end{aligned}
$$

Covariates Considered

$\mu_i(t) = \alpha_i + \beta_i t - \gamma_i \log(t+1)$

No.	Covariate	Longit. Param.	
1.	Baseline Age		
2.	Stage		
3.	Intercept	α	Drop subscript *i*
4.	Controls rise for large t	β	
5.	Controls initial decline	γ	
6.	Current value	$\mu(t)$	
7.	Derivative	$\mu'(t)$	
8.	Indicator of past nadir	$I(t > t_{\text{nadir}})$	

Optimal Model (SBC=2473.6)*

Longitudinal component: $\mu_i(t) = \alpha_i + \beta_i t - \gamma_i \log(t+1)$
Survival component:

	Predictor	Estimate	Hazard Ratio	p-value
η_1	β	.67	1.95	$< .0001$
η_2	γ	-.26	.77	$< .0001$
η_3	$\mu(t)$	.84	2.33	$< .0001$
η_4	$\mu'(t)$	-.45	.64	.04
η_5	$I(t > t_{\text{nadir}}) \times \mu'(t)$	1.34	3.82	$< .0001$
η_6	$\beta \times \mu(t)$	-.13	.88	$< .0001$

Drop subscript *i*

*Similar models with similar SBC; popular model using only $\mu(t)$ had an SBC value of 2682.3, 208.7 points above the optimal model.

Markov Chain Monte Carlo (MCMC)

Once $x(\cdot)$ is specified full joint model can be coded and fit in C.

- A combination of Gibbs and Metropolis-Hastings steps.
- 5,000 burn-in + 5,000 samples.
- 4 hours on a Dell computer.

Posterior Estimates of Survival Parameters

Most predictors no longer significant. Later on, retained only μ(t), μ´(t) to meet Gelman convergence criteria.

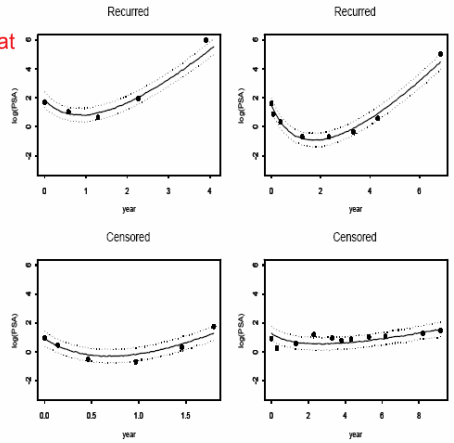
Predictor	log HR Median (HR)	log HR (2.5%,97.5%)	Length
β	.78 (2.18)	(.45,1.12)	.67
γ	-.17 (.84)	(-.43,-.02)	.41
$\mu(t)$	.89 (2.44)	(.67,1.11)	.44
$\mu'(t)$	.36 (1.43)	(-.47,1.93)	2.40
Int	.70 (2.01)	(-.99,1.95)	2.94
$\beta \times \mu(t)$	-.14 (.87)	(-.18,-.11)	.07

*Int=$I(t > t_{\text{nadir}}) \times \mu'(t)$

Posterior estimates of longitudinal parameters have no problems.

Goodness of fit checks show adequate individual fits for longitudinal component.
Check of Normal assumptions, outliers, influence in progress.

Currently finding that t_4 or t_6 for both random effects and measurement error may be more robust.



Prediction for a new patient

t^*: time to recurrence for new patient followed up until time c.

$y^* = \{y^*(t) : t \leq c\}$: all PSA's measured for new patient up until time c.

$\mathcal{D}$: all data from a training set, e.g. MGH

$$p(t^*|y^*, \mathcal{D}, t^* > c) \propto p(t^*, y^*|\mathcal{D}, t^* > c)$$

$$= \int p(t^*, y^*, \theta^*, \theta|\mathcal{D}, t^* > c) d\theta^* d\theta$$

$$= \int p(t^*|\theta^*, \theta, t^* > c) p(y^*|\theta^*, \theta) p(\theta^*|\theta) p(\theta|\mathcal{D}) d\theta^* d\theta$$

Note: θ includes σ^2 and η.

conditional independence

Prediction for a new patient

$$p(t^*|y^*, \mathcal{D}, t^* > c)$$

$$\propto \int p(t^*|\theta^*, \theta, t^* > c) p(y^*|\theta^*, \theta) p(\theta^*|\theta) p(\theta|\mathcal{D}) d\theta^* d\theta$$

Ordinarily, the posterior predictive distribution for t* can be drawn using the inversion algorithm, U ~ Unif[0,1], t* satisfies F(t*)=U.

Posterior draws

Posterior predictive dist. of random effect

However, this term in a joint model prevents that and has to be incorporated into the update of θ*.

Doable if θ* not transformed by log (linear in θ*) AND Normal+Normal case for y*, θ*

Prediction for a new patient

$$p(t^*|y^*, \mathcal{D}, t^* > c)$$

$$\propto \int p(t^*|\theta^*, \theta, t^* > c) p(y^*|\theta^*, \theta) p(\theta^*|\theta) p(\theta|\mathcal{D}) d\theta^* d\theta$$

Method 1: data augmentation

If patient* is included as a censored case in the full MCMC can treat t* as a missing parameter and draw at each iteration using the inversion method.

Can also be done for each censored case in the dataset and works well. What we currently do.

Problem is that it would be too time-consuming for an online calculator since you have to refit the whole dataset.

Prediction for a new patient

$$p(t^*|y^*, \mathcal{D}, t^* > c)$$

$$\propto \int p(t^*|\theta^*, \theta, t^* > c) p(y^*|\theta^*, \theta) p(\theta^*|\theta) p(\theta|\mathcal{D}) d\theta^* d\theta$$

Method 2: mini MCMC chain for t*,θ*, θ

Use the posterior distribution as a prior distribution for θ; works even if it has no analytical form when used as a proposal density in a Metropolis-Hastings update for θ.

Results in a chain for a very informative prior and non-informative likelihood so converges in 500 iterations.

Problem is that it is not feasible to check convergence in the online calculator—potentially outlying input PSA´s.

Prediction for a new patient

$$p(t^*|y^*, \mathcal{D}, t^* > c)$$

$$\propto \int p(t^*|\theta^*, \theta, t^* > c) p(y^*|\theta^*, \theta) p(\theta^*|\theta) p(\theta|\mathcal{D}) d\theta^* d\theta$$

Method 3: Sampling-importance resampling

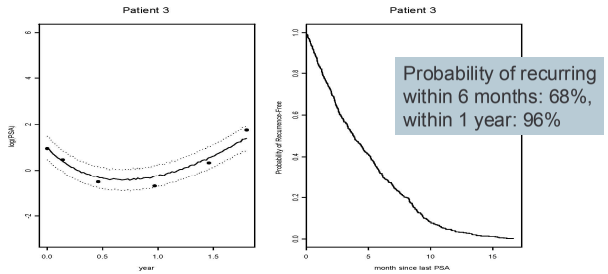
- Draw multiple pairs (θ*,θ) from the last two distributions as an approximate g(θ*, θ).
- Sample (θ*,θ) from the correct distribution by sampling and resampling without replacement from the sample above with weights w(θ*,θ) equal to 3rd to last term (other terms cancel).
- Using correct samples (θ*,θ) draw t* using inversion.

Problem: high-dimensions, unstable weights, may not sample all of posterior.

Example fit and prediction for new patient (Method 1 Data augmentation)

PSA History

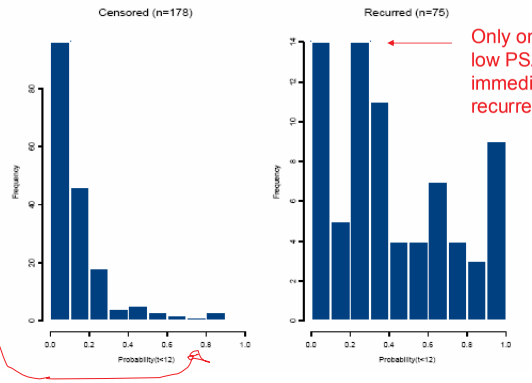
Predicted recurrence-free curve F(t* > t| t > 1.8 years)



Probability of recurring within 6 months: 68%, within 1 year: 96%

Individual Probabilities of Failing in Next Year (253 patients with censor indicators set to 0)

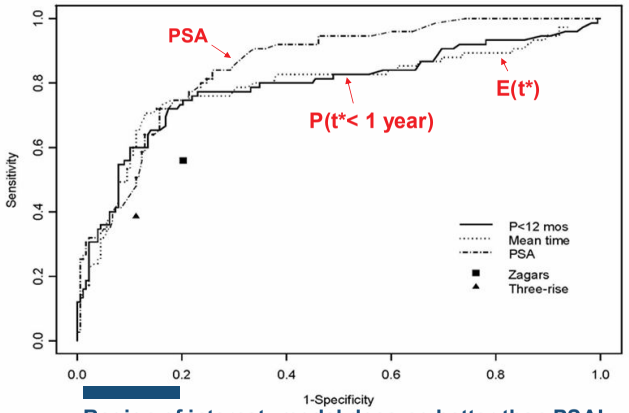
out of 1011 MGH patients



Only one or two low PSA's and immediately recurred.

Steeply rising PSA but did not recur

A ROC curve for 75 recurrences (cases) and 178 censored (controls) at event or censoring time



Region of interest: model does no better than PSA!

Some reasons for bad performance

PSA and Data

--PSA not elevated in all patients that recur and velocity does not add much. We found this for prostate cancer screening in the PCPT.

--Problems in the data: Many of the patients with a recurrence with a low PSA have very short follow-up: within 1 to 2 years. PSA may still be in the post-radiotherapy decline. Perhaps exclude these patients and require longer follow-up for prediction. Clean up outliers.

53

Some reasons for bad performance

Model

--Longitudinal model: A t-distribution with low degrees of freedom is more robust for outlying observations within a person and for outlying trajectories; autocorrelated within-subject error picks up deviations from individual trajectories. Other covariates, e.g. stage, Gleason, age.

--Survival model: Cure fraction (Yu et al, 2004), covariates, accelerated failure time or AUC-maximizing logistic regression instead of Cox proportional hazards model (Zheng Y et al, 2006).

54

Some reasons for bad performance

Evaluation criteria

--ROC for time-varying prediction and time-to-event. Work by Heagerty, Lumley, Zheng (Seattle) for a baseline marker. Our ROC extended this definition but averaged over different lengths of follow-up, which may not be appropriate. Need to explore defintion of a control.

--Positive predictive value: P(recurrence| risk greater c) also needs development in the time-varying/time-to-event case with many of the same issues as above.

55

Nevertheless...

A probability of recurrence is easier for a patient and doctor to understand and make informed decisions for intervention.

Doctors are confused and patients are alarmed by „bouncing PSA“ as referred to by one journal. Risk of recurrence should be smooth, less likely to fluctuate.

Translating risk for online web use will require a detailed understanding of how the risk responds to PSA fluctuation, outliers, adaption for missing covariates, and new data from different institutions/populations.

Work in progress to combine data from MGH, UMCC, RTOG.

56