Annual Lecture (SSA NSW) 2015

Malcolm Hudson

Department of Statistics, Faculty of Science,
Macquarie University
and
NHMRC Clinical Trials Centre,
University of Sydney

November 2015

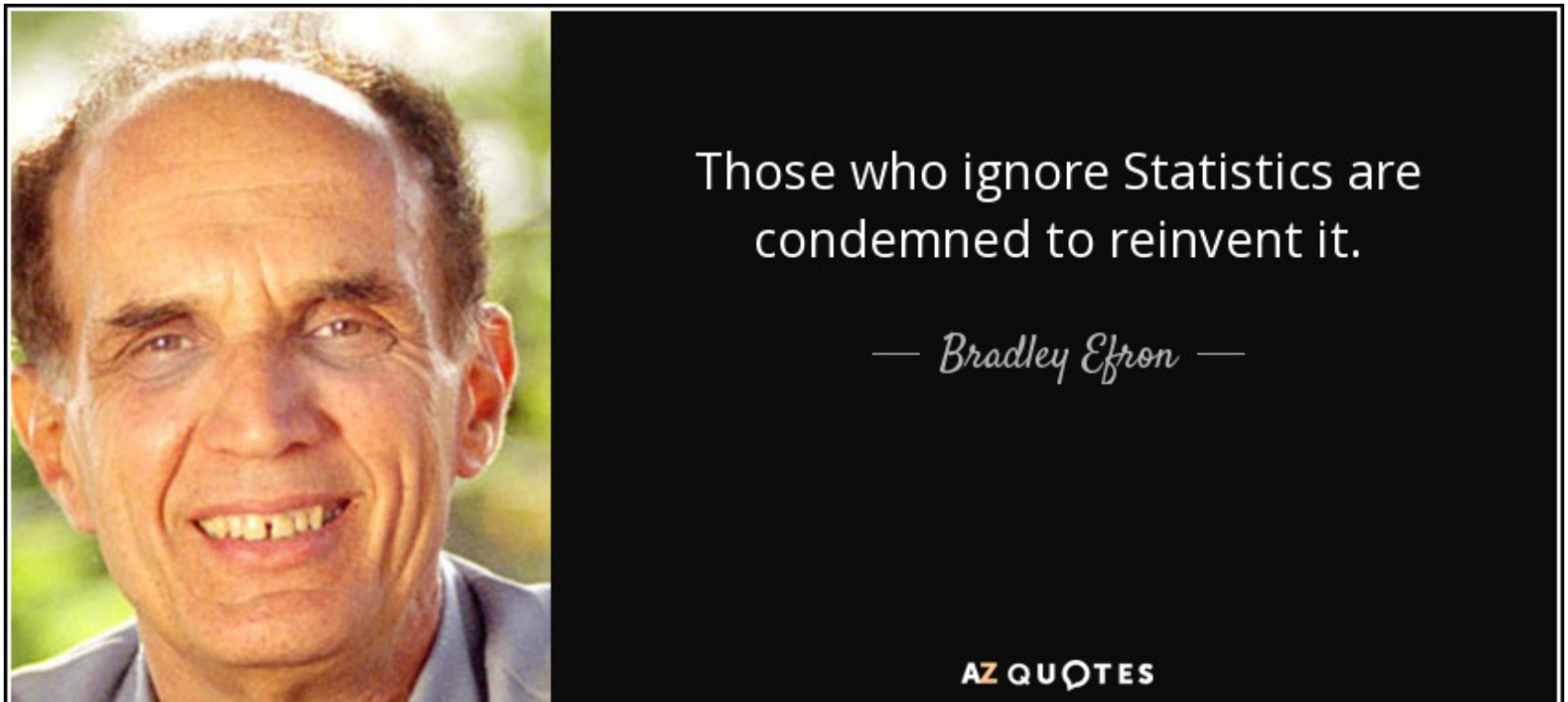
For your consideration

Hans Rosling

You use statistics all the time – for the weather forecast or calculating your income.
And whether you're talking about it with other academics or in the pub, these are topics that matter to people.

Brad Efron

Statistics has been the most successful information science.
Those who ignore statistics are condemned to reinvent it.



James B. Douglas, D.Sc. (UNSW Honoris Causa)

Jim Douglas is one of the founders of the field of Statistics in Australia, known internationally for his pioneering contributions to the discipline and the profession.

Beginning at the University of NSW even before its days as the NSW University of Technology, Mr Douglas was appointed Lecturer in Mathematics in November 1947, retiring as Associate Professor of Mathematical Statistics in 1983. However, his "retirement" never really happened: over an unbroken period of 57 years, and in his 81st year [then], he still attend[ed] the University, teaching, mentoring students, and maintaining his scholarly interest in the discipline and the profession.

Introduction

- Thank you for the invitation to speak tonight.
- Invokes memories of student days
  - Jim Douglas and Charles McGilchrist, teachers and exemplars
  - fellow students and teachers at UNSW (Mathematics) and Stanford University
- Acknowledge many close colleagues and fellow researchers over years since

My title:

Factors affecting treatment recurrence and death –
a case study with longitudinal hospital retreatment records

- Collaboration with Profs. M. Barton, UNSW and G Delaney, SWSHS
- Macquarie University PhD thesis (2011) of Dr Zhixin Luo
- Topic involves counting repeat visits after the first

Common research interest

Estimating under-counting in RT services in different cancers

- Liverpool Hospital cancer centre database
  - Prof Barton's innovative use of 'Treeage' software

Use of Treeage at CTC in medical decision making

- interest in developing capability for Markov modelling

Grant application

- goal: health services planning (administrative)
  - longitudinal modelling (UK studies cross-sectional)
- supported appointment of statistician (0.2) and data manager (0.2)
- deliverable: counting
  - John Graunt's medieval analysis of causes of death in the plague

Approach tonight

Methods for counting recurrent events

- compare several approaches estimating *mean numbers*

Case study of cancer retreatments in LMCTC

- Liverpool and Macarthur Cancer Therapy Centres (LMCTC) hospital database[1]
- our goal, estimation of mean numbers of retreatments
  - for Breast and Lung cancer patients
  - identify effects of factors on the mean number (e.g. trends)

Survival with intermediate events

- recurrent events ('retreatments') ended by a terminal event ('death')
- focus on the retreatment process rather than survival
  - do we need dates of death?

[1] acknowledge project DM: Stuart Allen

Specifics

Patterns of Retreatment by Radiotherapy in LMCTC

- 6200 cancer patients were followed after initial RT in the period 1997-2006
  - follow-up to March, 2011 (from 4- years to 12+ years f/u)
  - 1453 retreatments
  - 3066 deaths
  - 3127 remained alive at study end
- event outcomes retreatments and deaths
- supplemented by NSW State Cancer Registry mortality data
- descriptive analysis[2] available

[2] Barton et al, Clinical Oncology 23 (2011) 10–18

Descriptive approaches to recurrent events

Institutional comparisons[3]

- LMCTC, Royal Brisbane and Women's Hospital, Radiotherapeutic Institution Friesland (RIF), the Netherlands
- over the period 1991 to 2009
- patient outcomes
  - 2 and 5 year survival
  - proportion retreated
  - mean number of RTs
  - by gender, year of initial treatment, baseline age and tumour type
  - Poisson regression of number of RTs during first 2 years
- among patients with observed retreatment episodes
  - more than a single interval per patient
  - e.g. 6193 pts, 7646 intervals
  - characteristics of treatment in these episodes
    - retreatment number
    - retreatment site (primary, bone, etc)
    - time since last radiotherapy

[3] Barton et al, Clinical Oncology (2014)

Analysis options

First-event analysis

- Complication-free survival time (i.e. time to first event)

Competing risk analysis:

- model cause-specific hazard of time to first retreatment
- directed at outcome of interest, censor after others (death)

Multiple recurrence analysis

- mean numbers of events

Issues

- explain variability in mean numbers
  - fixed follow-up *or* adjust for length of follow-up
- association between recurring events and death


analysis
time
risks
events
censoring
failure
model
hazards
cumulative
estimation
incidence
multiple
event
dependent
counting
method
application
function
association
per
pwp
rprentice
hamiltonbased
cancer
lin
coxs
cook
geskus
biostatistics
curves
lee
lim
survival
left
june
estimators
gray
kelly
fine
aalen
cox
clinical
several
american
dmcneil
biometrics
jklein
table
correlation

First event (Survival Analysis)

- Distinguish cumulative incidence, cumulative mean function and prevalence:
  - cumulative incidence, $N(t)$, is the cumulative count as events occur
  - survival analysis - at most 1 event per record - $N(t)$ 0 or 1 step-up if event
  - prevalence $F(t) = P(T \leq t) = 1 - \exp(-H(t))$ aka CIF
  - links prevalence to cumulative hazard
    - constant hazard $h(t) = \lambda$ for $t > 0$
    - cumhaz $H(t) = \lambda t$
    - event prevalence $F(t) = 1 - \exp(-\lambda t)$
  - mean function $CMF(t) = E(N(t))$
  - at most 1 first event: $CMF(t) = F(t)$, event prevalence

Sampling records of recurring events

- For randomly sampled records $\{N_i(t) : N_i(t) \in \{0, 1\}; i \in 1, \ldots, n\}$
  - event prevalence $F(t) = P(T \leq t)$ is the proportion of records with (first) event by time t
  - contrast $CMF(t)$, the expected average $(\sum_1^n N_i(t)/n)$ number of events recorded by time t
- CMF permits comparisons (of events per-person)
  - despite long follow-up (1999 cohort) and short (2006 cohort)
  - is medical practice changing?
- Nelson-Aalen's (N-A) estimator of cumulative *hazard* $H(t)$
  - each individual remains at risk *after* an event (until end of follow-up)

Models for recurring events

- Assume independent censoring (endpoint indpendent of events)
- Anderson-Gill (AG) model for an individual record postulates events recurring with intensity $h(t)$
  - same $H(t)$ for all subjects, N-A estimator
  - event hazard is independent of `earlier` event times
  - independent of the number of `earlier` events
- Prentice-Williams-Petersons (PWP) model
  - hazard of next event may depend on previous event history
  - H(t) may differ depending on event number (i.e. *episode*)
  - gap-time models restart clock after each event

Practicalities

- Recurrent events
- R packages:
  - survival
  - cmprsk, Epi, ...
- Counting Example: 2 subjects with 2 and 3 events
  - events at times 1, 2 years, with an additional event at 0.5 years for subject 2.
  - Calculate the CIF. Note the effect of adding a third interval, censored at 0.5 years.
- Method:
  - 'survival' package using start,stop times
  - time origin is at study entry

Counting example (plot commands)

- two patients with 2, 3 recurrences respectively

```
> tdata2
```

```
  id  t2  t1 status enum
1  1 1.0 0.0      1    1
2  1 2.0 1.0      1    2
3  1 3.0 2.0      0    0
4  2 0.5 0.0      1    1
5  2 1.0 0.5      1    2
6  2 2.0 1.0      1    3
7  2 3.0 2.0      0    0
```

- N.B. patients have same follow-up time - 3 years
- count using AG model analysis of intervals

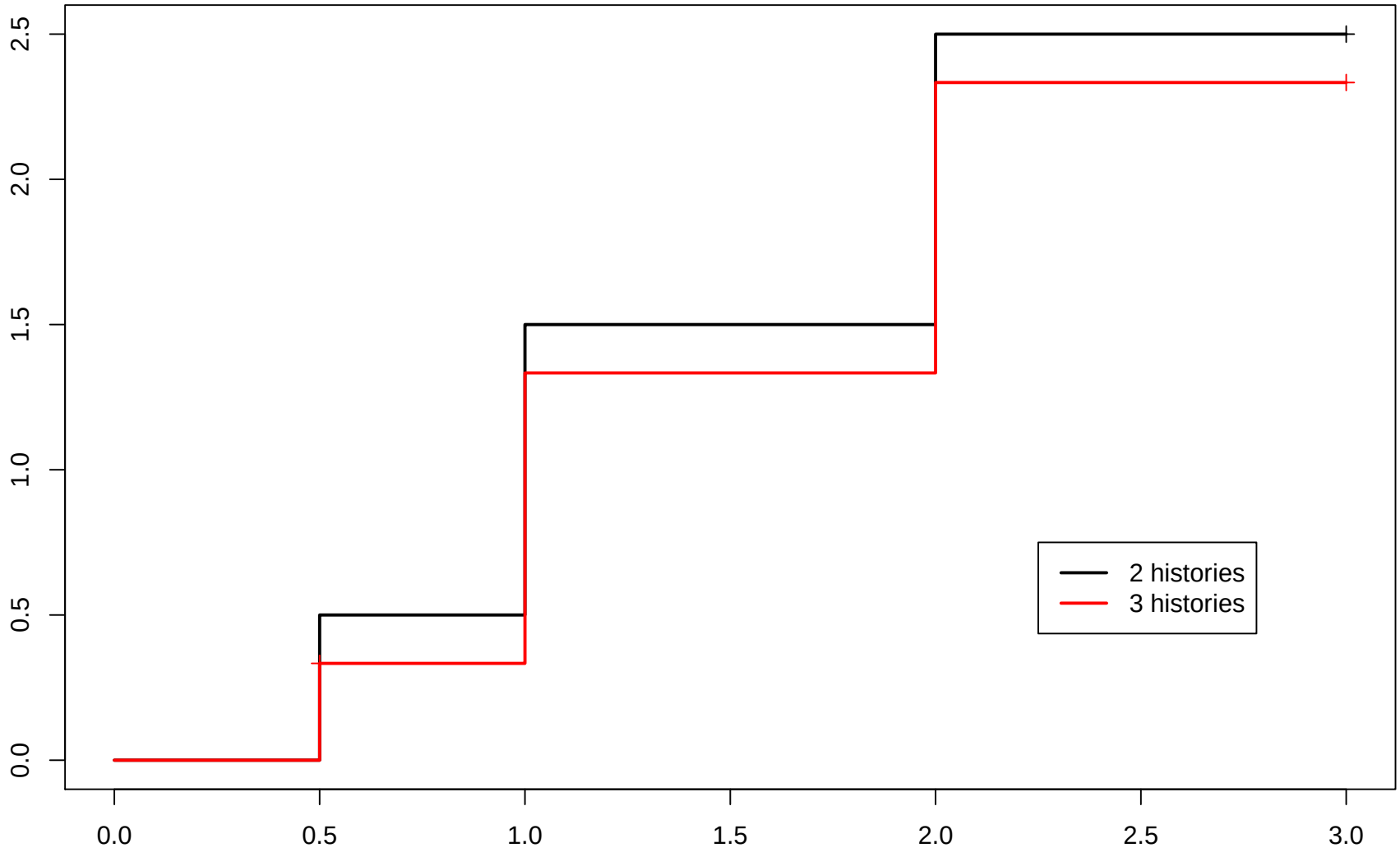
```
> require(survival)
> fit.sfit2 <-coxph(Surv(t1,t2,status)~1, data=tdata2, ties="breslow")
> plot(survfit(fit.sfit2), fun='cumhaz', conf=F, lwd=2)
```

Counting example (extra case)

- Extra case has a single interval no events, censored at 6 months

```
> censorcase <- c(id = 3, t2 = 0.5, t1 = 0, status = 0, enum = 0
> tdata3 <- rbind(tdata2, censorcase)
```

Counting example (`cumhaz` plot)



About 1,540,000 results (0.51 seconds)

Scholarly articles for **event history data age period cohort survival**

... for displaying individual fertility **data** and **cohort survival** ... - Carey - Cited by 44
... Risk for Worsening Cardiovascular **Event-Free Survival** ... - Shaw - Cited by 379
Event history analysis - Yamaguchi - Cited by 1627

[PPT] MSc ASR: SR04 Lecture 1, Introductory data analysis (...

www.tlrp.org/rcbn/capacity/Activities/Themes/Secondary/Lambert.ppt

Event history datasets. 4. **Cohort** studies. 3. ... Distinguish **age**, **period** and **cohort** effects; Causality and residual heterogeneity ..... Eg 5.1 : Kaplan-Meir **survival**.

Wiley: Survival and Event History Analysis - Niels Keiding ...

www.wiley.com › ... › Applied Probability & Statistics › Survival Analysis

Aging Models. Age-of-onset Estimation. **Age-Period-Cohort** Analysis. Bayesian Approaches to Cure Rate Models. Bayesian Model Selection in **Survival** Analysis ...

[PPT] MSc ASR: SR04 Lecture 1, Introductory data ... - ReStore

www.restore.ac.uk/Longitudinal/talks/pl_5approaches.ppt

Event history datasets ... Introducing quantitative longitudinal **data** analysis ... Distinguish **age**, **period** and **cohort** effects; Career trajectories / life course sequences ..... Descriptive: compare times to **event** by different groups (eg **survival** plots) ...

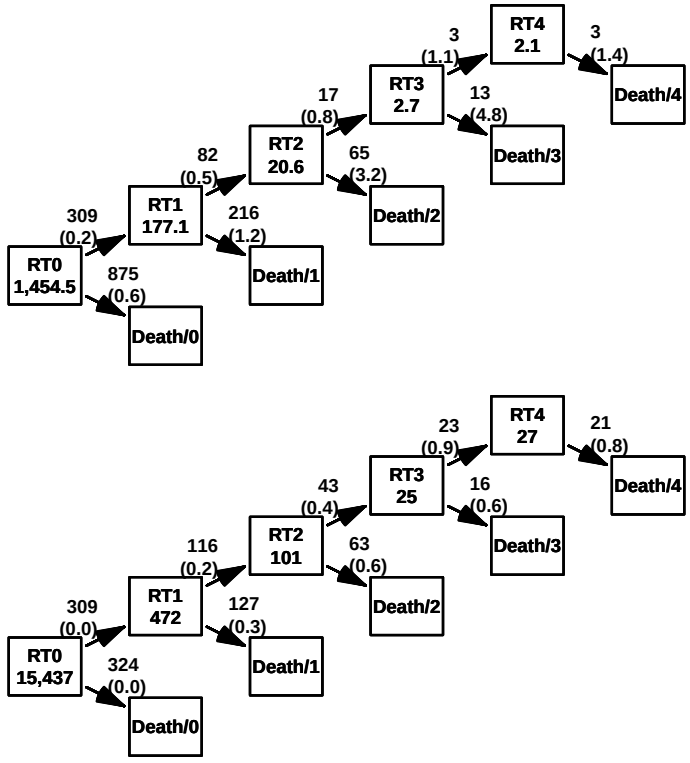
[PDF] Event History Models for Life Course Analysis - New Yo...

sociology.as.nyu.edu/docs/IO/320/chap22.pdf

by LL WU - Cited by 52 - Related articles

spent in selected life course statuses) has changed for successive **cohorts** ... of time, for example, **age**, **duration** in various statuses, and historical ... In this chapter, I review methods relevant to life course research when **event history data**— that is ..... heavy use of nonparametric estimators of quantities such as the **survival** ...

MSM diagram: 1.Lung, 2.Breast cancers



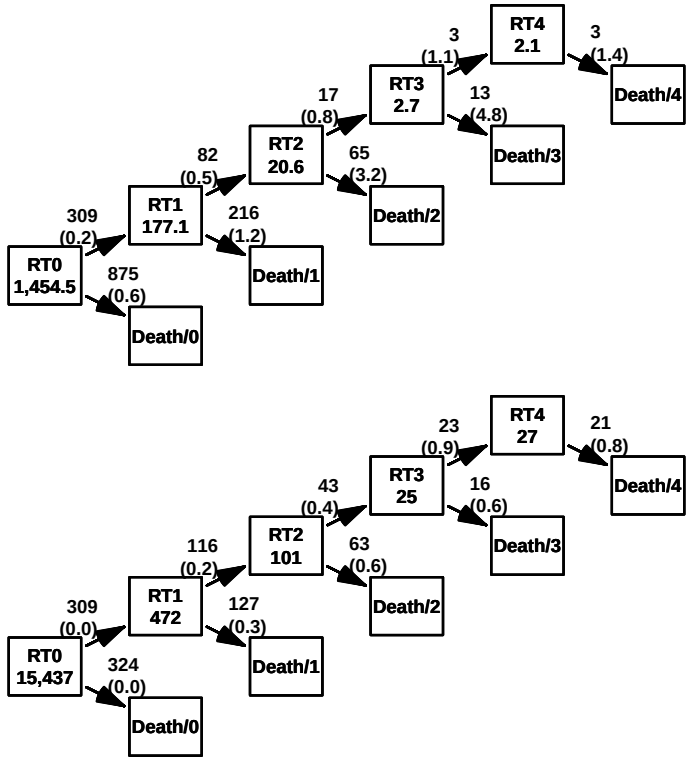
- State transition diagram and statistics.
  - Numbers of transitions from each state
  - [in box] person years (p.y.'s) at risk

Example: Lung cancer (top)

$\Rightarrow$ ratios *observed deaths to retreatments* remain around **3 to 1**

$\Rightarrow$ *event rates p.a.* rise from **1 in 10** after RT0 to **1** after RT2+

MSM diagram: 1.Lung, 2.Breast cancers



- State transition diagram and statistics.
  - Numbers of transitions from each state
  - [in box] person years (p.y.'s) at risk

Example: Lung cancer (top)

$\Rightarrow$ ratios *observed deaths to retreatments* remain around **3 to 1**

$\Rightarrow$ *event rates p.a.* rise from **1 in 10** after RT0 to **1** after RT2+

Mean Estimation defined by Events and follow-up

Method	Events data	Intervals	censor time	Reference
OCI	retreatments	RT0-RTk, RT0-eof	censor at eof	KP, ZM
C-L	retreatments and death	RT0-RTk, RT0-death	death (or eof)	CL 4.1
Pepe2	retreatments and death	RT0-RTk, RT0-death	death (or eof)	CL 4.2
N-A	death/RT (composite)	RT0 -'event'	eof	*
A-J	retreatments and death	inter-events-death	death (or eof)	CL 4.3

eof = End date of follow-up
RT0 = Date of initial radiotherapy
RTk = Date of k-th retreatment

CL = Cook, Lawless et al, JASA, 2009
KP = Kalbfleisch & Prentice (text)
ZM = Zhang-Salomons and Mackillop, Comp.Meth.Prog.Biomed., 2008

Other cause ignored (OCI)

```
> oci <- subset(d1mcut[,c("id","dob","dst2","type",
+                         "episode","t1","t2","status")],
+               type=="Breast")
> head(oci)
```

```
        id         dob        dst2   type episode t1   t2 status
2  1011145 1938-01-27 1997-05-05 Breast       1  0 5062 censor
3  1011148 1936-03-25 1997-05-06 Breast       1  0 5061 censor
5  1011157 1944-03-22 1997-05-05 Breast       1  0 5062 censor
9  1011159 1951-03-19 1997-05-05 Breast       1  0 5062 censor
15 1011162 1941-01-27 1997-05-07 Breast       1  0   49     RT
16 1011162 1941-01-27 1997-05-07 Breast       2 49 5060 censor
```

```
> with(oci, table(status))
```

```
status
censor     RT
  2271    518
```

A proportional hazards model for the subdistribution of a competing risk
Fine, Jason P;Gray, Robert J
Journal of the American Statistical Association; Jun 1999; 94, 446; ProQuest Central
pg. 496

3.2 Censoring Complete Data

In smartly designed clinical trials, censoring results only from administrative loss-to-follow up; that is, patients have not failed by the time the data are analyzed. Under this condition, the potential censoring time is always observed, even on individuals who die prior to the time of analysis. We call these data *censoring complete*. We redefine the risk set for the ith individual to include the censoring information

$$R_i = \{j : (C_j \wedge T_j \geq T_i) \cup (T_j \leq T_i \cap \varepsilon_j \neq 1 \cap C_j \geq T_i)\},$$

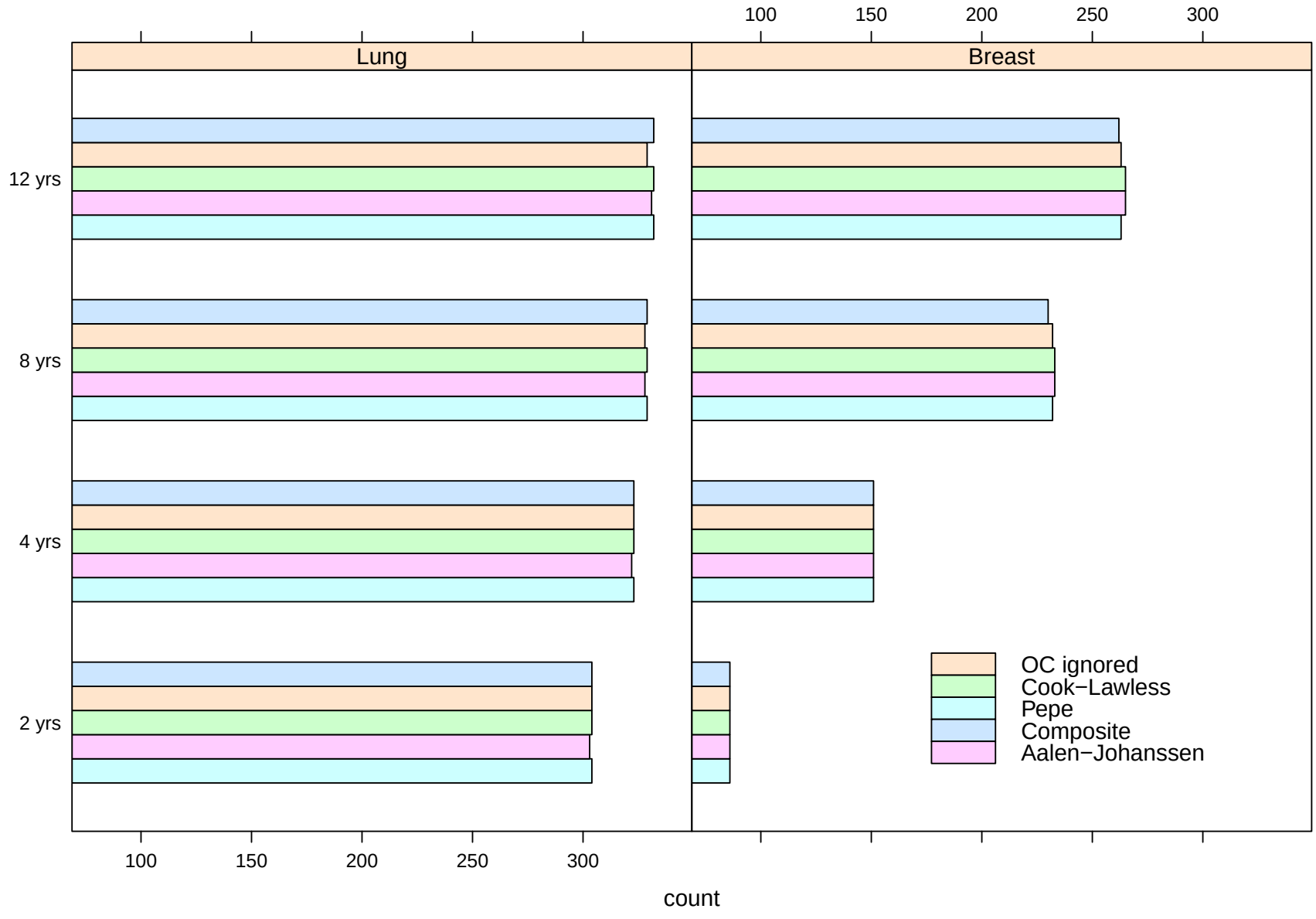
where $i \wedge j$ denotes $\min(i, j)$. In our hypothetical cohort, an individual with $\varepsilon \neq 1$ is still "at risk" for failure from the cause of interest until time C, when $T < C$. If (T, ε) and C are conditionally independent given the covariate, then the "crude" subdistribution hazard function with censoring-complete data, $\lambda_{1*}\{t; \mathbf{Z}\}$, is equivalent to the "net" subdistribution hazard function with complete data, $\lambda_1\{t; \mathbf{Z}\}$. This

Results – Subgroups

Cumulative mean numbers: retreatments per 1000 RT patients

LUNG CANCER				
Method	Year			
	2	4	8	12
OCI	304	323	320	332
C-L	303	322	328	331
Pepe	304	323	329	332
Composite	304	323	328	329
A-J	304	323	329	332
BREAST CANCER				
Method	Year			
	2	4	8	12
OCI	86	151	232	263
C-L	86	151	233	265
Pepe	86	151	233	265
Composite	86	151	232	263
A-J	89	151	230	262

Above Table as graph



Findings for mean estimation

- Concerning methods
  - Since all patients experience at least 4 years follow up, all methods provide the same mean number of events up to time $t = 4$.
  - Thereafter, some estimates differ.
  - But differences are small, for mean retreatments to $t = 8$ and $t = 12$ years
  - even for Breast Cancer, with continuing incidence of new retreatments to 12+ years.
- **Is follow-up of *deaths* necessary in this context?**

Theorem: Multiple Cohorts

Assume longitudinal data is available on first recurrence

- homogeneous patient cohorts $1, 2, \ldots, I$:
- a common entry date in each cohort;
- a common exit date (other than death);
- cohort i has a pre-specified length of follow up τ_i;
- this **administrative censoring** is the only source of censoring.

The empirical CIF of time from entry to *first* recurrence, allowing for death as competing cause, is the empirical CIF of first recurrences *alone*, and so is independent of times of death.

Theorem Consequences (Corollaries)

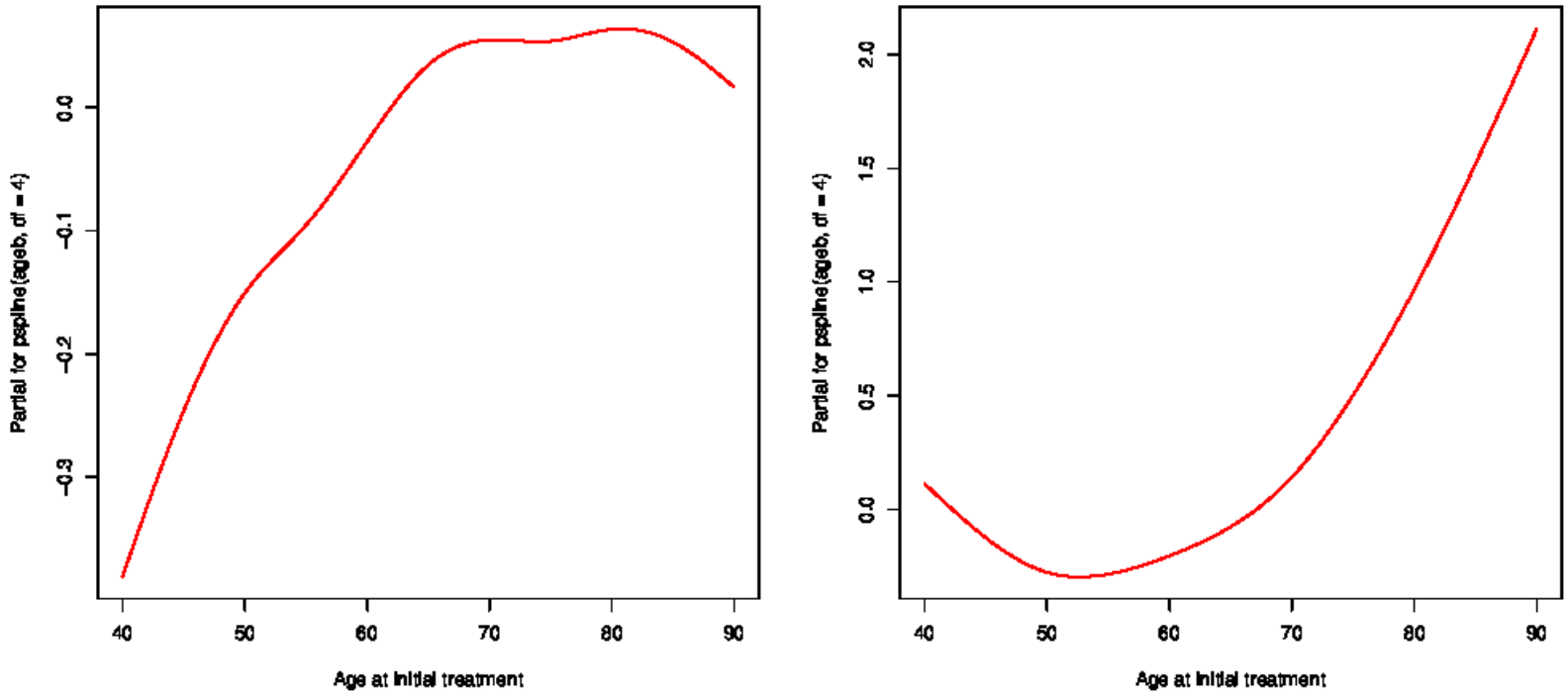
- Distinct cohorts convenient for thinking about Theorem proof.
  - proof by induction on number of cohorts
- Every individual patient can compose a new cohort

  $\Rightarrow$ Theorem applies to any study with censoring dates known in advance
- The *event* can be defined to be second, third, ... recurrence.

  $\Rightarrow$ Theorem applies to whichever event, all event numbers
- empirical CMF is calculated from these CIFs

  $\Rightarrow$ CMF is independent of times of death

Factors affecting survival: Lung and breast cancer

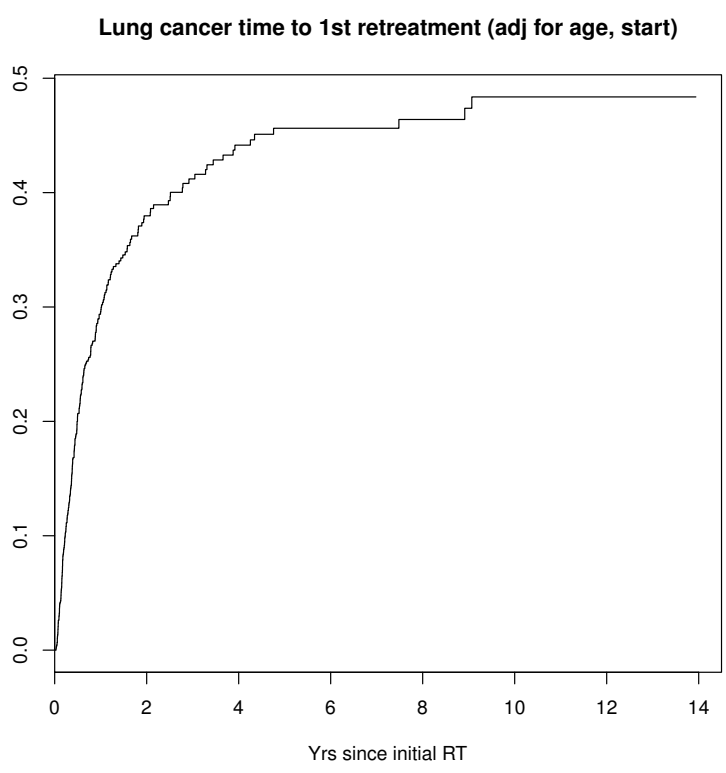


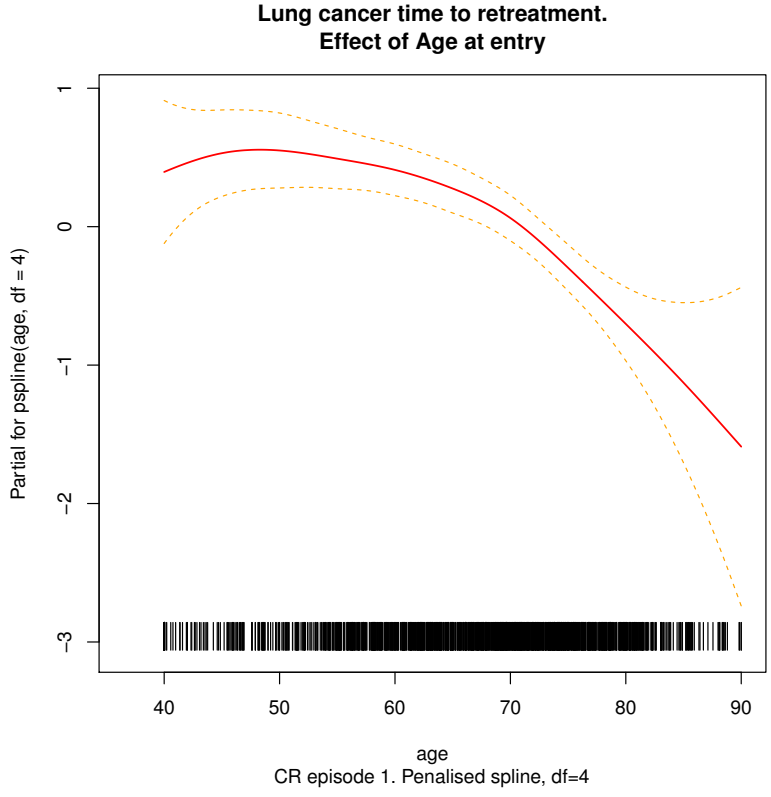
- For patients treated for Lung Cancer, shorter *survival* at older ages.
- In Breast Cancer *survival*, age effect is non-monotonic, hazard bottoms at age 50 and accelerates beyond 70.
- In Breast cancer, age and cohort effects are strongly significant.

Factors affecting mean retreatments

- Data analysis
  - Theorem suggests OCI (ignoring death): analysis of patient records: retreatments only
  - CMF will still depend on risk factors for competing events: retreatment(s) and death
  - retreatment of Lung cancer patients curtailed by death
  - less so for Breast cancer
  - to illustrate: explore effects of fixed covariates **age**
  - *Is follow-up of deaths necessary in this context?*

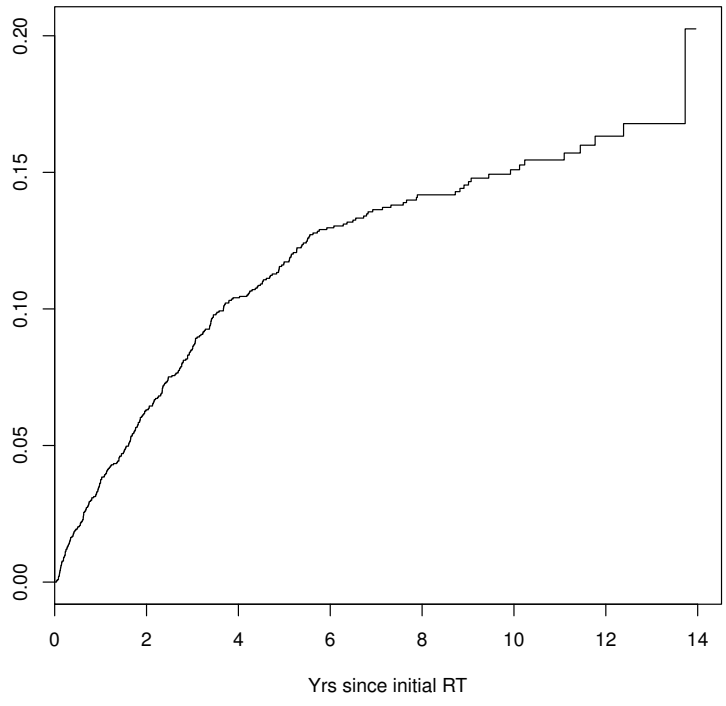
Factors affecting retreatment prevalence in Lung Cancer

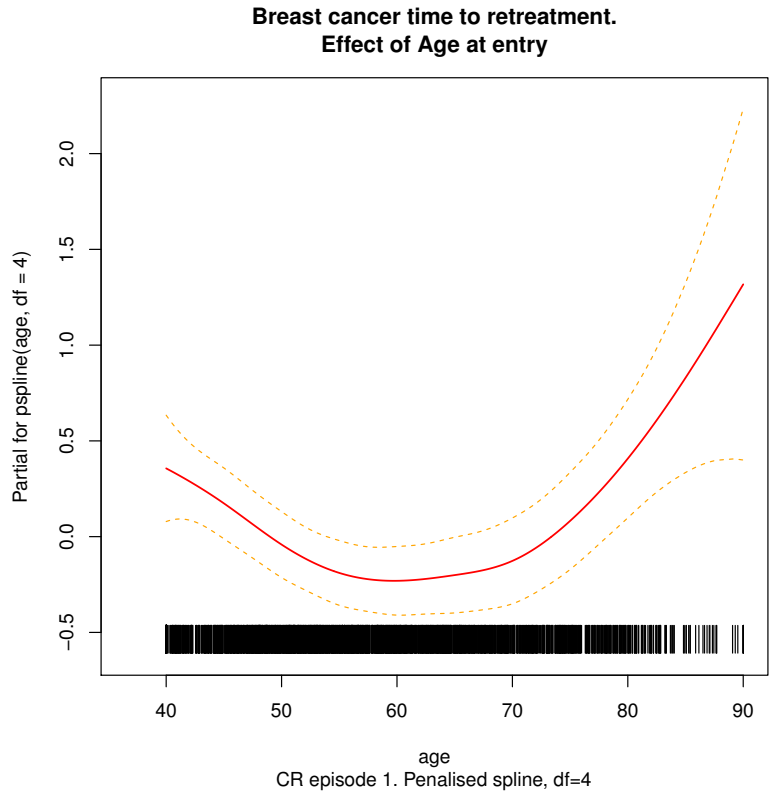




- Cox model of time to first retreatment, censoring death.
- Older Lung Cancer patients not utilising retreatment as early as others in Lung Cancer, when death has not intervened

Factors affecting retreatment prevalence in Breast Cancer





- *Younger* and *older* Breast Cancer patients utilise *retreatment* earlier

Linear and quadratic coefficients of Age

Method	Age (linear)		Age (quadratic)	
	$\hat{\beta}_1$	P-val	$\hat{\beta}_2$	P-val
LUNG CANCER				
OCI ep 1	-0.045	0.02	-0.122	P<0.001
CR ep 1	-0.040	0.04	-0.138	P<0.001
CR ep 2+	**-0.006**	NS	-0.119	P<0.001
PWP (all)	-0.032	0.07	-0.143	P<0.001
BREAST CANCER				
OCI ep 1	0.009	NS	0.133	P<0.001
CR ep 1	0.017	NS	0.173	P<0.001
CR ep 2+	**-0.016**	NS	**0.000**	**NS**
PWP (all)	0.000	NS	0.095	P<0.001

LMCTC findings: covariate effects

- *Coefficients*, their *SE*s and *P-values* differ little between OCI (ignoring deaths) and competing risk analysis (retreatment 1 versus death).
- Follow-up of *deaths* does not add much to findings in LMCTC data.
  - This suggests we may dispense with registry data on deaths:
  - revert to recurrent event model methods for a single event type
- We found no evidence of efficiency gain in estimating CMFs and risk factor effects using death data.

Conclusion

- Longitudinal cohort event histories are common
  - these track transitions (events) from state to state
  - cohorts often differ in length of follow-up
    - we may wish to forecast the future for a recent cohort
    - using knowledge from earlier cohorts with longer follow-up
    - e.g. predict mean number of events in 10 years
- Remaining length of life may also predict mean numbers of events
- We have shown that when censoring time is predictable there is no need to know who is alive /dead
  - if time of death *is* known, survival methods should *not* censor at time of death
  - the individual should remain at risk until their prespecified end-of-study
  - more complex statistical modelling will provide the same mean estimates

References

- Barton, Hudson, Delaney et al, Clinical Oncology (2011, 2014)
- Cook & Lawless, The Statistical Analysis of Recurrent Events, Springer, 2007
- Cook, Lawless et al JASA 2009
- Fine, Gray JASA, 94: 496-509, 1999
- Geskus, Biometrics 67, 39–49, 2011
- Gooley, Statistics in Medicine, 18, 695-706, 1999
- Kalbfleisch & Prentice, The Statistical Analysis of Failure Time Data, Wiley, 2002
- Lawless, Statistical Models and Methods for Lifetime Data, 2003
- Therneau & Grambsch, Modeling Survival Data: Extending the Cox Model, Springer, 2000

Outline

Jim Douglas

Topic tonight

Survival analysis with competing risks

Application: Cancer radiotherapy retreatment is South-West Sydney

Theorem

Survival risk factors

CMF covariates

Conclusion

CIF and CMF terminated by death

- consider first retreatment (C=1) with competing risk death (C=2)
- $CIF_1(t)$: subdistribution $P(T < t, C = 1)$
- Recurrent events terminated
  - now CMF, counting recurrences, is attenuated by the probability of death

Estimators (first event, recurrence or death)

- *wrong* to use KM by censoring follow-up at death
  - there will be no more retreatments
  - similarly CIF not estimable by 1- KM(t)
- cumulative mean can be estimated by methods of Cook and Lawless[4]

[4] Cook, Lawless, et al JASA 2009

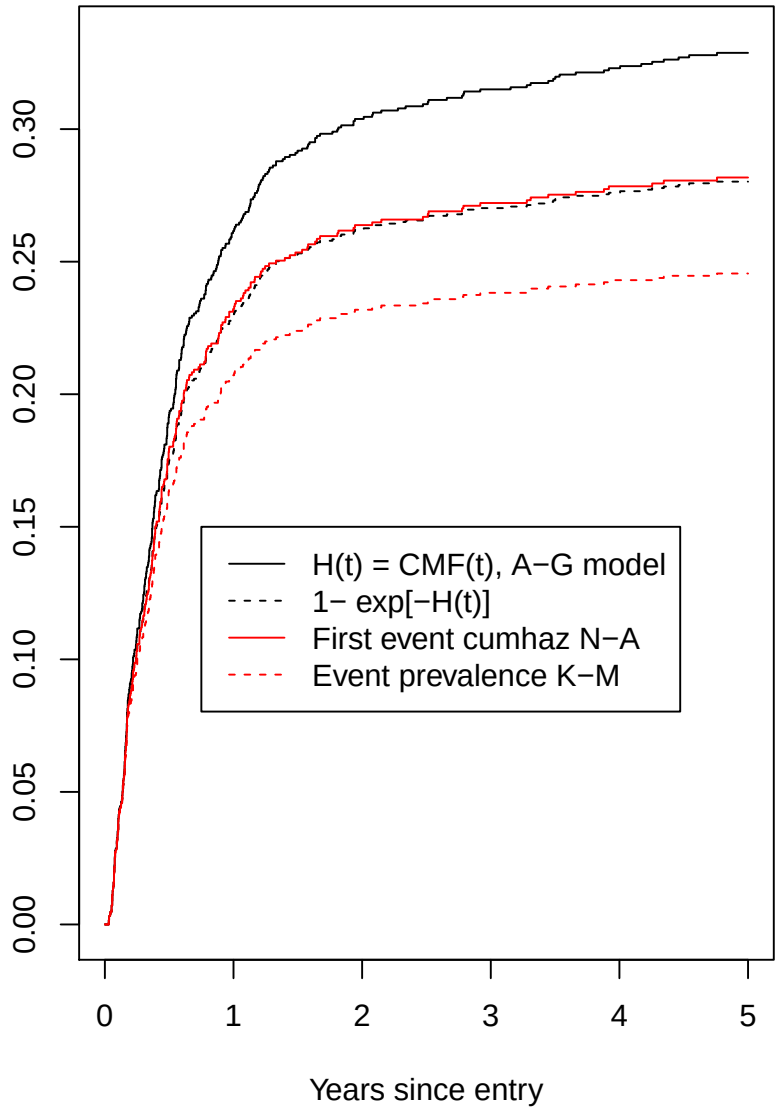
Adjusting Nelson-Aalen with composite events

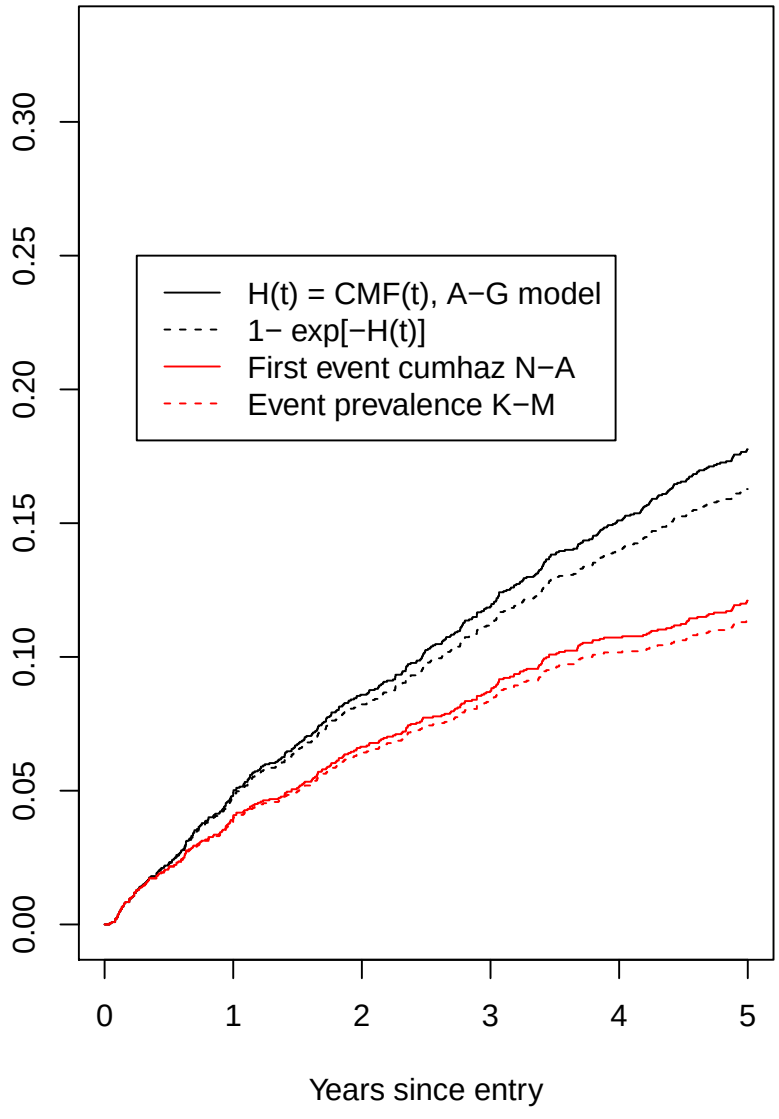
- N-A estimator provides CMF of composite recurrence/death
- subtract an estimator of cumulative incidence of death, e.g. (1-KM(t)) $\Rightarrow$ CMF of recurrences alone

Methods and data for analysing retreatments

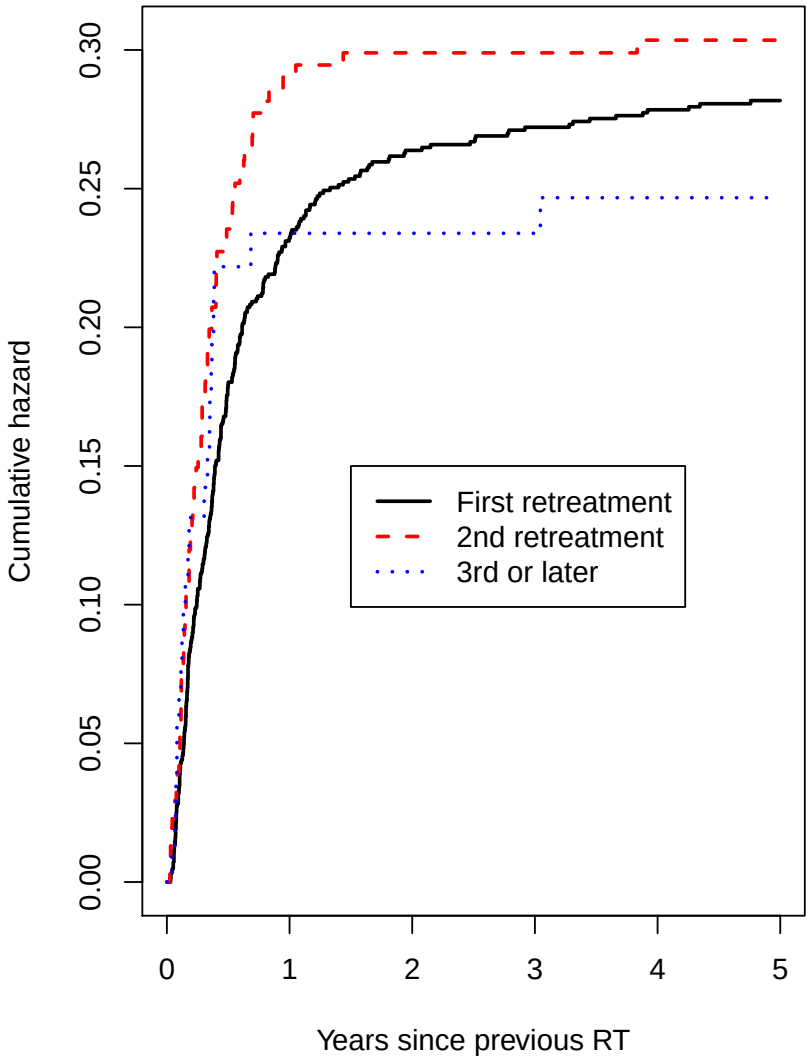
- A competing risks model separates interpretation of effects on recurrent events and terminal event.
- Some factors affect death and event incidence (sometimes in opposite directions).
- Difficult to integrate effects on mortality with effects on event numbers
  - Can we understand the net effect of a covariate on the CMF?
- OCI methods, for administrative censored data, simplify analysis to a single (recurring) event.
- Our Theorem justifies using Fine and Gray's risk set (i.e. OCI) in estimating net event incidence.

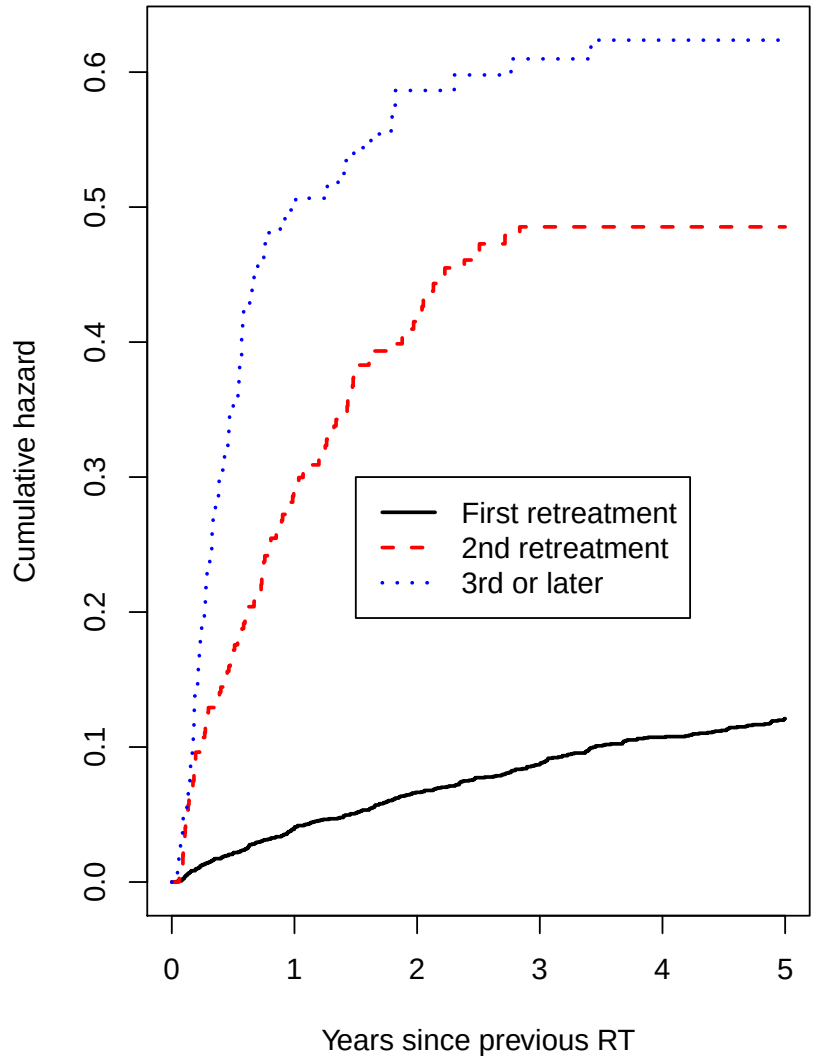
AG model fits of prevalence and CMF





PWP model





- N.B. gap time scale

PWP: interpretation

- PWP hazards vary by event number
  - the Figure provides evidence this is the better model
  - metastatic disease; curative vs palliative treatment intent?
  - use of PWP to estimate mean numbers of retreatments is hard!
  - convolutions, MSM fits