Biometrics Hobart 2015

Malcolm Hudson

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

December 2015

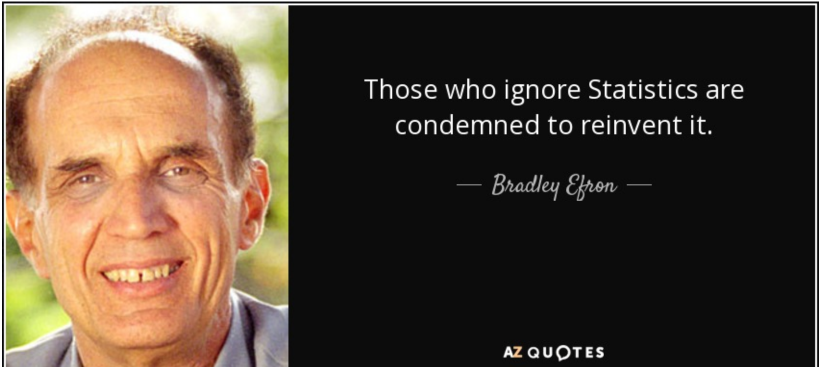
For your consideration

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.



Factors affecting treatment recurrence

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

Case study

Patterns of Retreatment by Radiotherapy in Liverpool Hospital (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[1] available

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

Analysis options with competing events

First-event analysis

- Complication-free survival time (i.e. time to first event)
- $F(t) = P(T \leq t)$, prevalence of event of either cause

Competing risk analysis:

- cause specific $CIF(t) = P(T \leq t, \delta = 1)$
- model covariate effects on cause-specific hazard of time to first retreatment
- directed at outcome of interest, censor after others (death)

Multiple recurrence analysis

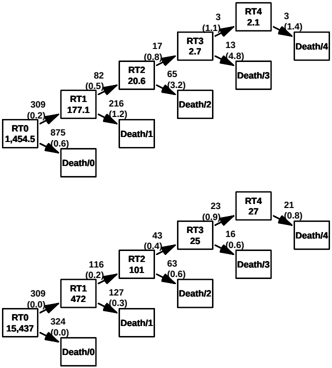
- mean numbers of events
- mean function $CMF(t) = E(N(t))$
- if $N(t) \in \{0, 1\}$: $CMF(t) = F(t)$, event prevalence

Records of recurring events

Concerns

- explain variability in mean numbers
  - fixed follow-up *or* adjust for length of follow-up
- association between recurring events and death
- CMF permits comparisons (of events per-person)
  - despite long follow-up (1999 cohort) and short (2006 cohort)
  - is medical practice changing?

MSM diagram: 1.Lung, 2.Breast cancers



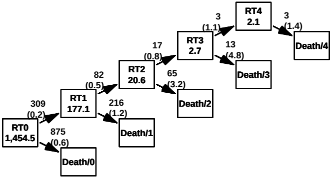
- State transition diagram and statistics.
  - Numbers of transitions from each state
  - [in box] person years (p.y.'s) at risk

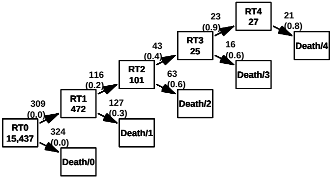
Example: Lung cancer (top)

⇒ ratios *observed deaths to retreatments* remain around **3 to 1**

⇒ *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	CL = Cook, Lawless et al, JASA, 2009
RT0 = Date of initial radiotherapy	KP = Kalbfleisch & Prentice (text)
RTk = Date of k-th retreatment	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
```

Mean estimates obtained from **survival** package, (start, stop] interval data (AG)

- using OCI risk set

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

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

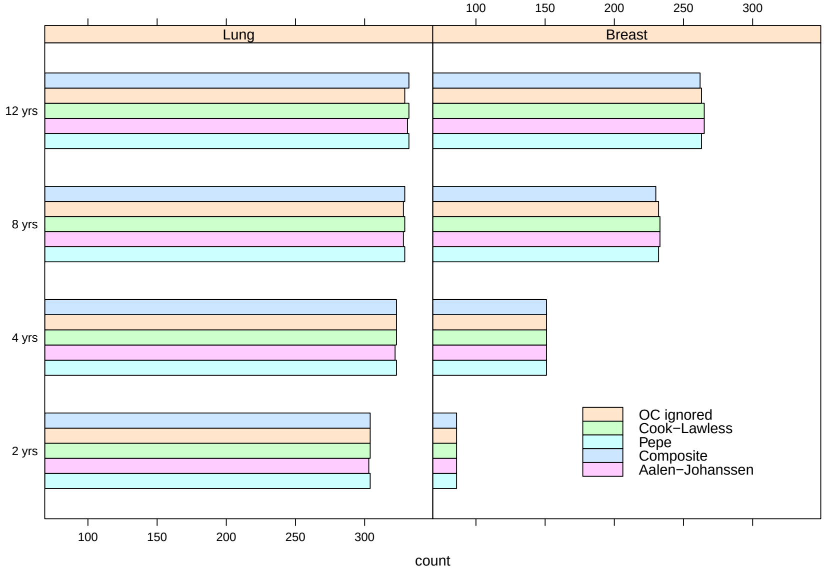
where $i \wedge j$ denotes $\min(i, j)$. In our hypothetical cohort, an individual with $\varepsilon \neq 1$ is still "at risk" for failure from the cause of interest until time C, when $T < C$. If (T, ε) and C are conditionally independent given the covariate, then the "crude" subdistribution hazard 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

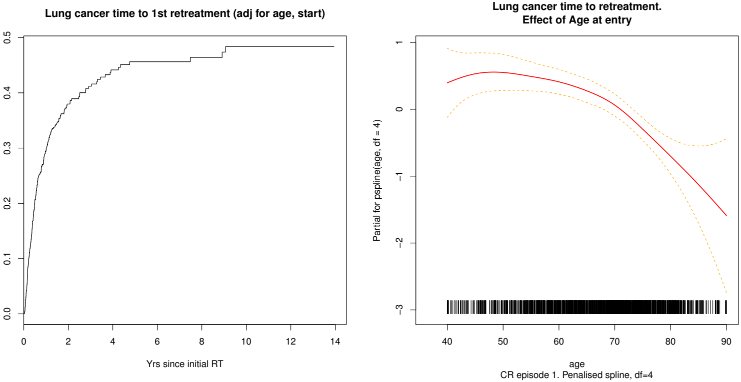
  $\Rightarrow$ CMF is independent of times of death

Factors affecting mean retreatments

Data analysis

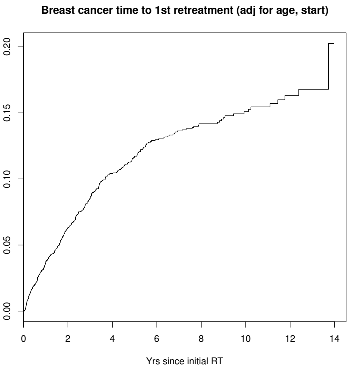
- Theorem motivates analysis of patient records using OCI: retreatments only
- relevant risk factors for retreatment(s) *and* death
  - retreatment of Lung cancer patients curtailed by death
  - less so for Breast cancer
- counting process model or stratified Cox for the recurrent retreatment times alone
- to illustrate: explore effects of fixed covariate **age**
  - the model also adjusts for **cohort** (spline function of year of entry)
  - *Is follow-up of deaths necessary in this context?*

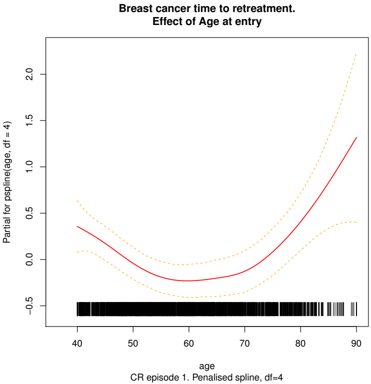
Factors affecting retreatment prevalence in Lung Cancer



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


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

Outline

Topic

Application: Cancer radiotherapy retreatment is South-West Sydney

Theorem

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[2]

[2] Cook, Lawless, et al JASA 2009

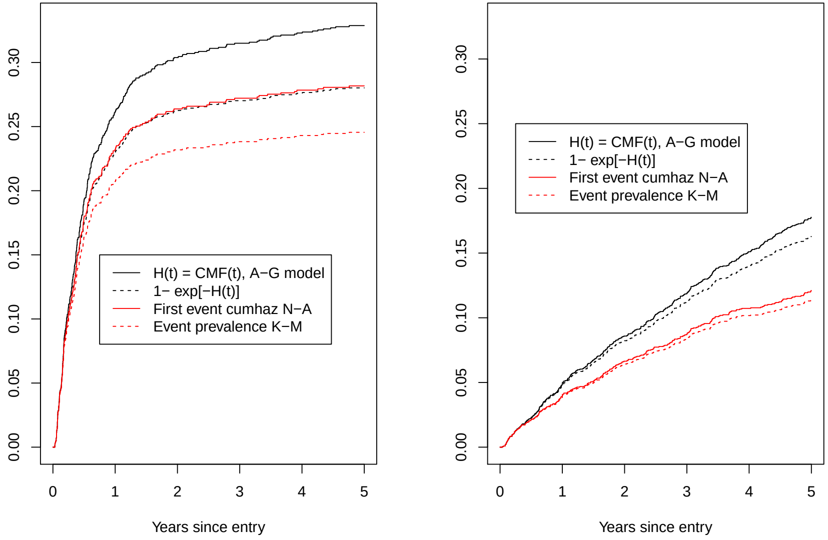
Adjusting Nelson-Aalen with composite events

- 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

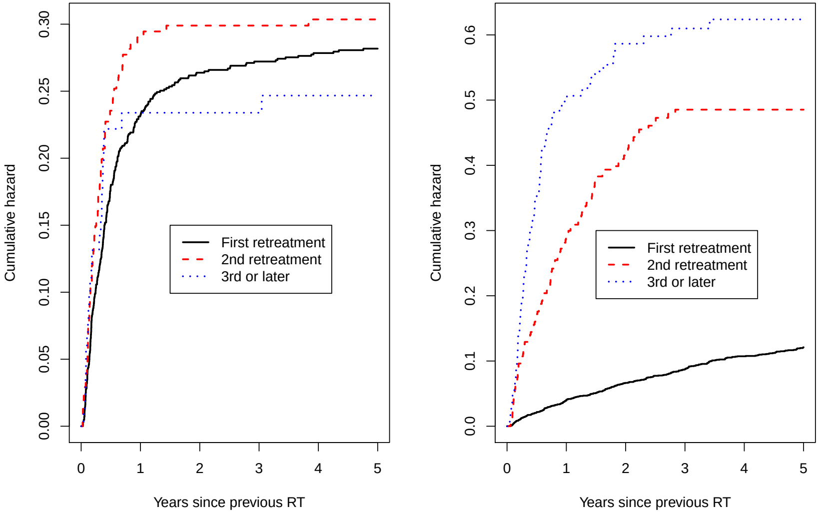
Methods and data for analysing retreatments

- 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

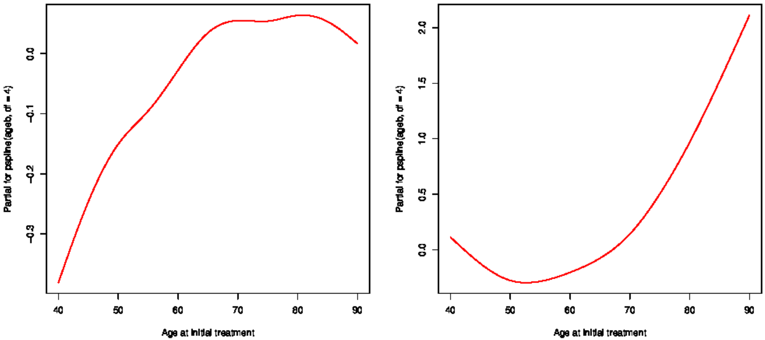


- N.B. gap time scale

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

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.