Estimating and comparing cancer progression risks under varying surveillance protocols: moving beyond the "Tower of Babel"

Jane Lange

March 22, 2017

Acknowledgements

Many thanks to the multiple project contributors.

- Coauthors on paper/feedback on presentation
  - Ruth Etzioni, Roman Gulati, Amy Leonardson
- Consultation
  - Vladimir Minin, Lurdes Inoue
- Data
  - UCSF (Janet Cowan, Peter Carroll)
  - PASS (Daniel Lin, Lisa Newcomb)
  - Toronto (Laurence Klotz, Alexandre Mamedov)
  - JHU (H. Ballentine Carter, Bruce Trock)

Introduction

- Many outcomes of cancer diagnosis and progression are identified at discrete times via diagnostic examination.
  - Prostate cancer progression following primary surgery identified by rising PSA
  - Breast cancer recurrence after diagnosis of in-situ disease identified by surveillance mammography

Surveillant-dependent outcomes

- A continuous-time failure outcome tracked by diagnostic tests or biomarker measurements that occur at discrete times (patient visits).
  - Sensitive to frequency of patient visits
  - Subject to misclassification error

Why are surveillant-dependent outcomes problematic?

- Comparisons of studies, patient populations, or treatment groups with different surveillance schema are confounded by differences in visit frequencies.
- Integrating information across studies is challenging.
- Target of inference may be event that occurs in continuous time, rather than detection of that event.

- Gignac (2008) identified the problem in clinical trials of drugs for preventing bone metastasis in prostate cancer, which are detected by bone scans.

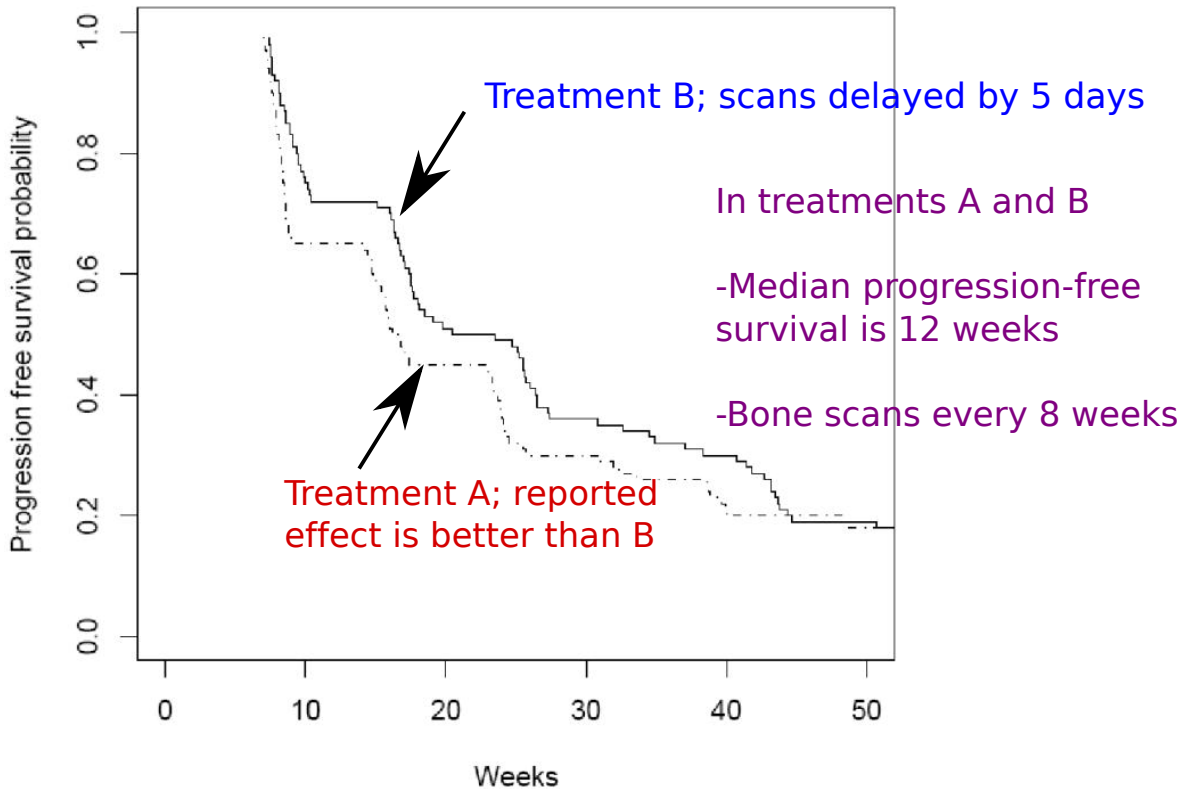


Figure: Simulation study from "Assessing Outcomes in Prostate Cancer Clinical Trials: A 21st Century Tower of Babel", Gignac 2008

Our focus: prostate cancer active surveillance

- AS is now the preferred approach for managing low-risk prostate cancer.
- At diagnosis, men are assigned to series of biopsies
  - referred to treatment if a biopsy detects progression.
- Progression=increase in grade (Gleason score) or tumor volume.
- Many single institution AS studies, but no clinical trials.

Tower of Babel problem for prostate cancer AS studies

Monitoring protocols and triggers for intervention.

	Intervals of surveillance				Triggers for intervention*				
	PSA (mo.)	Exam (mo.)	Mandatory confirmatory biopsy (≤ 1 yr.)	Subsequent biopsies (yrs. from previous)	Gleason score	Positive cores	Max % core with cancer	PSAV	PSADT (yr)
Johns Hopkins#	6	6	Yes	1	>6	>2	>50		
Sunnybrook	3 (x2 yr) then 6		Yes	3–4	Upgrade				< 3†
Göteborg	3–6	3–6	No	2–3	Progression in PSA, grade, or stage (not strictly defined)				
UCSF	3	6	Yes	1–2	>6	>33%	>50		
Royal Marsden	3–4 (x2 yr) then 6	3–4 (x2 yr) then 6	No (≤ 2 yrs.)	2	≥4+3	>50%		>1	
St. Vincent's	3 (x3 yr) then 6	6 (x3 yr) then 12	Yes	1–2, then 3–5	>6	>20%	>8 mm	>0.75	< 3
PRIAS	3–6		Yes	3	>6	>2			< 3
University of Copenhagen	3	3	Yes	Variable	≥4+3	>3			< 3
University of Miami	3–4 (x2 yr) then 6	3–4 (x2 yr) then 6	Yes	1	>6	>2	Increase		

Figure: "Active Surveillance for Prostate Cancer: Contemporary State of Practice", Tosoian 2016

- Different AS studies have different surveillance protocols.

Active Surveillance Scientific Questions

- What is the underlying risk of progression on AS?
- Are different risks that are reported across studies artifactual or real?
- Can we use published results to inform development of surveillance policies?

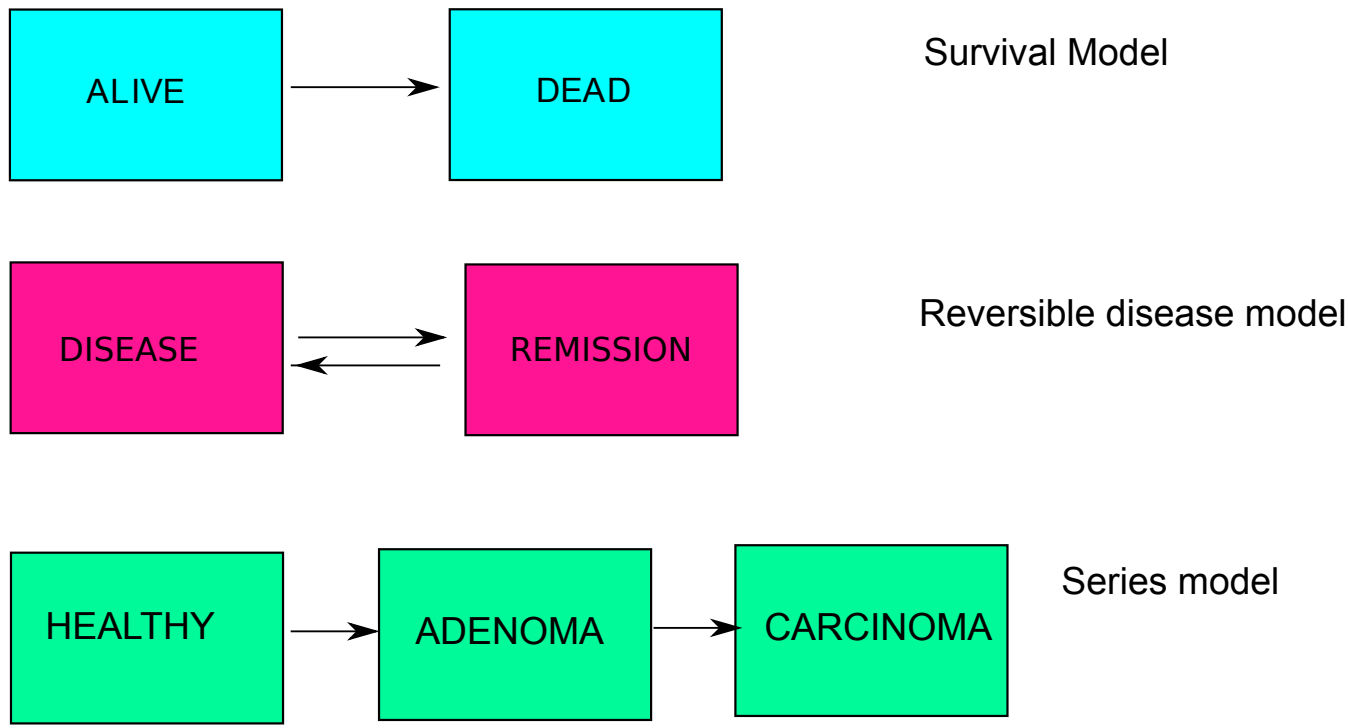
Statistical Issues: How do we move beyond the "Tower of Babel"?

- Require that all studies standardize their followup protocols in terms of visit frequency.
  - Not practical, especially when exposure of interest is the surveillance protocol.
- Acknowledge that clinical observations represent discrete realizations of an underlying continuous time process.
- Use modeling methods that characterize the underlying process, enabling comparison across populations.

Multistate models (MSMs)

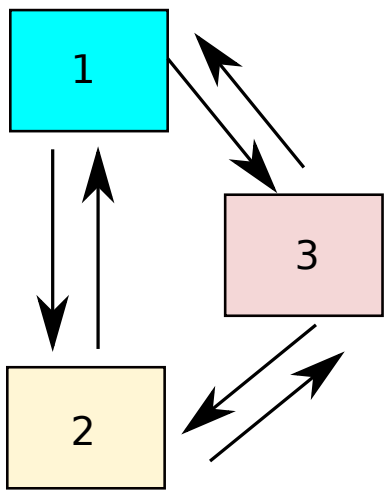
- Our approach is based on assuming that the underlying events of interest are captured by a multistate model.

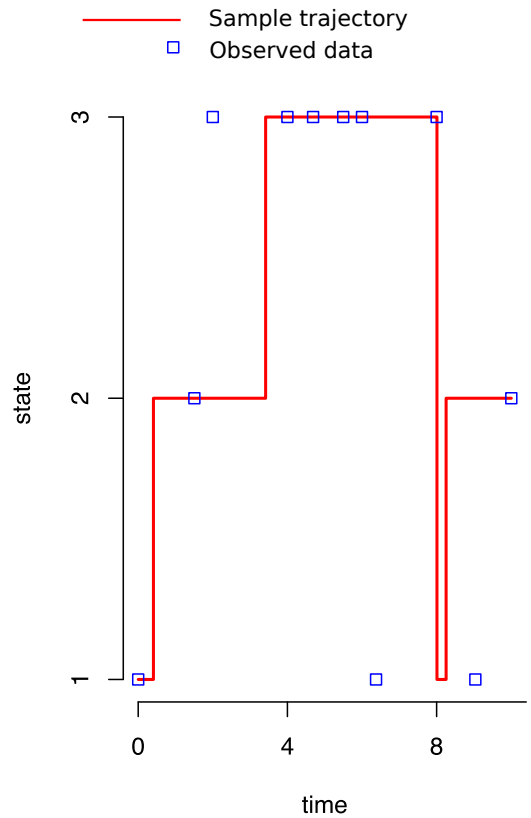
- MSMs characterize an underlying process consisting of transitions over time through a discrete state space.



Observed data

- Example: multistate model with 3 states, discrete observations, and misclassification error.
- Note that multiple transitions can occur between successive observations (not the same as interval censored data).





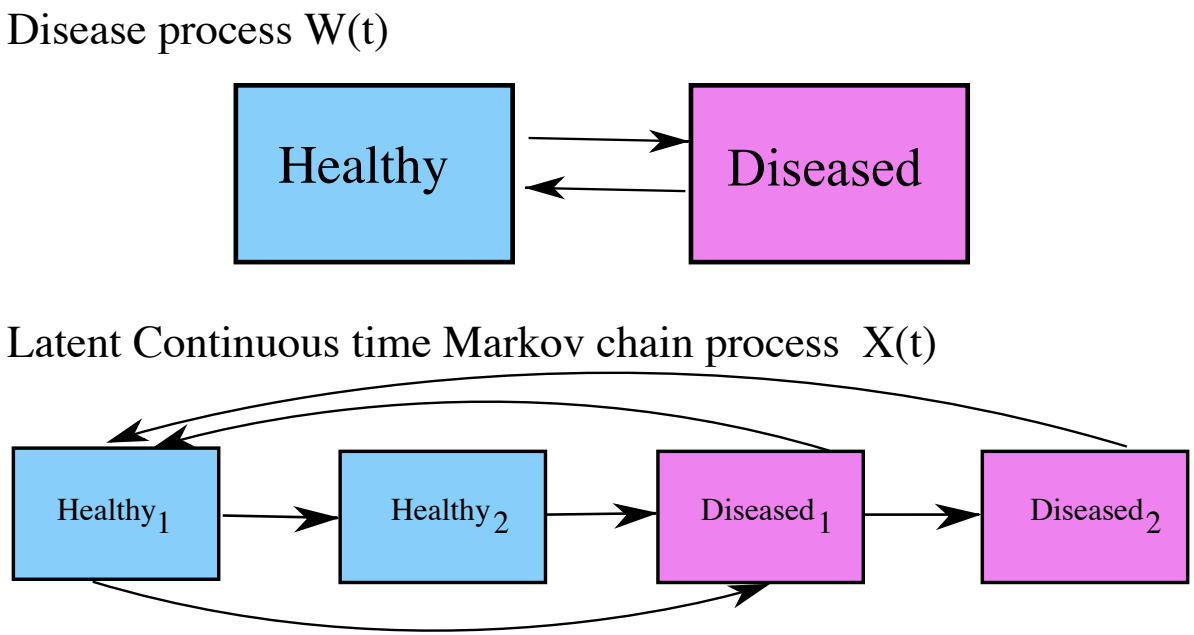
Methodology for discretely observed MSMs

- Fully observed transitions present multiple options for MSMs, both parametric and non-parametric.
- Discretely observed MSMs pose more challenges for estimation, particularly those with reversible transitions.
- We've developed a stochastic modeling approach that is both tractable and flexible.

Underlying disease process model

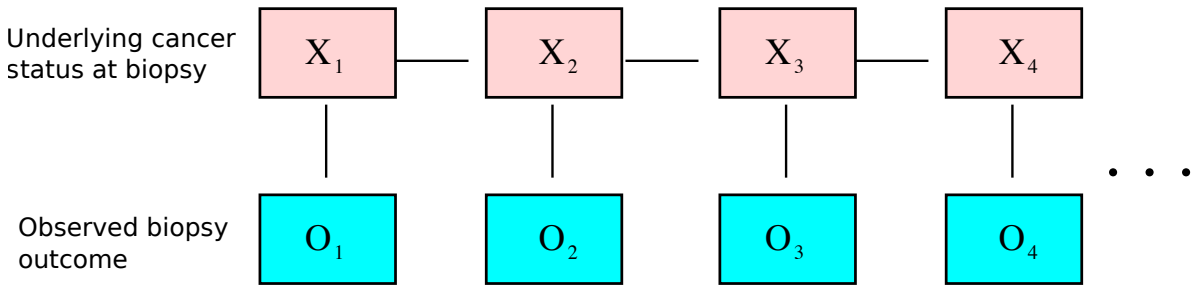
- Typical assumption: the disease process is a time-homogeneous continuous time Markov chain (CTMC)
  - Rates of transitions between states are constant with respect to the time spent in the state.
- This constant hazard assumption is rarely realistic.

Our approach: latent CTMC model



- Disease process $W(t)$ is the trajectory through the states in the model.
- Underlying $W(t)$ is a time-homogeneous CTMC $X(t)$.
- Latent CTMCs permit flexible hazard functions.
- Structured, Coxian transitions prevent over-parameterization.

Incorporating misclassification error: Hidden Markov Models



- $x_1, \ldots, x_k$ are states in underlying process at observation times.
- $o_1 \ldots o_k$ are observed data
- Given conditional independence, observed and underlying data at time t are related via emission probabilities $\boldsymbol{E} = \{e(i,j)\}$

$$e(i,j) = Pr(O_t = j | X_t = i)$$

.

Observed data likelihood

- S is state space for $X(t)$.
- x_0 is initial underlying state at entry.
- $x_1, \ldots, x_n$ are states in underlying hidden process at observation times $t_1, \ldots, t_n$.
- $o_1 \ldots o_n$ are corresponding observed data.
- $P(X_0 = i)$ is initial state probability.
- $P_{[t_i, t_{i+1}]}(x_i, x_{i+1})$ is probability of transitioning between states x_i and x_{i+1} between t_i and t_{i+1}.

The observed data likelihood marginalizes the joint probability of $x_0, x_1, \ldots, x_n$ and the observed data at $t_1, \ldots, t_n$ over $x_1, \ldots, x_n$.

$$P(o_1 \ldots o_n) = \sum_{x_0 \in S} \sum_{x_1 \in S} \ldots \sum_{x_n \in S} Pr(X_0 = i) \prod_{i=0}^{n} P_{[t_i, t_{i+1}]}(x_i, x_{i+1}) \prod_{i=1}^{n} e(x_i, o_i).$$

Estimation

Model parameters

- $\mathbf{\Lambda}$, the transition intensity matrix governing the latent CTMC transition probabilties.
- The vector of initial state probabilities
- The matrix with misclassification probabilities $e(i,j) = Pr(O_t = j | X_t = i)$
- All components may be parameterized with covariates.
- In prior work (Lange 2013), I developed an EM algorithm for parameter estimation.
- Implemented in R package, cthmm (on Rforge)

Prostate Cancer Active Surveillance Application

- PROMISS (Prostate Modeling to Identify Surveillance Strategies)–Fred Hutch R01 (PIs Etzioni, Lin, and Penson)
- Objective is to determine best practices for AS.
- Project integrates data from multiple AS cohorts.
- Models downstream outcomes given different AS protocols.
- Provides recommendations for policy makers.

Four active surveillance cohorts

			Surveillance			Triggers for intervention: Grade	Triggers for intervention: Volume	
Cohort	Years	Enrollement criteria*	PSA	Confirmatory biopsy	Subsequent biopsy intervals	Gleason score	Positive cores	Max % core with cancer
PASS	2008-2013	Low risk	4 mo.	Yes	2yr	>6	>33%	
Toronto	1995-2015	Low risk + select intermediate risk	3 mo.	Yes	4yr	>6		
JHU	1994-2014	Very low risk and low risk (older men)	6 mo.	Yes	1yr	>6	>16%	>50%
UCSF	1990-2015	Low risk + select intermediate risk	3 mo.	Yes	2yr	>6	>33%	>50%

*Risk based on Gleason score, clinical stage, tumor volume, PSA; JHU also included low PSA density criterion.

- Cohorts differ in terms of
  - surveillance frequency
  - inclusion criteria
  - definition of progression (trigger for intervention)

Description of cohorts

- We standardized inclusion criteria (Gleason ≤ 6, age at enrollment < 80, entry 1995+) and the definition of progression on AS to mean an increase in tumor grade (Low grade=Gleason ≤ 6; high grade=Gleason > 6).

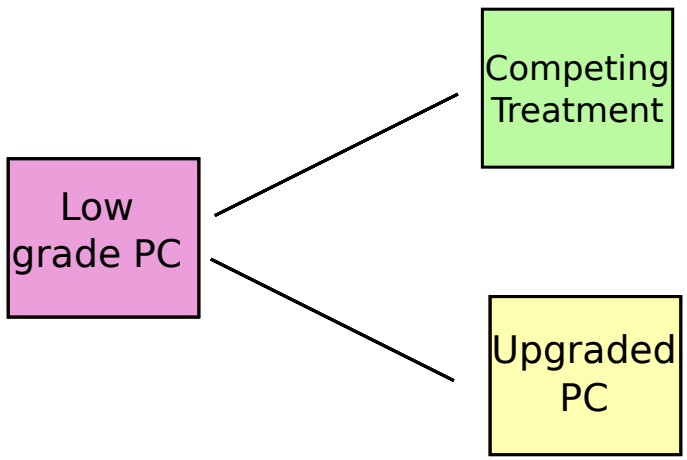
Description of cohorts with common inclusion criteria

	JHU (N=699)	PASS (N=613)	Toronto (N=421)	UCSF (N=843)
Duration of follow-up (years), median (IQR)	4.3 [2.4, 6.5]	2.7[1.5,4.4]	4.8 [2.4,7.7]	3.0 [1.4,5.3]
Number of PSA measumrents, median (IQR)	6 [4,10]	9 [5,14]	12 [7,18]	9 [5,15]
Number of biopsies, median (IQR)	4 [3,6]	1 [1,2]	1 [1,1]	1 [1,2]
Mean number of biopsies/year, median (IQR)	1.1 [.93, 1.3]	.55 [.35,.83]	.26 [.16,.50]	.60 [.35,.91]
Mean number of PSA measurements/year, median(IQR)	1.7 [1.3, 1.8]	3.5 [2.9, 4.1]	2.6 [2.0,3.5]	3.3 [2.4, 4.1]
Age at diagnosis	66 [62,69]	63 [58,67]	65[60,70]	62 [57,66]

Model for prostate cancer upgrading

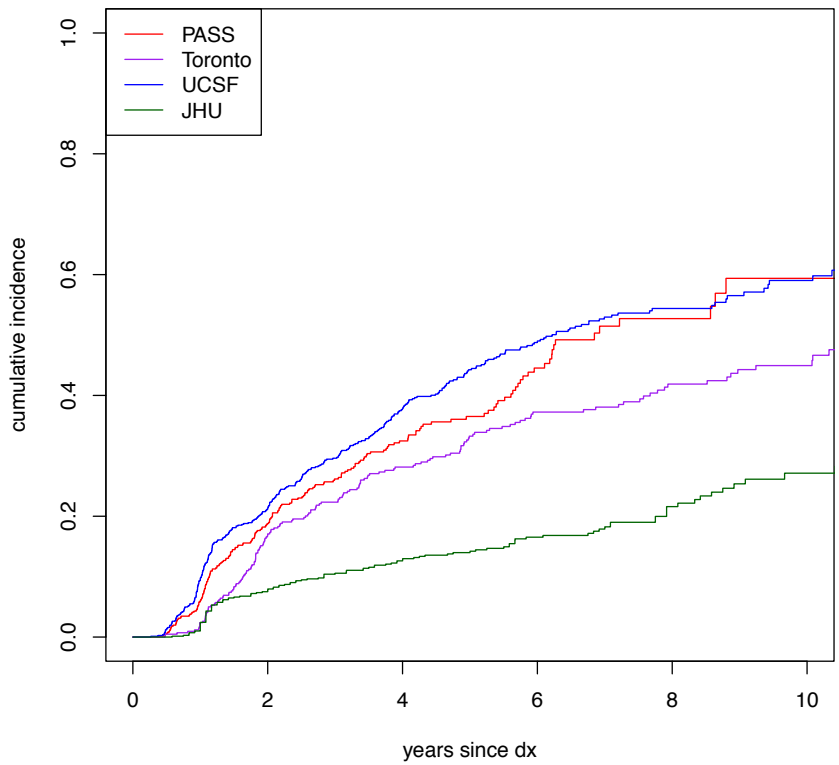
- Treatment prior to upgrade is a competing event that prevents us from observing the natural history of the disease.

- Thus we use a competing risks model framework to characterize the natural history of grade progression during AS.

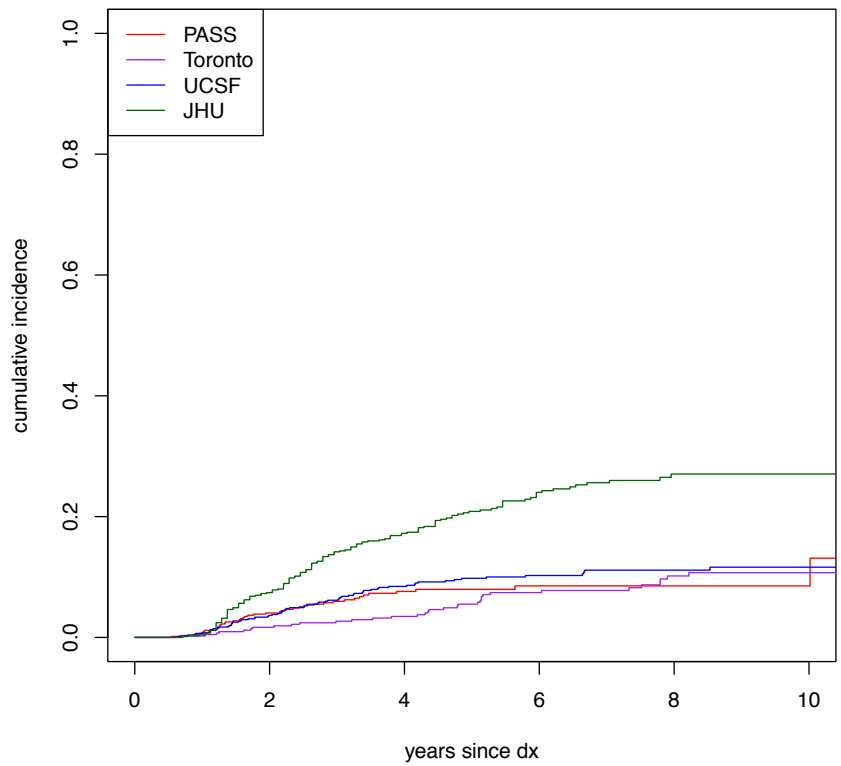


Risk of biopsy upgrading over time and risk of competing treatment

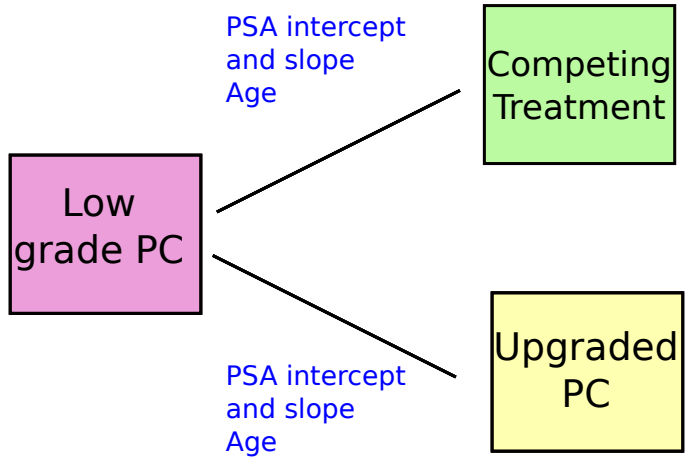
A. Biopsy upgrade



B. Competing treatment



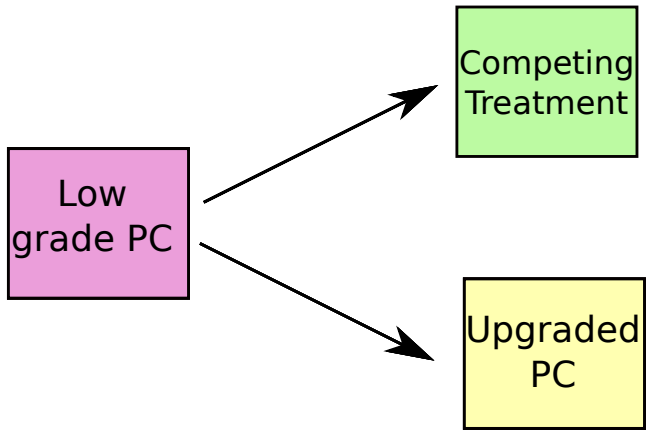
Capturing correlation between upgrade times and times of competing treatment



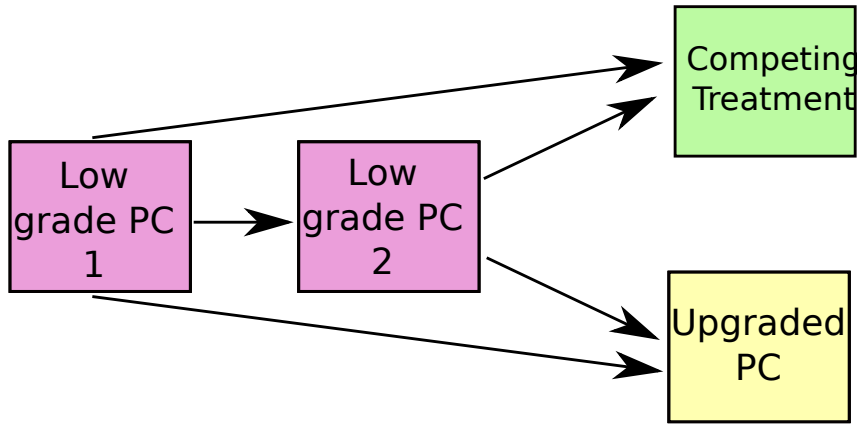
- Times of underlying upgrade and times of competing treatment may be correlated.
- We assume this correlation is fully captured by baseline age, PSA at entry and PSA velocity, and include these as covariates in the transition model.
- Of interest is the distribution of upgrade time in absence of competing treatment–obtained by setting treatment rates to zero.

Latent structures

Standard CTMC



Latent CTMC



- Additional latent states in CTMC model add flexibility to the sojourn time distribution.
- Considered models with 1 and 2 additional latent states for each cohort.
- Selected best fitting model for each cohort via Bayesian information criterion.

Misclassification component

- Biopsies may misclassify tumor grade (have false negatives or false positives)
- For these analyses, we assume 100% specificity (low grade cancers do not yield positive (high grade) biopsies).
- We considered models with imperfect sensitivity (biopsies may not detect high grade disease).
- Empirical tests with models fit with varying sensitivity suggested these data are not able to estimate it reliably.
- Therefore we fixed biopsy sensitivity at 75%, 90%, and 60%, 100% and estimated disease progression parameters.

Specific analysis goals

- Study differences in times of underlying progression in absence of competing treatment
  - by cohort
  - under different assumptions about biopsy sensitivity

Results: model selection for each cohort

Table: Model selection using Bayesian information criterion assuming 100% biopsy sensitivity

PASS

Model	Log likelihood	N params	N sample	BIC
CTMC	-688.1	8	613	1398.5
Latent-2	-677.1	11	613	1384.9
Latent-3	-674.8	14	613	1388.6

JHU

Model	Log likelihood	N params	N sample	BIC
CTMC	-1086.4	8	699	2195.6
Latent-2	-1069.6	11	699	2170.5
Latent-3	-1055.8	14	699	2151.4

Results: model selection for each cohort

Table: Model selection using Bayesian information criterion assuming 100% biopsy sensitivity

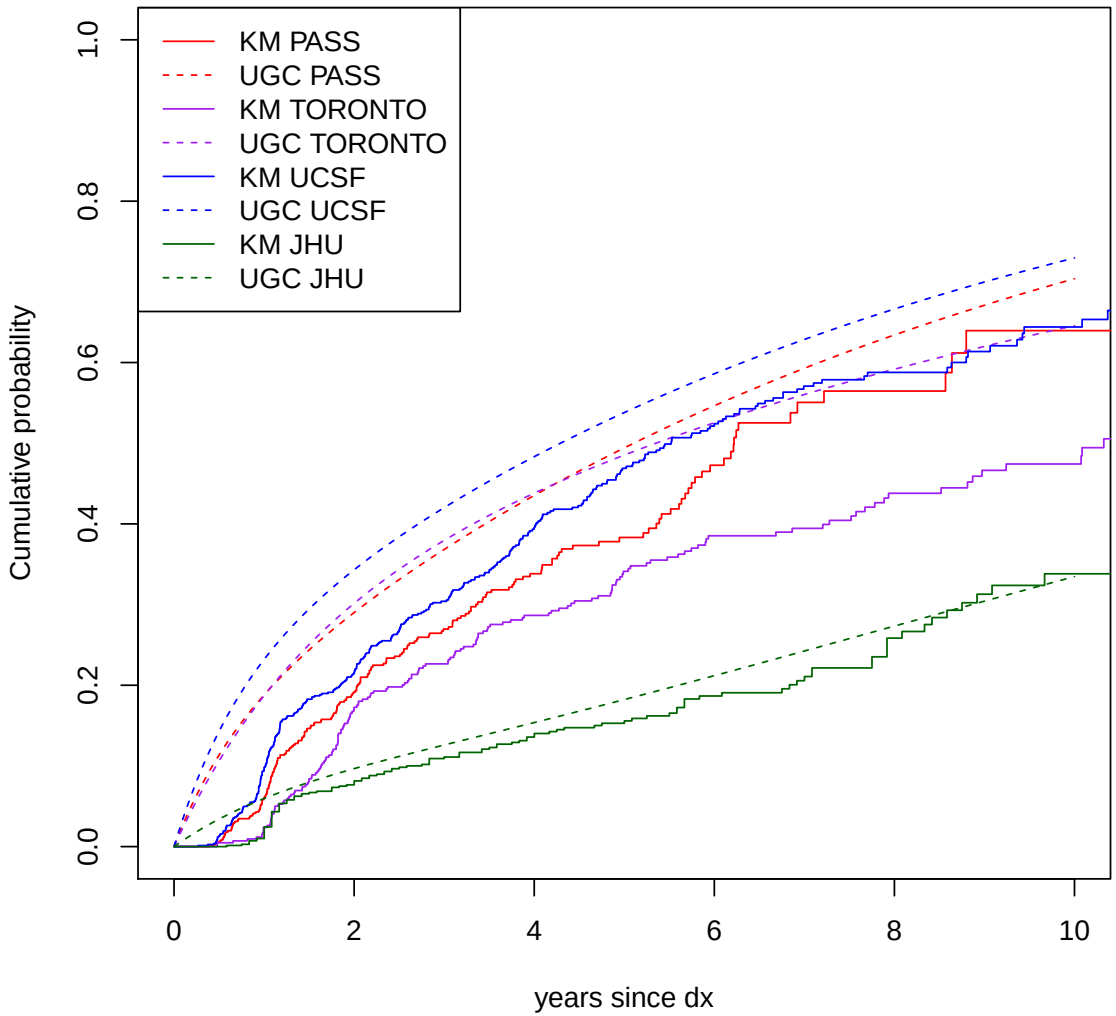
Toronto

Model	Log likelihood	N params	N sample	BIC
CTMC	-495.3	8	421	1011.6
Latent-2	-480.4	11	421	989.7
Latent-3	-479.2	14	421	995.1

UCSF

Model	Log likelihood	N params	N sample	BIC
CTMC	-1149.4	8	764	2321.8
Latent-2	-1122.0	11	764	2275.7
Latent-3	-1122.0	14	764	2284.4

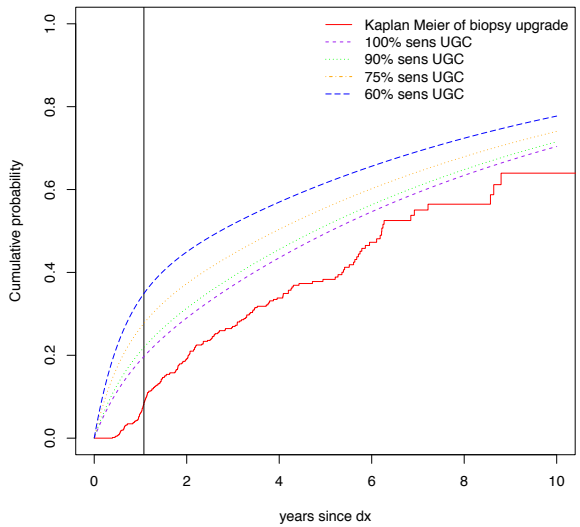
Results: Predicted distributions of times of upgrading absence of treatment across cohorts



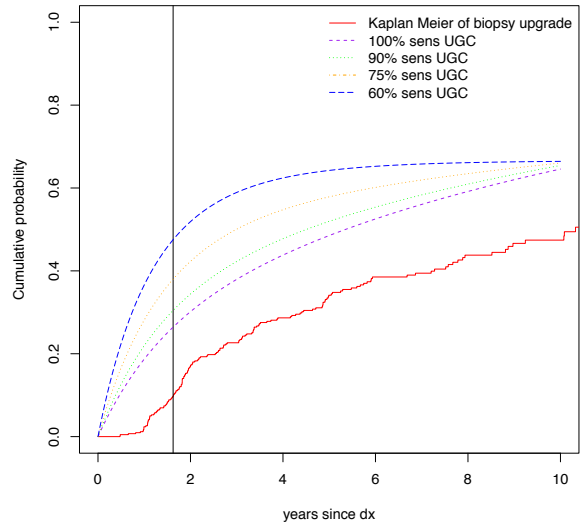
- All cohorts have distributions of underlying upgrade shifted left of diagnosis time curves, but the degree varies depending on biopsy frequency.
- PASS and TORONTO may be pretty similar in terms of risk (p=.31 for difference in combined analysis)
- JHU has considerably lower risk.
- UCSF has somewhat higher risk.

Results: effects of biopsy sensitivity on predicted distributions of times of upgrading in absence of treatment

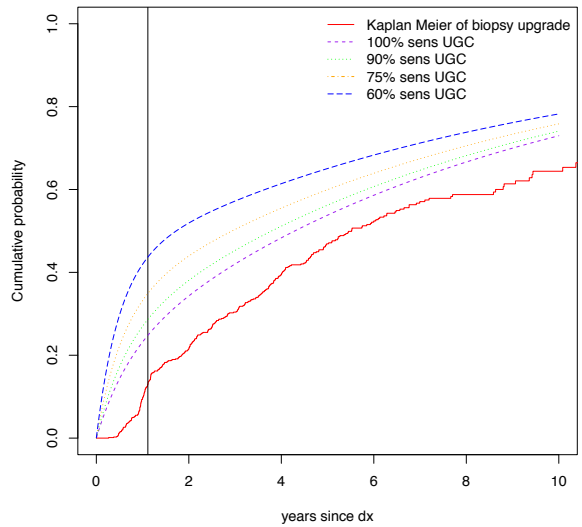
A. PASS



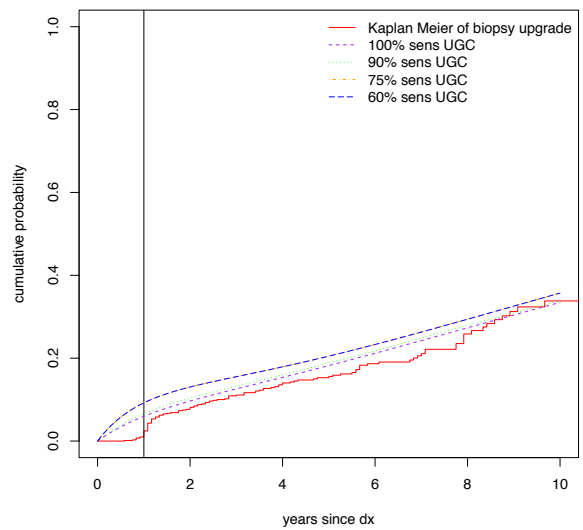
B. Toronto



C. UCSF



D. JHU



Clinical implications of AS study

- On the underlying upgrade scale, we conclude PASS, Toronto, and UCSF may be reasonably comparable in risk of grade progression, but JHU is still considerably lower risk.
  - Partial explanation: JHU has stricter PSA density criterion.
  - Using any one cohort to make absolute risk predictions regarding grade change may be problematic.
- Biopsy sensitivity affects the projections of distribution upgrade times; lower biopsy sensitivity suggests that many enter the cohort with higher grade disease rather than progressing over time.
  - Assumptions about sensitivity may suggest different biopsy screening strategies.
- We plan to use these models to simulate downstream outcomes with different surveillance schedules.

Overall summary

- The latent CTMC approach avoids the "Tower of Babel" problem of comparing surveillant dependent outcomes by treating such data as discrete observations of an underlying continuous time process.
- The latent parameterizations enables flexible sojourn time distributions, but retains analytic tractability of standard CTMCs.
- The models enable dynamic prediction of a patient's underlying status based on his prior history of testing results.
- While there are other methods for interval censored data and panel data, this methods applies flexibly to a variety of scenarios.

Limitations

- Models still make parametric assumptions about upgrading distribution, although latent structure provide added flexibility.
- It is not always possible to simultaneously estimate misclassification probabilities.
- Latent parameters not always fully identifiable (but are not themselves target of inference).
- Complexity of latent structure is constrained by the frequency of observations.
- This method conditions on visit times, and assumes they are non-informative. Lange (2015) considered an extension to informative visit times, useful when patient initiate visits based on symptoms.

Questions?

Additional Slides

Initial parameter estimation with latent and standard CTMC