Multiple and Correlated Events

Terry M. Therneau
Mayo Clinic

Spring 2009

SURVIVAL ANALYSIS

Traditional

- ”Time to death” is the endpoint of interest
- It’s time to write the paper
- Not everyone has died yet
  - ⋆ Someone enrolled 03/01/1987
  - ⋆ Analysis on 04/05/2001
  - ⋆ We only know that survival $>$ 5149 days
- A particular kind of partial information
- Multiple specialized methods have been derived

Extensions

Methods apply to any ”time to” endpoint

- recurrence of tumor after cancer chemotherapy
- visual loss for diabetic patients
- revision of a hip replacement
- duration of unemployment

Multiple events are allowed

- Repeated infections in immunocompromised children
- Multiple fatal/non-fatal myocardial infarctions
  - MDPIT Study
    - $2.5 million
    - 2466 patients
    - 255 first events: $100,000 per event
    - 40+ second events
- Multiple toxicities
- Social interactions

Methods

- Focus on the event *time* itself
  - ⋆ Kaplan-Meier
  - ⋆ Accelerated failure time models
- Or on the event *rate*
  - ⋆ Nelson cumulative hazard estimate
  - ⋆ Cox model

Let

$$
\begin{aligned}
\lambda(t) &= \text{event rate (assume continuous) or hazard} \\
\Lambda(t) &= \int_0^t \lambda(s)ds = \text{cumulative hazard} \\
F(t) &= \text{cumulative distribution function of (first) events} \\
S(t) &= 1 - F(t) = \text{Survival function} \\
&= e^{-\Lambda(t)}
\end{aligned}
$$

Event Rates

In biological data, covariates normally affect the rate of event

- ”Add two years” : linear model
- ”Halve the recurrence rate” : proportional hazards
- ”Age half as fast” : accelerated failure time

My examples are focused on SAS and S-Plus. SAS for the data manipulation, S for the graphics, and either for the analysis. (These are the packages that I know).

SAS

Largest | usage
revenue
manuals
code | system.

Deserved reputation for reliable and accurate procedures.
(Undeserved reputation of infallibility.)

S-Plus

Commercial and supported release of S, a package originally developed at Bell Laboratories.

Includes:

- documentation and support
- data management (e.g. import from SAS)
- graphical user interface
- add-on libraries
- Web interface to a central engine

R

An open-source implementation of S.

- Base package: very solid, fast, free
- Add-ons: variable
  - survival: same code as S-Plus
  - cutting edge material, rapid growth
  - ...
- Much activity in the university realm

The Kaplan-Meier

The KM estimates a probability at each event time t_i

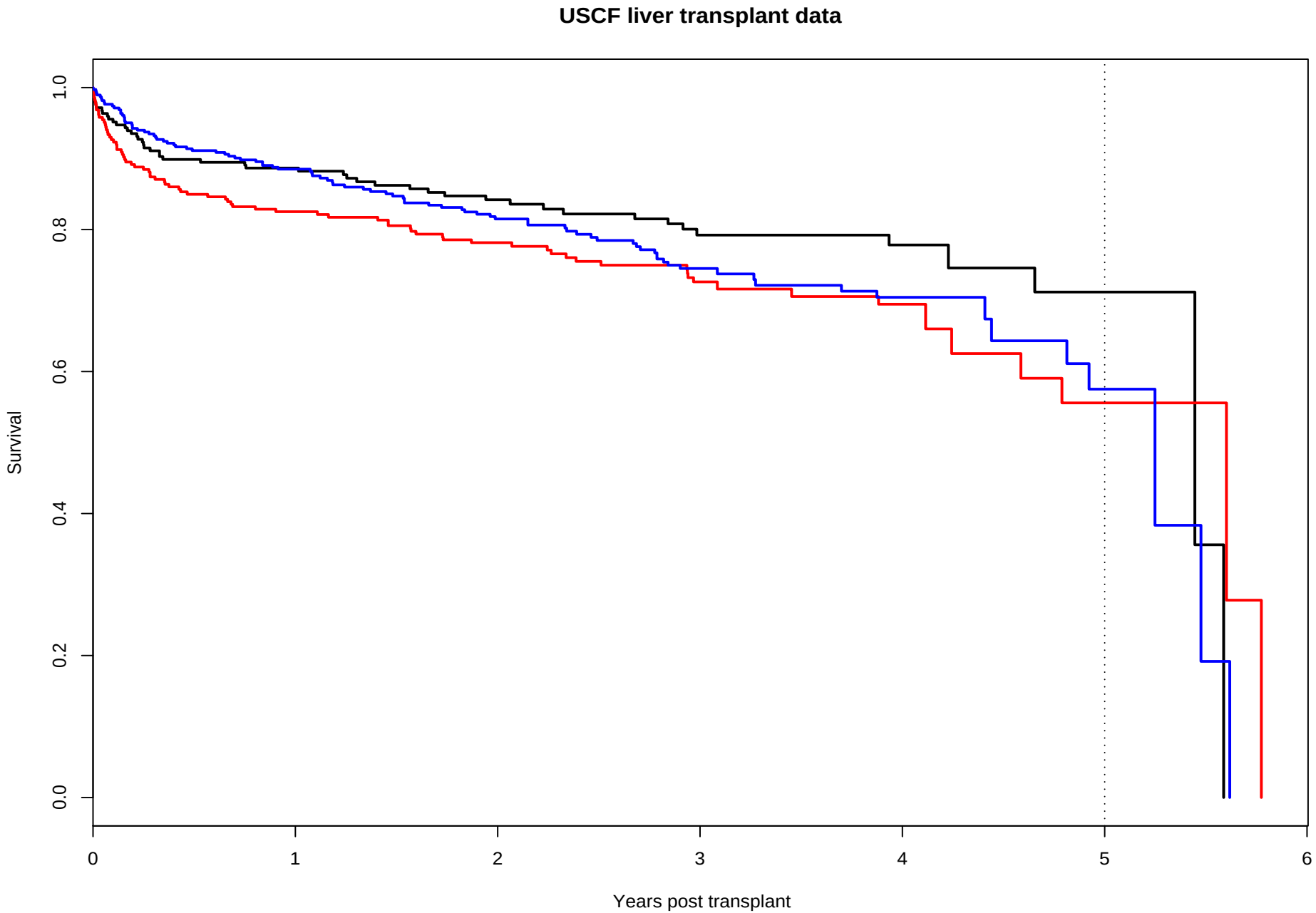
$$p_i = \frac{\text{number of events at time } t_i}{\text{number of subjects at time } t_i}$$

Then the overall curve is

$$Pr(\text{still alive}) = (1 - p_1)(1 - p_2) \ldots$$

The concept is well known, any code is fully debugged: what could go wrong?

Risk sets


USCF liver transplant data
Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
Years post transplant

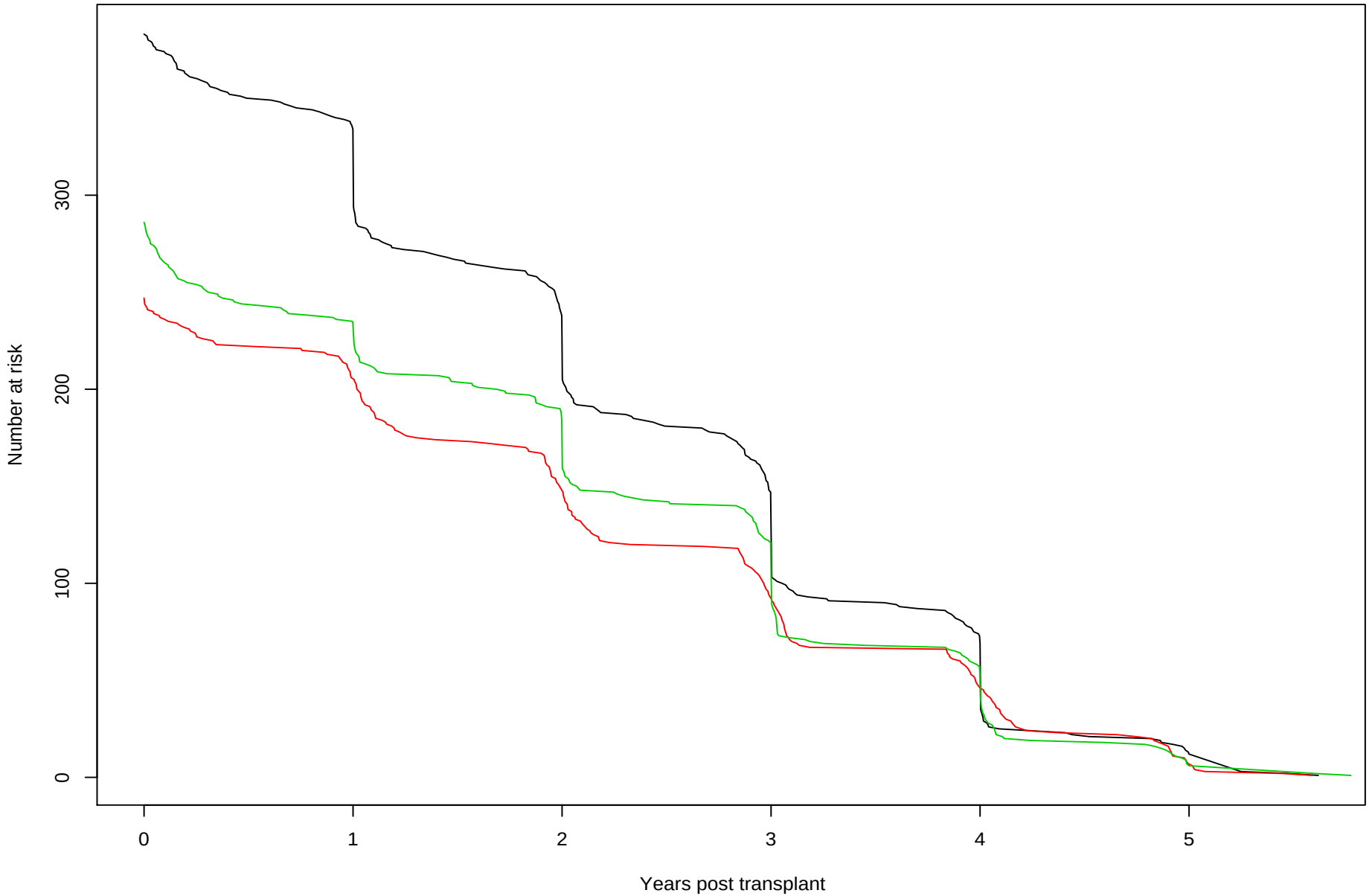
UCSF transplant data

- Nearly complete follow-up until 5 years
- Study funding ended after 5 years, minimal study assistant time
- Joint analysis with another group at year 6
- Between years 5 and 6, 6 deaths were noted via unsolicited follow-up, and added to the data set.

The correct denominator for a KM at some time t is the number of subjects for whom

- If an event occured, we would hear about it, and
- it *would be counted* in the numerator

Risk sets



Cox model

Let $r_i = e^{X_i\beta}$ be the risk score for each subject i.

If there is an event at time t, then the partial likelihood contribution at time t is

$$PL(t) = \frac{\sum r_i \text{ for those with an event at time } t}{\sum r_i \text{ for those at risk for an event at time } t}$$

The increment to the hazard function for a (hypothetical) subject with risk r_0 is

$$p_0(t) = \frac{\text{number of events at time } t}{\sum r_i/r_0 \text{ for those at risk for an event at time } t}$$

Risk sets

The *risk set* for a KM or Cox model at some time t is the set of subjects who could have a recorded event at time t, no more and no less.

Sometimes this is not so easy

- Deaths found through outside sources.
- Patients with primary care elsewhere.
- Fortuitous early detection.

Sometimes it just requires thought

- Multiple events
- Disallowed events

Cox proportional hazards model

Accelerated failure time (AFT) models predict the event time $\log(T)$ directly, and are similar to non-survival regression in that sense.

The Cox model models the event *rate*. The model assumes that the risk for subject i is

$$\lambda_i(t) = \lambda_0(t) e^{X_i(t)\beta},$$

where λ_0 is an unspecified baseline hazard. Because the hazard ratio for two subjects with fixed covariate vectors X_i and X_j

$$\begin{aligned} \frac{\lambda_i(t)}{\lambda_j(t)} &= \frac{\lambda_0(t) e^{X_i\beta}}{\lambda_0(t) e^{X_j\beta}} \\ &= \frac{e^{X_i\beta}}{e^{X_j\beta}} \end{aligned}$$

is constant over time, the model is also known as the *proportional hazards* model.

Elements of the model

The basic premise is to model the death (or event) *rate*.

- This avoids the main problem of censored data. Someone who is followed for 18 months is a part of the rate computations for months 0–18, and not a part of them thereafter.

- Why exp(linear predictor)? To avoid negative death rates.
  - ⋆ Implies that factors are multiplicative, e.g., treatment reduces the death rate by 28%.
  - ⋆ Two covariates multiply in effect.
  - ⋆ For acute disease the model seems to fit well.
  - ⋆ The model $\lambda_0(t)[1 + X\beta]$ is sometimes proposed. Programs are unreliable.

- The baseline hazard is arbitrary.
  - ⋆ Scientifically comforting
  - ⋆ Does not extrapolate.

Computation

Mathematically, the Cox model ends up as a “lottery” model.

Risk score for each subject = $r_i = \exp(X_i\beta)$.

Say a death occurs at day 174, subject 24. The likelihood term is

$$\frac{r_{24}}{\sum_i Y_i(174) r_i}$$

Where Y is the “at risk” indicator. It is 1 for subjects who are

- still alive
- still under observation
- not excluded

Probability(person who did win the lottery, should have won the lottery, given the set of people who were participants in the drawing).

Stratified Cox model

An extension of the proportional hazards model is to allow for multiple *strata* in the fitting procedure. That is, assume that the subjects can be broken into multiple groups, and the hazard for subjects in the kth group is

$$\lambda_k(t)e^{X_i\beta}.$$

A common use of stratification is in multi-center trials. Because of different patient populations and referral patterns, different institutions in the trial may have quite different hazard rates, yet a common treatment effect across institutions. In this way, strata play a similar role to multiple intercept terms in an analysis of covariance model.

Counting Process Style

Think of each subject as the realization of a very slow Poisson process.

Censoring $\neq$ “incomplete data”

rather

“the Geiger counter just hasn’t clicked yet.”

The computing idea here is not new. Laird and Olivier, JASA 1981, show how to use a Poisson regression program to fit a Cox model.

Also see Whitehead, 1980, Applied Statistics.

Computer Implementation

Cox model program with just *one* more variable. Replace

time status strata x1, x2, . . .

with

(time1, time2] status strata x1, x2, . . .

- Over the intervale time1 $< t \leq$ time2
- $x1$, $x2$, /ldots are the covariates that apply
- strata = the strata for that interval
- status=1 if the interval ended with an event

This simple extension allows

- time dependent covariates
- time dependent strata
- left truncation
- multiple time scales
- multiple events per subject
- independent increment, marginal, and conditional models for correlated data
- various forms of case-cohort models

Time dependent covariates

Consider the famous Stanford Heart Transplant study. All patients start on medical treatment. When a heart becomes available, the selected patient "converts" to the transplant treatment arm.

Patient 1 has died after 102 days; patient 2 was transplanted at day 21 and died on day 343. The data file for these two patients would be

Interval	Status	Transplant	Age at Entry	Prior Surgery
(0,102]	1	0	41	0
(0,21]	0	0	48	1
(21,343]	1	1	48	1

Note that static covariates such as age are simply repeated for a patient with multiple lines of data.

It is worthwhile to pursue this example in detail.

- 103 subjects
  - 69 received a transplant
  - 34 did not
  - The data set will have 69*2 + 34 = 172 rows.
    - maybe: 2 subjects were enrolled on the day of their transplant
- One subject died on the day of entry
  - (0,0] is an illegal interval
  - So treat them as a death on day 0.5
- A subject transplanted on day 10 is on
  - the medical treatment arm days 0 - 10
  - the transplant treatment arm days 11 - last contact
  - (transplants happen later in the day than deaths)
  - Except patient 38, who died during surgery on day 5
    - medical treatment days 0 - 4.5
    - surgical treatment arm days 4.5 - 5

- The data is somewhat collinear
  - ⋆ to get the exact same coefficients as are found in table 5.2 of Kalbfleisch and Prentice one must use the exact same variable definitions.
  - ⋆ Age is (age in days)/365.25 - 48 years
  - ⋆ Entry time is (days since study began)/365.25

- (The data in the appendix of K&P will not give the “correct” answers since age is rounded to the nearest year).

- The data found in the JASA article of Crowley and Hu contains the actual dates.

- Note that using 0.5 or 0.1 (or any number that does not occur naturally in the data) instead of 0.9 will give the same result.

```
data temp;
    infile 'data.jasa';
    input id  @6 birth_dt mmddyy8. @16 entry_dt mmddyy8.
              @26 tx_dt mmddyy8.
              @37 fu_dt mmddyy8.  fustat  prior_sx ;
    format birth_dt entry_dt tx_dt fu_dt date7.;

data stanford;
    set temp;
    age =  (entry_dt - birth_dt)/365.25  - 48;
    year = (entry_dt - mdy(10,1,67))/ 365.25;  *time since 1 Oct 67;
    wait = (tx_dt - entry_dt);
    if (id = 38) then wait = wait - .5;

    if (tx_dt =.) then do;
        rx = 0;                * standard therapy;
        start = 0;  stop  =  fu_dt - entry_dt;
        if (stop =0) then stop = .5;
        status= fustat;
        output;
        end;

    else do;
        rx =0;
        start = 0;    stop  = wait;
        status= 0;
        output;
```

```
        rx =1;
        start = wait;  stop  = fu_dt - entry_dt;
        status= fustat;
        output;
        end;

proc print;
```

S-Plus code to fit the model

```
coxph(Surv(start, stop, status)  ~ age + year + rx,
data = stanford)
```

SAS code to fit the model

```
proc phreg data=stanford;
    model (start,stop) * status(0) = age year rx;
```

Four steps

1. Think through the problem.

2. Create the (start, stop] data set
   tedious but straightforward

3. Check the data set for sanity

   - PRINT OUT some or all of the cases
   - Read the printout carefully

4. Fit the model
   — trivial

Multiple Time Scales

$$\lambda_i(t) = \lambda_0(t)e^{X\beta}$$

Here λ_0 is the baseline hazard, which we are not attempting to model. But what is the best time scale t?

- time since entry into the study?
- time since birth?
- time since diagnosis?
- calendar time?

For all but the first the counting process form may be necessary.

Parkinson's disease patients at Mayo

For scientific reasons, we prefer to use time since diagnosis. That is, in assessing the effect of L-Dopa, a patient who died at 2 years after diagnosis is compared to other patients who are two years from diagnosis.

Not everyone comes to Mayo immediately. Time from diagnosis to referral may be a function of distance, past association, affluence, and many other factors which have little to do with the disease state.

On this time scale, a patient diagnosed in 1985 and referred in 1987 is not at risk for death ("observed death in my study") in the interval 0-2 years.

This is known as left truncation.

Time alignment

Without the counting process form all intervals implicitly start at zero. The older, and unsatisfactory solution is to create a dummy variable d = delay time from diagnosis to entry, and enter d into the model as a covariate.

			Traditional		Correct
Dx	Referral	Event	Interval	d	Interval
9Nov84	10Mar85	1Mar92	(0,2548]	121	(121,2669]
8Aug82	28Apr85	18Jun90	(0,1877]	994	(994,2871]
3Jul83	22Jul85	1Mar92	(0,2414]	750	(750,3164]
28Nov84	1Aug85	12Feb88	(0,925]	246	(246,1171]

Epidemiology of Breast Cancer

A pending grant to look at risk factors for breast cancer, particularly early onset disease (Dr. Tom Sellers).

- 426 families, identified as sequential cases in 1944–52, follow-up to 1995.
- 9073 females: 1134 first degree, 3939 second, 4000 marry-ins
- Covariates include diet, smoking, estrogen, . . .

Age is the appropriate time scale for the model, since it is the largest effector of *baseline* cancer risk.

Mortgage defaults

The rate and timing of loan defaults (foreclosure) is of great interest to firms that “bundle” loans to create an investment. Some time scale issues are:

- Loans do not become a part of the portfolio on day 0 (the local bank may hold them for a time). This induces left truncation: the loan is not at risk for “observed failure” until it is acquired. Failures before that point are neither observable nor of interest.

- Two natural time scales are time since initiation and calendar time.

The baseline hazard for a Cox model is very flexible, and plays a similar role to blocking effects in linear models — “effects that should be adjusted for, but are not modeled directly”.

The default rate might vary with time since initiation, e.g., loans held for more than 10 years represent stable families. Not all these effects can be modeled or need to be modeled.

Calendar time can affect defaults as well, e.g., an economic downturn.

Stratification by geographic region would allow these effects to be modeled per region.

The flexible baseline hazard λ_0 is a strength of the Cox model — you don't have to model everything.

One problem with $\lambda_0(t)$ is that *you only get to pick one time scale*. Either time-since-initiation or calendar-time, but not both. If both are important (likely), then one must be modeled explicitly using time-dependent covariates. Suggestion: model the simpler one.

Another problem is that time cannot be extrapolated, the Cox model can only predict within the range of $\lambda_0(t)$ for the data at hand.

Zero length intervals

Zero length intervals such as (11,11] are an error.

- SAS silently ignores such data lines.
- S-Plus generates an error message.

(I prefer the S-Plus behavior, since such intervals usually mean that I have made a coding error).

The most common problem with this convention is a death on day 0.

Suggestion: have the subject die on day 0.5 (or .1 or .7 or . . .).

- The actual value of the number will not effect the model results, as long as the value is before the first actual death or censoring event.
- The value will effect the survival curve based on the Cox model (but requires good eyesight to see it), as well as the tests for proportional hazards.

Summary

- Easy to set up many different models using available software
  - ⋆ available since 4/91 in S
  - ⋆ available in SAS version 6.10 or later (e.g. 3/95 for PC)
- Nearly all of the work is in creating the data set.
  - ⋆ Because the user creates the data, it is easy to code variants not thought of by the Cox program's authors.
  - ⋆ Because the user sets up the data, it is possible to mess up, and fit an absolutely idiotic model.
- The created data set can be carefully examined before the fit
  - ⋆ Avoid errors
  - ⋆ Assurance of exactly which model is being fit

This is a real study, analysis done in my department, names changed to protect the guilty.

- Placebo controlled trial of a new drug, survival endpoint.
- Some patients on the active arm were expected to have, and did have, an adverse reaction
  - small numbers (4–5%)
  - early onset
  - cross-over to placebo
  - no such reactions on placebo
  - variable `weeks on rx` added to form
- Time dependent treatment covariate

SAS code

```
data anal; set main;
    futime = last_dt - entry_dt;
    if (rx=1 and futime > weeks*7) then do;
```

Counting Process Form

```
start=0; stop = weeks*7; status=0; output;
start=stop, stop=futime; status=death; rx=2; output;
end;
    else do;
start=0; stop=futime; status=death; output;
end;
```

What could be wrong with 10 lines of code?

- assume death exactly 1 year after enrollment, no crossover
- `futime=365`, `weeks=52`, `weeks*7 = 364`
- Due to round off, about 1/2 of the treatment patients are ‘crossed over’ to placebo 1–3 days before death or last follow-up
- Cox model showed a 50% reduction in deaths due to treatment; correct analysis gives $\hat{\beta} \approx 0$.
- Discovered when working up the example for a course
- Paper was in process

Four steps

1. Think through the problem.

2. Create the (start, stop] data set
   — tedious but straightforward

3. Check the data set for sanity
   - PRINT OUT some or all of the cases
   - Read the printout carefully

4. Fit the model
   — trivial

Time dependent covariates

Computation

The Cox likelihood is a series of terms, one for each event

$$\frac{r_e}{\sum_{\text{those at risk}} r_j}$$

where r_j is the (modeled) risk for subject j and r_e is the risk for the person who actually had the event.

This is essentially a sequential lottery model, where each term is

$$Pr(\text{person who won the drawing, should have won the drawing})$$

(Think of r_j as the number of tickets that each person purchased).

- At each event time t, $r_j(t)$ depends only on the jth person's covariates *at that time*.

- Past history can be included, as well as the present, with suitable variables

- The simple "at this moment" structure makes it easy to code time-dependent covariates

Warnings

- You must not look into the future.
- Avoid prophetic variables.
- Bad things happen if you look into the future.
- Duration variables work surprisingly rarely.
- It’s all too easy to look into the future.
- Short term prediction is uninteresting.
- You must not look into the future.
- It is challenging to draw survival curves.

Responders vs non-responders

This particular incorrect analysis is re-discovered every few years in oncology. (Or I should say re-published).

Like the mythical hydra, it never seems to go away.

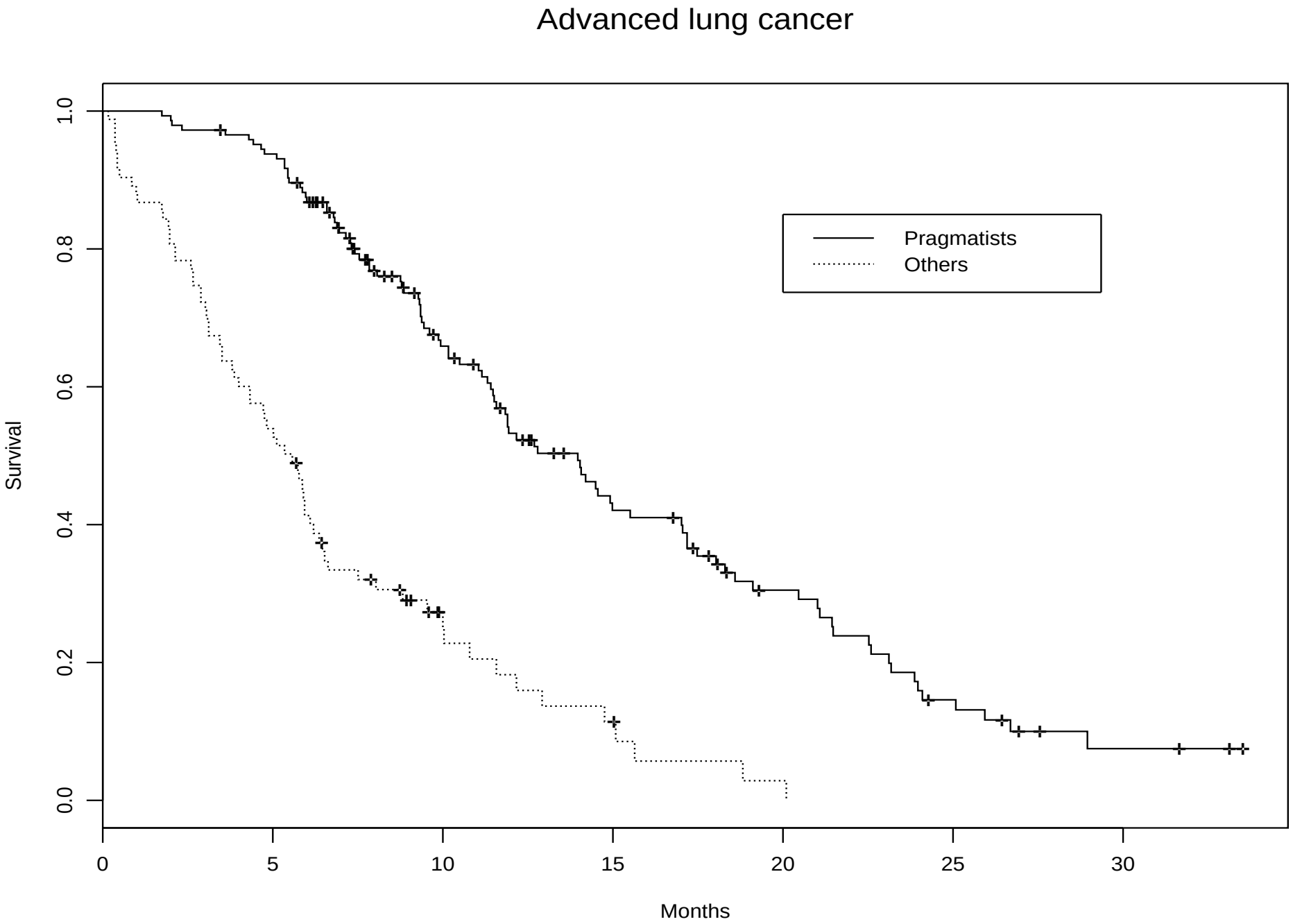
Group people, at baseline, according to whether they eventually had a complete or partial response to therapy (shrinkage of tumor), and then draw the survival curves. Surprise – responders always do better! Why?

- Assume patients are evaluated every 4 weeks.
- Response, if it occurs, will happen by week 12.
- Anyone who dies before week 4 is a non-responder, and most of those who die before week 8.
- You have to live longer to be called a responder.

Simple simulation:

- At each visit, ask the patient if they filled up their car with gas that day.
- 1/5, on average, answer ‘yes’.
- Anyone who answers ’yes’ at some time is a pragmatist.
- Do pragmatists live longer than others?
- Using an advanced lung cancer data set, we randomly added the pragmatist variable to each patient. (True if they ever had filled up).

Counting Process Form



Prophetic variables

Some time-dependent covariates are not predictors of an event as much as they are markers of a failure-in-progress:

- Multiple-organ failure
- Ventilation
- “Have called the priest”
- Medication changes
  - ⋆ Cessation of diuretics in heart failure
- PSA and prostate cancer, if measurement and declaration occur on the same visit

These will tend to be phenomenal predictors.

So what?

Basic rule

At any time point, the covariates can be anything that you want, *as long as* they use only information that would have been available on that day, had analysis been done then.

The perils of pathology:

- Patients have been classified into A/B/C based on an expert reading.
- During follow-up, some patients in the low-risk group have early progression.

- The pathologist asks “did I miss something”, and reviews the slides of those subjects.
  - ⋆ This is good clinical practice
- If something is spotted, you can *not* go back and change the diagnosis group. Results will be badly biased.
- Blinded review of an entire cohort is acceptable.

Lagged variables

Often it is preferable to not incorporate a result that is found on day x immediately. Rather lag it to appear at $x + lag$.

- Multi-day visits
  - ⋆ If a patient has a multi-day appointment (common at Mayo), the dates on the electronic record may show $a > b > c$ when actually they are concordant.
  - ⋆ Solution – we often have enter time-dependent lab test variables in the model at day + 3 or day + 7.
  - ⋆ Model checking: don't trust anything that happens ¡ 1 week before an event.

- Treatment delays
  - UDCA is a drug used in liver disease, that works by being incorporated into the bile reclamation loop.
  - It requires 4-6 months to reach saturation levels
  - In a randomized trial of UDCA vs placebo, the end of the study contained an open label period.
  - On the day of cross-over, prior placebo patients are not yet getting benefit — should I count them as treated?
  - The treatment group was changed 3 months after the drug.

- Short term prediction
  - Often it is just not *interesting* that one can predict events within 1–2 weeks.
  - More relevant question: can variable ”x” predict things that happen 4 or more weeks away?
  - Defer the time-dependent covariate’s change for 4 weeks.

Insidious look-ahead

Smoothed continuous variables

- A particular lab test has values of
  - 120 on day 0
  - 150 on day 90
  - 180 on day 120
- What should we use for the value at day 100?
- It is tempting to use 160 (1/3 of the way between 150 and 180).
- Bad idea

Persistence

- Patients with a solitary plasmacytoma are treated with local radiation
- The tumor produces an immunoglobulin spike

- If the spike is still present at the 1 year evaluation, this is a bad thing. (It mean that the ‘solitary’ lesion likely was not solitary.)
- Want to draw a curve for “survival, post 1 year”.
- Does the patient evaluated at 13 months (with persistence) go in the ‘persistent spike’ or ‘other’ group?
- We know that the spike would have been present at 12 months, if the test had been done then.
- Still, it’s a bad idea.

In the Cox model, a time-dependent variable does not change until it is observed to change.

Multiple Events

Methods

- Time to first event
- Special time-dependent covariates
- Random-effects models (frailty)
- Marginal models

Answers

- Time to first event is simple, safe, and wastes data.
- Specially constructed time-dependent covariates may mislead.
- Random effects models are interesting, and our understanding is beginning to mature.
- Marginal models are simple to use, interpretable, and flexible. They can be fit with standard software.

Fitting the models

There are 3 major things to think through

1. Creation of the appropriate risk sets
   - Parallel vs sequential
2. Separation of event types into strata
3. Accounting for correlation
   - Marginal (GEE) approach
   - Random effects (frailty) approach

Two computational things

- Building the data set
- Running the code

Risk sets

- Parallel
  - unordered events
  - multiple toxicities for a single subject
  - correlated family members
- Sequential
  - ordered events
  - repeated infections in immunocompromised children
  - treatment - response - relapse
  - any events that are *labeled* as 1, 2, 3, . . .

Sorting events into strata

$$\lambda_i(t) = \lambda_{0k}(t)e^{X_i\beta}$$

Each baseline hazard captures the baseline *rate* for an event.

When events are of different types, one should have different baselines.

Are first and second heart attacks the same type of event?

Marginal models, heuristic rationale

Consider a weighted linear model, either the case of subject weights or a fully known variance matrix Σ.

If we fit the model ignoring the weights

- $\hat{\beta}$ usually changes very little
- $\mathrm{Var}(\hat{\beta})$ may be badly incorrect

Robust variance

The robust variance is based on a “sandwich” estimate

$$V = \mathcal{I}^{-1} B \mathcal{I}^{-1}$$

where

- $\mathcal{I}^{-1}$ is the usual variance estimate of a Cox model, the inverse of the information matrix $\mathcal{I}$.
- B is a correction factor.

There are several ways to motivate this estimate:

1. The proper variance when a likelihood for distribution f is fit, but the data come from g.

2. Multivariate form of a variance inflation factor.

3. Approximation to the jackknife.

Conceptual derivation

- An honest estimate of variance, given possible model misspecification
  $\Longrightarrow$ jackknife estimator

- Limited computer budget
  $\Longrightarrow$ approximate jackknife estimate $D'D$

- Correlated data
  $\Longrightarrow$ grouped jackknife

Correlated Data

- The jackknife is a consistent estimate of variance as long as the observations left out are independent of those left in.

- If the correlation is confined to disjoint groups, then use a grouped jackknife: leave out one *group* at at time.

Diabetic Retinopathy Trial

- Two treatments.
- Randomly assigned to left and right eye of each patient.
- Leads to a data set with two observations (eyes) per subject.

$$
\underset{D_{2n\star p}}{\begin{pmatrix} d_{11} & d_{12} & \dots & d_{1p} \\ d_{21} & d_{22} & \dots & d_{2p} \\ d_{31} & d_{32} & \dots & d_{3p} \\ d_{41} & d_{42} & \dots & d_{4p} \\ d_{51} & d_{52} & \dots & d_{5p} \\ \vdots & \vdots & & \vdots \end{pmatrix}} \quad \Longrightarrow \quad \text{add} \quad \underset{\widetilde{D}_{n\star p}}{\begin{pmatrix} \tilde{d}_{11} & \tilde{d}_{12} & \dots & \tilde{d}_{1p} \\ \tilde{d}_{21} & \tilde{d}_{22} & \dots & \tilde{d}_{2p} \\ \vdots & \vdots & & \vdots \\ & & & \end{pmatrix}}
$$

D has one row per eye, $\widetilde{D}$ has one row per subject.

$\widetilde{D}'\widetilde{D}$ is a grouped jackknife estimate of variance.

- For large group sizes this may be a poorer approximation than D.
  - ⋆ Imagine 100 subjects in 5 families.
  - ⋆ The approximate jackknife will be the sum of ≈ 20 rows.
  - ⋆ Since this is not necessarily a "small" change in $\hat{\beta}$, the approx may not be as accurate, numerically.

- The number of subjects per group need not be identical.

- For a generalized estimating equations model (GEE), this is the *working independence* estimate.

- For a Cox model, this is the correlated data variance estimate of Wei, Lin and Weissfeld.

Random Effects models

$$\begin{aligned} \lambda(t) &= \lambda_0(t)e^{X\beta+Zb} \\ b &\sim G(0,\Sigma) \end{aligned}$$

- Assume: conditional on b, observations are independent — all the correlation is through b
- Then estimating the RE model is sufficient

Marginal or RE?

- Marginal
  - ⋆ Available in multiple packages
  - ⋆ Works in all cases (jackknife connection)
  - ⋆ Only the *working independence* variance is available
- Random Effects
  - ⋆ Can accomodate complex variance relationships
  - ⋆ Estimates of Σ and/or b may be of interest
  - ⋆ Assumption that you have the right model
  - ⋆ Young software
  - ⋆ Debate over the form of the random effect

All of the arguments between the ‘GEE’ and ‘Mixed’ camps are available.

S Example

```
mfit <- coxph(Surv(futime, status) ~ rx + age +  cluster(id),
  data=diabetes)

rfit <- coxme(Surv(futime, status) ~ rx + age,
              random= ~(1|id), data=diabetes)

rfit <- coxme(Surv(futime, status) ~ rx + age + (1|id),
      data=diabetes)
```

SAS

```
proc phreg data=diabetes covsandwich(aggregate);
    model futime * status(0) = rx age;
    id id;

No random effects
```

Example: Doubled data

```
> fit1 <- coxph(Surv(futime, status) ~ rx + karno +cluster(id),
data=veteran,  method='breslow')
> fit1

         coef exp(coef) se(coef) robust se      z       p
   rx  0.1736     1.190  0.18309   0.19233  0.903 3.7e-01
karno -0.0338     0.967  0.00508   0.00483 -6.989 2.8e-12

Likelihood ratio test=42.5  on 2 df, p=5.84e-10  n= 137

> temp <- rbind(veteran, veteran)
>                       # in SAS: data temp; set veteran veteran;
> fit2 <- coxph(Surv(futime, status) ~ rx + karno + cluster(id),
                data=temp, method='breslow')
> fit2

         coef exp(coef) se(coef) robust se      z       p
   rx  0.1736     1.190  0.12946   0.19233  0.903 3.7e-01
karno -0.0338     0.967  0.00359   0.00483 -6.989 2.8e-12

Likelihood ratio test=85  on 2 df, p=0  n= 272

> coxme(Surv(futime, status) ~ rx + karno, random= ~1|id,
             data=temp, ties='breslow')

Fixed effects: Surv(futime, status) ~ rx + karno
```

```
            coef exp(coef)   se(coef)     z       p
   rx  0.8404165 2.3173319 0.63830866  1.32 1.9e-01
karno -0.1358233 0.8729969 0.01586364 -8.56 1.1e-16

Random effects:  ~ 1 | id
                id
Variance: 7.428457
```

Multiple Event Cox Models

First: are the events ordered or unordered?

- Ordered events: First, second, third, . . . infection
- Unordered events: Time to death, time to transplant
  Competing risks
  Correlated subjects, such as family groups

Unordered events

- Each observation is entered into the data set as a single line, no differently than it would be if the correlated observations were not present.
- Time scale questions are just as for any single event model (time from diagnosis versus time from enrollment versus . . .).
- If multiple events are of different types, then each event type may be placed into a separate strata.

Single strata examples

- Correlated data from families, where each subject has “ordinary” single event survival data.

- Multiple concurrent measurements on a subject, if they are all of similar type (rarely). Adverse effects might be an example.

Diabetic Retinopathy Trial

Between 1972 and 1975 seventeen hundred forty-two patients were enrolled in the study to evaluate the efficacy of photo coagulation treatment for proliferative diabetic retinopathy; photocoagulation was randomly assigned to one eye of each study patient, with the other eye serving as an untreated control. A major goal was to assess whether treatment significantly delayed the onset of severe visual loss.

The data below is for the 197 subjects who form a high risk subset.

What model to fit?

- Hazard is equal for left and right eyes. (Stratify on left/right, dominant/non-dominant, lighter/darker shade, . . . ?) No stratification based on eye or outcome.
- When eye X fails, the appropriate risk set is *all* eyes that have not yet failed.
- Hazard for the left eye is independent of the status of the right eye, given all appropriate covariates and the right model. Since we never have this, the observations will be correlated.

Combine all of the data into a single data set with $2n$ rows.

```
> dfit <- coxph(Surv(time, status) ~ adult + trt + cluster(id),
    data=diabetes)
> summary(dfit)

 n= 394
          coef exp(coef) se(coef) robust se      z    p
 adult  0.0539     1.055    0.162     0.179  0.302  .76
   trt -0.7789     0.459    0.169     0.149 -5.245 <.0001

Rsquare= 0.055   (max possible= 0.988 )
Likelihood ratio test= 22.5  on 2 df,   p=1.31e-05
Wald test            = 27.9  on 2 df,   p=8.94e-07
Score (logrank) test = 22.4  on 2 df,   p=1.4e-05,
              Robust = 26.4  p=1.89e-06
 (Note: the likelihood ratio and score tests assume independence of
 observations within a cluster, the Wald and robust score tests do not).
```

Because each subject received both treatments, the robust variance $\widetilde{D}'\widetilde{D}$ for treatment is *smaller* than the naive estimate. The effect is similar to a paired t-test.

SAS output

```
proc phreg data=diab.data3 covsandwich(aggregate);
    model time * status(0) = trt adult / ties=efron;

    id id;
```

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	22.4825	2	<.0001
Score	22.3596	2	<.0001
Modified Score	26.3492	2	<.0001
Wald	27.8672	2	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq
TRT	1	-0.77889	0.14847	0.879	27.5225	<.0001
ADULT	1	0.05388	0.17849	1.101	0.0911	0.7627

Variable	Hazard Ratio
TRT	0.459
ADULT	1.055

Random effect

```
> coxme(Surv(time, status) ~ adult + trt, random= ~1|id, diabetes)

Cox mixed-effects model fit by maximum likelihood
  Data: diabetes
  n= 394
  Iterations= 4 46
                    NULL Integrated Penalized
Log-likelihood -867.9858  -851.1137 -804.3837

  Penalized loglik: chisq= 127.2 on 74.75 degrees of freedom, p= 0.00015
 Integrated loglik: chisq= 33.74 on 3 degrees of freedom, p= 2.2e-07

Fixed effects: Surv(time, status) ~ adult + trt
             coef exp(coef)  se(coef)     z       p
adult  0.06070263 1.0625829 0.2105922  0.29 7.7e-01
  trt -0.90246282 0.4055696 0.1743538 -5.18 2.3e-07

Random effects:  ~ 1 | id
                id
Variance: 0.7979162
```

Random effects

- Conditional vs marginal
  - The estimate treatment effect is larger, this is expected
  - $E[\exp(X\beta)] > \exp(E[X\beta])$
  - The per-subject effect is larger than the population effect
  - The lower limit of the confidence interval is a about the same
- Testing
  - (33.7 on 3 df) - (22.5 on 2 df) = 11.2 on 1 df

Random effects (gamma frailty)

```
> coxph(Surv(time, status) ~ adult + trt + frailty(id), diabetes)

               coef se(coef)   se2  Chisq DF     p
     adult    0.041 0.221    0.166   0.03  1   .85
       trt   -0.911 0.174    0.171  27.31  1  <.0001
frailty(id)                        113.79 84   .017

Iterations: 6 outer, 24 Newton-Raphson
Variance of random effect= 0.851
Degrees of freedom for terms=  0.6  1.0 84.0
Likelihood ratio test=201  on 85.56 df, p=2.77e-11  n= 394
```

Parallel events, Multiple strata

Competing risks is the most common example

- A subject who is still at risk is assumed to be at risk for all of the event types, e.g., one can die without first having progression of disease.
- If there are k event types, then each subject will have k observations in the data set. (Often, $k-1$ of them will be censored).
- Usually the model will be stratified by event type. That is, we do not assume that cardiac death, stroke death, infectious death, . . . , have the same baseline hazard rate.

Long-term outcomes in MGUS

- “Monoclonal gammopathy of undetermined significance” (MGUS) is the presence of a monoclonal peak in the immunoglobulin profile of a subject, without evidence of overt disease
- Population prevalence increases at about 1.4%/decade from age 40 onward
- Detected incidentally, with an average of 16.5 years between incidence and detection in Olmsted County
- What is the prognostic significance?

Kyle (1993) studied all 241 cases of MGUS identified at the Mayo Clinic before 1/1/1971.

	Alive	Dead
multiple myeloma	2	37
amyloidosis	0	8
macroglobulinemia	1	6
lymphoproliferative	2	3
none	52	139
	57	184

Covariates are

- age (mean 63.4, range 34–90)
- size of M-spike (.3–3.2)
- hemoglobin (6.8–16.6)
- sex (140 male, 101 female)

We use two data sets; the first data set `mgus` has one observation per subject and corresponds to the usual “time to first outcome” analysis. The second data set `mgus2` allows for a competing risks analysis. It contains one stratum for each outcome type, with all 241 subjects appearing in each stratum. For simplicity, collapse into “death,” “multiple myeloma,” and “other.”

The first two subjects in the study experience death without progression at 760 days and lymphoproliferative disease at 2,160 days, respectively.

id	time	status	endpoint	sex	age	hgb	creat	mspike
1	760	1	death	2	79	11.5	1.2	2.0
1	760	0	myeloma	2	79	11.5	1.2	2.0
1	760	0	other	2	79	11.5	1.2	2.0
2	2160	0	death	2	76	13.3	1.0	1.8
2	2160	0	myeloma	2	76	13.3	1.0	1.8
2	2160	1	other	2	76	13.3	1.0	1.8

Time to first event:

```
>  coxph(Surv(time, status) ~ sex + age + mspike + hgb +
             cluster(id), data=mgus)

          coef exp(coef) se(coef) robust se      z       p
   sex -0.3494     0.705  0.15611   0.15277 -2.287 2.2e-02
   age  0.0516     1.053  0.00732   0.00725  7.119 1.1e-12
mspike -0.0945     0.910  0.18673   0.18292 -0.517 6.1e-01
   hgb -0.1669     0.846  0.04418   0.03506 -4.759 1.9e-06

Likelihood ratio test=76.8  on 4 df, p=7.77e-16  n=238
(3 observations deleted due to missing)
```

Competing risks, common coefficients:

```
> coxph(Surv(time, status) ~ sex + age + mspike + hgb
       + cluster(id) + strata(endpoint), data=mgus2)

          coef exp(coef) se(coef) robust se      z       p
   sex -0.3493     0.705  0.15611   0.15276 -2.287 2.2e-02
   age  0.0516     1.053  0.00732   0.00725  7.121 1.1e-12
mspike -0.0947     0.910  0.18673   0.18289 -0.518 6.0e-01
   hgb -0.1669     0.846  0.04418   0.03505 -4.761 1.9e-06

Likelihood ratio test=76.8  on 4 df, p=7.77e-16  n=714
(9 observations deleted due to missing)
```

If one assumes common coefficients, then the competing risks model has *identical* results to a time to first event model (other than approximations for ties).

Random Effects

A random effects model is not feasable for this data.

- There are n random coefficients b_i, we need more than one event per b to do estimation.
- In theory the model is identifiable with 1 obs/subject, but this is one case where “as n goes to infinity” really means *bigger than any n I’ll ever see*.

The advantage of the larger data set is that it allows for easy estimation of within-event-type coefficients.

Fitting separate models for each endpoint is the same as assuming all possible covariate * event type interactions.

Is the effect of age is identical for both outcomes, while controlling for a common effect of hemoglobin?

```
> age1 <- mgus2$age * (mgus2$endpoint=='death')
> age2 <- mgus2$age * (mgus2$endpoint!='death')
>  coxph(Surv(time, status) ~ hgb + age1 + age2
              + strata(endpoint), data=mgus2)

         coef exp(coef) se(coef)      z      p
 hgb -0.14057     0.869  0.04410 -3.187 0.0014
age1  0.07873     1.082  0.00928  8.486 0.0000
age2  0.00511     1.005  0.01212  0.421 0.6700
```

The median age of the study subjects is 64 years, and it is not surprising that age is a significant predictor of the overall death rate.

Age is, however, of almost no importance in predicting the likelihood of a plasma cell malignancy. (About 1%/year convert, regardless of age.)

If we look at the variable `mspike`, we will see that it is a major predictor of conversion to malignancy, with no effect on death rates.

Ursodeoxycholic Acid in the Treatment of Primary Biliary Cirrhosis

UDCA is a natural bile acid normally found in humans in minimal amounts, but which naturally occurs in bears. When given orally it becomes incorporated into the recycled pool of bile salts. The potential benefit of UDCA was thought to result from its hydrophilic properties, i.e., it is less hepatatoxic when retained in the liver than the more hydrophobic bile acids. Accordingly, several randomized trials have now been undertaken to assess its efficacy. Mayo Clinic coordinated a randomized double-blind trial of the drug starting in 1988, one hundred eighty patients were enrolled.

89	UDCA
91	Placebo
3	centers

Randomized from 5/1/88 to 5/31/92.

Endpoints

	Placebo	UDCA
Death	8	5
Transplant	8	7
Withdrawal	18	11
Drug toxicity	0	0
Histological progression	11	8
Development of varices	17	8
Development of ascites	4	1
Development of encephalopathy	1	3
Worsening of symptoms	9	6
Doubling of bilirubin	15	2

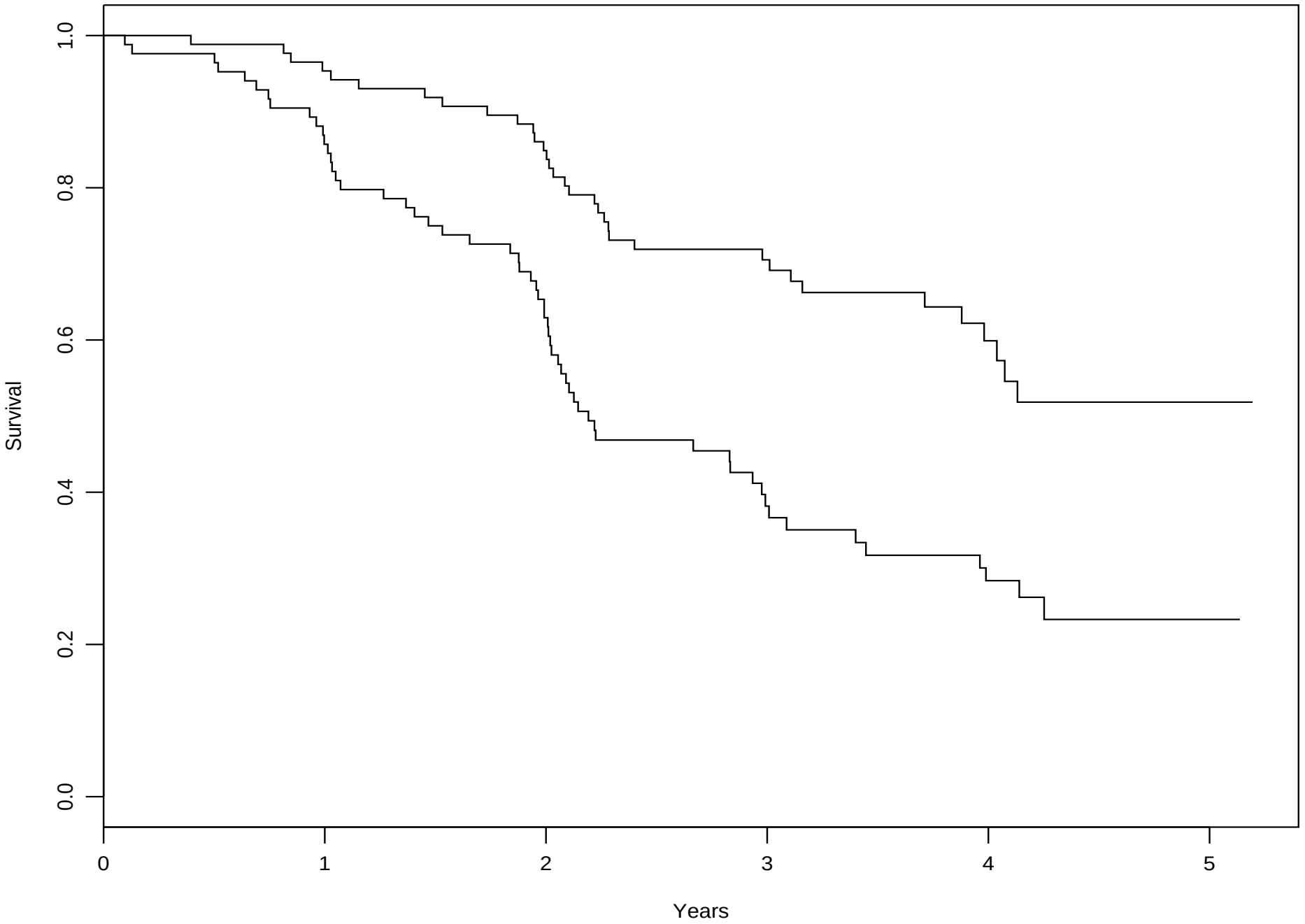
Total events

A decision was made between the study design and first analysis to not use voluntary withdrawal or doubling of bilirubin as endpoints, the first due to possible bias, the second due to laboratory variation.

	0	1	2	3	4
UDCA	63	17	6	3	0
Placebo	50	28	11	0	2

	First Events	Total Events	Gain
UDCA	26	38	12
Placebo	41	58	17

This is an apparent gain in “information” of 43%.


Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
Years

Choices for the UDCA data

- Time to first event

- Multiple events
  - Parallel events. It so happens that no one had 2 events of the same type, thus there is no ordering. At any given moment that a subject is at risk, any event could happen.
  - Separate strata per event type, as the baseline rates for each event are quite different. Also, some events can not happen twice (worsening of symptoms), so a subject may cease to be at risk for one event but not for another.

The time to first event data set has 180 observations.

The multiple data set has 180 * 7 = 1260 observations.

First event

```
> coxph(Surv(futime, status) ~ drug + age + log(bili) + cluster(id),
data=first)

      coef exp(coef) se(coef) robust se      z       p
     drug -0.92949     0.395   0.2597    0.2614 -3.556 3.8e-04
      age  0.00979     1.010   0.0127    0.0125  0.781 4.3e-01
log(bili)  0.64791     1.912   0.1509    0.1542  4.203 2.6e-05
```

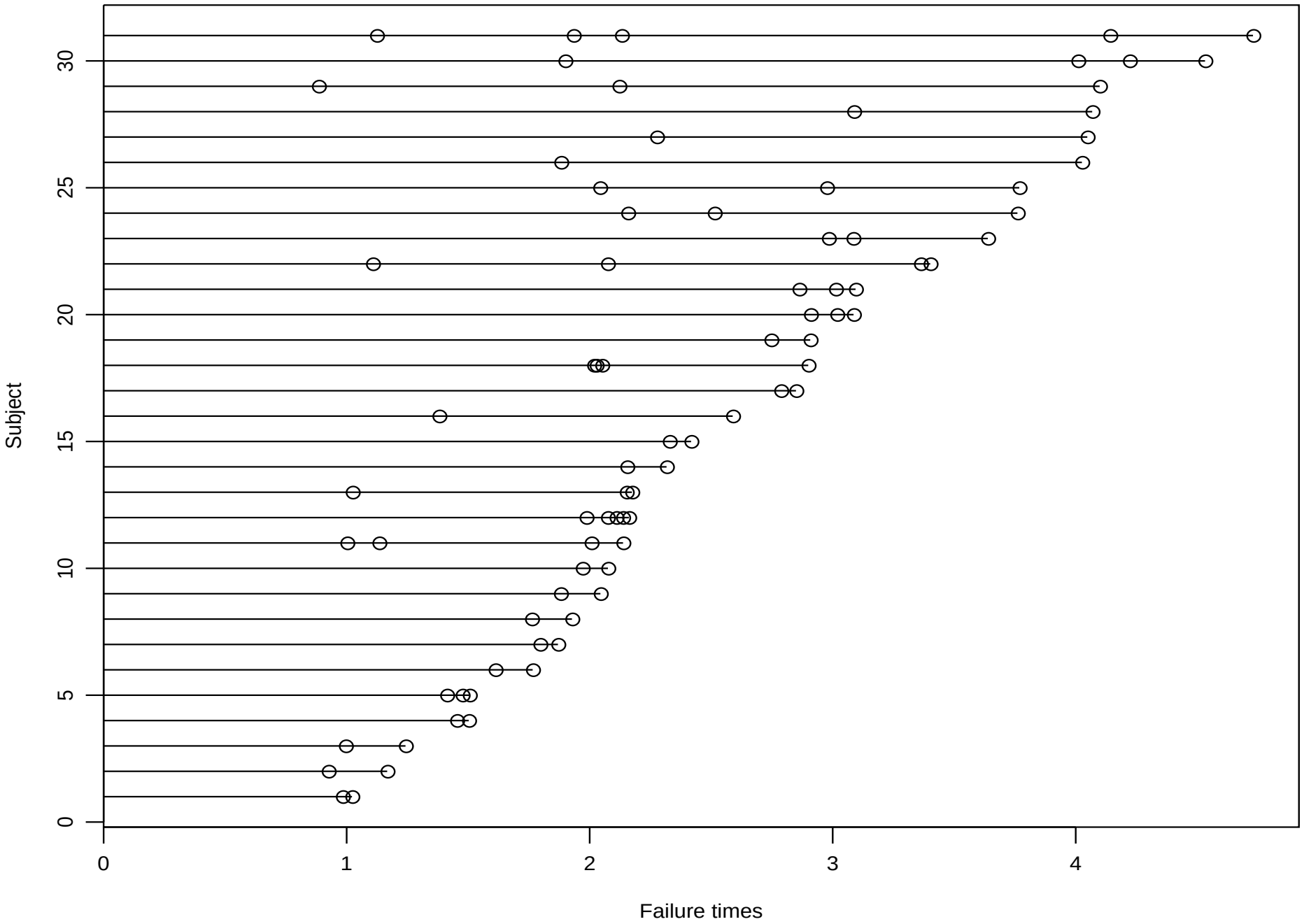
Multiple events, multiple strata

```
> coxph(Surv(futime, status) ~ drug + age + log(bili) +
cluster(id) + strata(event), data=multi)

      coef exp(coef) se(coef) robust se      z       p
     drug -0.80750     0.446   0.2167    0.2539 -3.181 1.5e-03
      age  0.00627     1.006   0.0109    0.0133  0.471 6.4e-01
log(bili)  0.66312     1.941   0.1221    0.1401  4.732 2.2e-06
```

The anticipated se for the second analysis is $.26\sqrt{67/96} = .22$. The actual gain is far less (effectively about 4 events).

- Patients were evaluated once a year.
- Adverse events were all assessed at evaluation, with the exception of death or transplant.
- All the endpoints are outcomes of a general degeneration in status.

Multiple event analysis does not always lead to gains.


Subject
0
5
10
15
20
25
30
0
1
2
3
4
Failure times

Colon cancer data

- 929 patients
  - 315 Observation
  - 310 Levamisole
  - 304 Levamisole + 5FU
- Time to death and progression for each subject
  - 423 No events
  - 92 One event
  - 414 Two events
- Up to 9 years of follow-up

Setup

Parallel (unordered) events. At any given moment when a subject is still at risk, either event could happen.

Separate strata for the two event types. There is no guarrantee that they have the same rate, or the same shape of baseline rate.

Data set will have $929 * 2 = 1858$ observations. The first 929 are the data set one would have built for a “time to death, ignoring progression” analysis, the next 929 similar for progression.

Time to death

```
> fitd <- coxph(Surv(time, status) ~ rx + extent+ node4, data=colon,
+       subset=(etype==1))
> fitd
             coef exp(coef) se(coef)     z       p
    rxLev -0.031     0.969   0.1071 -0.29 7.7e-01
rxLev+5FU -0.518     0.596   0.1187 -4.36 1.3e-05
   extent  0.538     1.713   0.1135  4.74 2.1e-06
    node4  0.845     2.328   0.0957  8.83 0.0e+00

Likelihood ratio test=127  on 4 df, p=0  n= 929
------------------------------
data temp; set colon;
    rx_lev = 1*(rx='Lev');
    rx_lev5= 1*(rx='Lev+5FU')
    if (etype=1);

proc phreg data=temp;
    model time * status(0) = rx_lev rx_lev5 extend node4/
ties=efron;
```

Marginal fit

```
> coxph(Surv(time, status) ~ rx + extent + node4 +
                              cluster(id) + strata(etype),
data=colon)

              coef exp(coef) se(coef) robust se      z       p
    rxLev -0.0362     0.964   0.0768    0.1056 -0.343 7.3e-01
rxLev+5FU -0.4488     0.638   0.0840    0.1168 -3.842 1.2e-04
   extent  0.5155     1.674   0.0796    0.1097  4.701 2.6e-06
    node4  0.8799     2.411   0.0681    0.0961  9.160 0.0e+00

------
data temp; set colon;
    rx_lev = 1*(rx='Lev');
    rx_lev5= 1*(rx='Lev+5FU')

proc phreg data=temp;
    model time * status(0) = rx_lev rx_lev5 extend node4/
ties=efron;
    strata etype;
```

Comparison

	Death	Progression	Combined
Levamisole vs. Obs	-0.031	-0.042	-0.036
Lev +5FU vs. Obs	-0.518	-0.379	-0.449
Extent of disease	0.538	0.493	0.516
> 4 Nodes	0.845	0.915	0.880

Interactions

One important issue to consider is strata by covariate interaction. Do we wish to assume that the effect of "> 4 nodes" is the same for the death and progression risks? Assume that node4 should be different

```
coxph(Surv(time, status) ~ rx + extent + cluster(id) +
node4* strata(etype),
data=colon)

                        coef exp(coef) se(coef) robust se      z       p
             rxLev -0.0364     0.964   0.0768    0.1056 -0.345 7.3e-01
          rxLev+5FU -0.4490     0.638   0.0840    0.1168 -3.844 1.2e-04
             extent  0.5154     1.674   0.0796    0.1097  4.700 2.6e-06
              node4  0.8460     2.330   0.0956    0.0994  8.512 0.0e+00
node4:strata(etype)  0.0688     1.071   0.1359    0.0534  1.287 2.0e-01

Likelihood ratio test=249  on 5 df, p=0  n= 1858
```

In SAS, simply code appropriate dummy variables.

WLW, computational note

In their paper, each strata was fit separately, and then the estimates are combined. Assume 3 strata, then

$$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$$

$$V = \begin{bmatrix} D_1'D_1 & D_1'D_2 & D_1'D_3 \\ D_2'D_1 & D_2'D_2 & D_2'D_3 \\ D_3'D_1 & D_3'D_2 & D_3'D_3 \end{bmatrix}$$

- This implicitly forces all strata*covariate interactions to be present.
- It is equivalent to fitting all of the data at once, followed by $V = \widetilde{D}'\widetilde{D}$.
- The latter is easier to program, and it allows for different codings of the interactions (in particular, leaving them out).
  It also allows the data event sets to be a different sizes, e.g., if one of the outcomes were missing for some subjects.

Ordered outcomes

Three main approaches

- Sequential events, no strata or “Andersen-Gill”
- Sequential events, stratifiy on event number: conditional or “Prentice, Williams, Peterson”
- Parallel events or “Wei, Lin and Weissfeld” (pretend it is competing risks).

Plus either a GEE variance or a random effect for the correlation.

Andersen-Gill model

"An event is an event is an event"

1. Time since entry or perhaps since diagnosis.

2. Counting process form is necessary; one observation per time interval or event.
   (0,t1), (t1,t2), . . .

3. The model may have strata, but they have no relation to the event history, e.g., stratify by gender.

Closest in spirit to Poisson regression

Conditional risk sets

Essentially the AG setup, stratified by event number.

1. Each event number or event type is a separate strata.

2. A subject is not at risk for a third event until they have had a second event. Both “time since entry” and “time since last event” are commonly used time scales.

3. Time-dependent covariates are as in a normal Cox model.

4. The counting process form will be needed unless we use time from last event.

Once my first choice, then my last, now in the middle.

WLW

Treats an ordered data set as though it were an unordered, competing risks problem.

1. Each event or event type is in its own strata.

2. Within strata 3, the data is “what I would have if the data recorder ignored all information except event type 3”.
   All patients appear in all strata.

3. In their paper WLW include all the strata*covariate interactions. This is not necessary.

4. Normally all time intervals start at 0.

Strata

Assume a subject with events at t_1, t_2, t_3 and no further follow up.

	Interval	Strata
AG	$(0, t_1]$	1
	$(t_1, t_2]$	1
	$(t_2, t_3]$	1
WLW	$(0, t_1]$	1
	$(0, t_2]$	2
	$(0, t_3]$	3
conditional	$(0, t_1]$	1
	$(t_1, t_2]$	2
	$(t_2, t_3]$	3
conditional	$(0, t_1]$	1
	$(0, t_2 - t_1]$	2
	$(0, t_3 - t_2]$	3

Risk sets

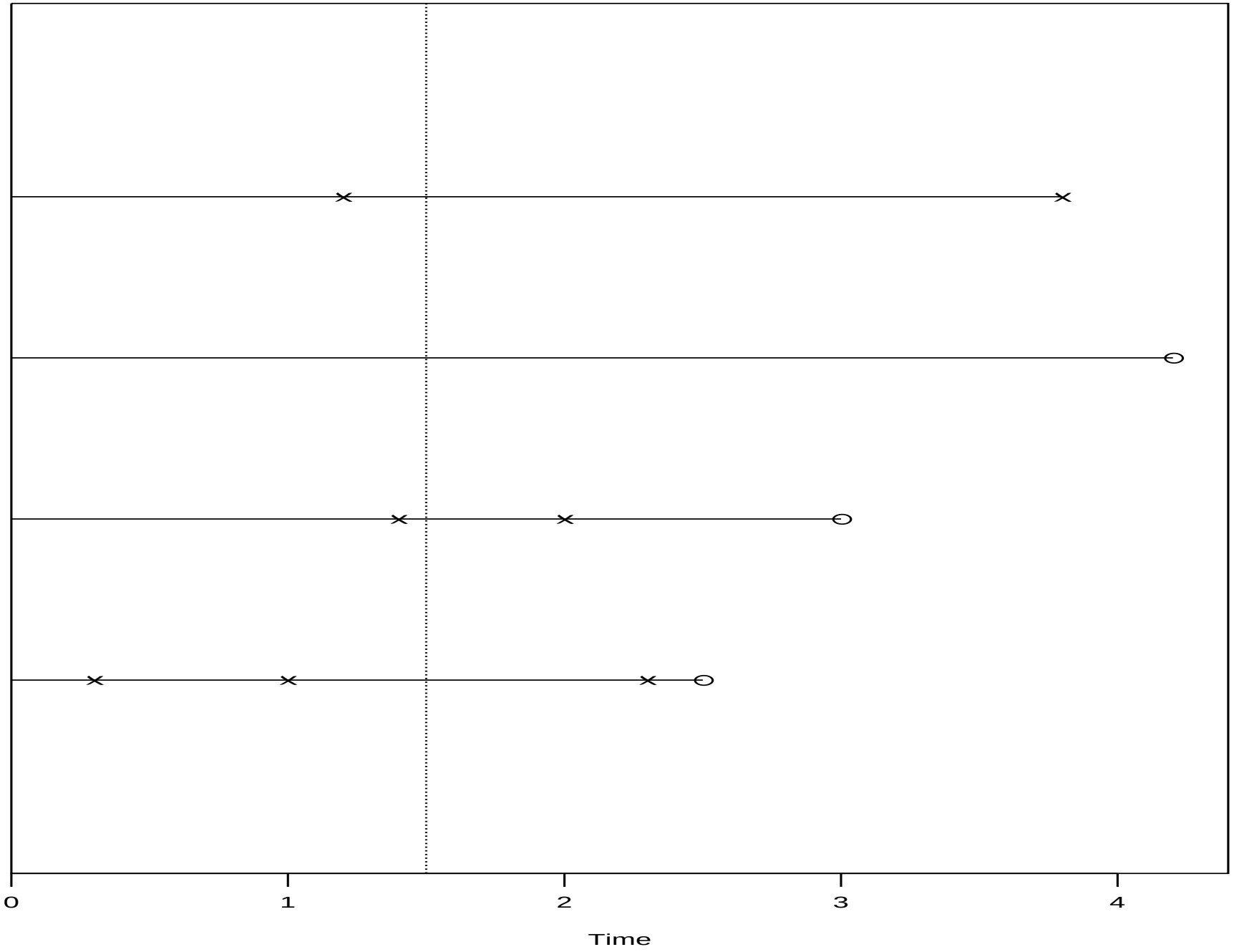
Subject Smith has experienced his second event on day 32. Who is at risk?

AG All subjects who were under observation on day 32.

WLW Subjects who were under observation on day 32,
and have not yet had a second event.

Cond Subjects who were under observation on day 32,
who have not yet had a second event,
but have experienced a first event.

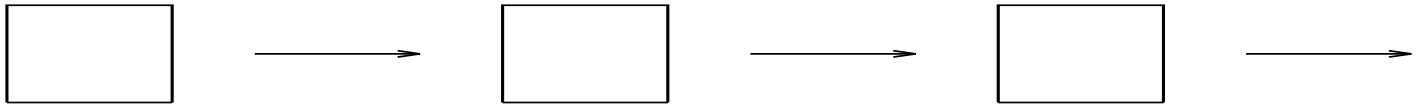
Ordered multiple events



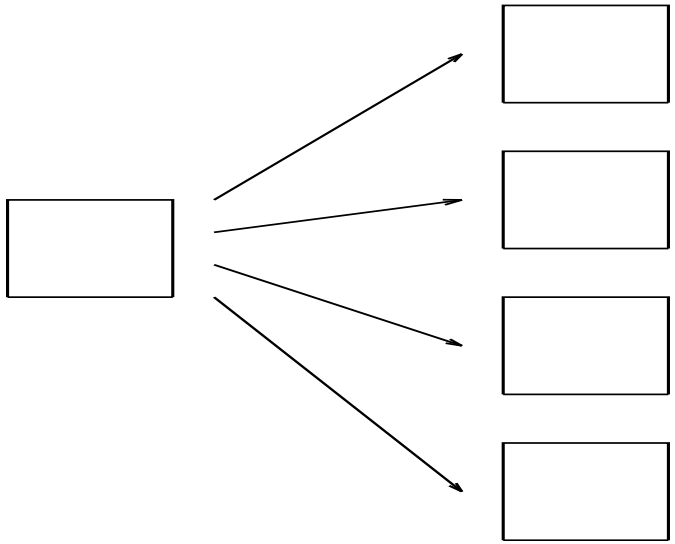
How many are at risk at time 1.5?

- Andersen-Gill: 4
- WLW:
  - Strata 1: 1
  - Strata 2: 3
- Conditional:
  - Strata 1: 1
  - Strata 2: 2

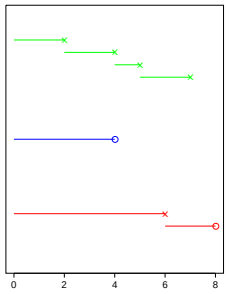
Data 1: Semi-Markov

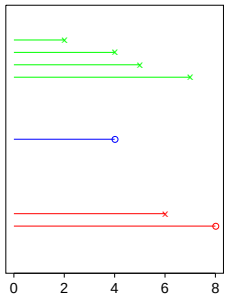


Data 2: Competing risks



Ordered multiple events





T_1	T_2	status	enum	time	status	enum
0	2	1	1	2	1	1
2	4	1	2	4	1	2
4	5	1	3	5	1	3
5	7	1	4	7	1	4
0	4	0	1	4	0	1
				4	0	2
				4	0	3
				4	0	4
0	6	1	1	6	1	1
6	8	0	2	8	0	2
				8	0	3
				8	0	4

	Data 1	Data 2
Stratified	PWP (conditional)	WLW (competing risks)
Not Stratified	Andersen-Gill (independent increments)	Lee, Wei, Amato

Andersen-Gill:

```
> fit <- coxph(Surv(time1, time2, status) ~ rx + age +
   cluster(id), data=data1)

proc phreg data=data1 covsandwich(aggregate);
    model (time1,time2) * status(0) = rx age;
    id id;
```

PWP or conditional: Add a `strata` statement to the above.

WLW:

```
> fit <- coxph(Surv(time, status) ~ rx + age +
   cluster(id) + strata(enum), data=data2)

proc phreg data=data2 covsandwich(aggregate);
    model time * status(0) = rx age;
    strata enum;
    id id;
```

Risk sets

For any setup make sure that the risk sets are correct.

- For each strata separately: if at some time t the subject is at risk then they are in the risk set.
  - ⋆ "At risk": if the subject were to have an event at t, my analysis would count it as an event in this strata.
  - ⋆ The WLW setup breaks this rule
- For any time t and any given strata, there is at most one copy of any subject.

Hidden Covariate Example

- $N = 2000$, large enough to not need simulation.
- True hazard for a subject is $\lambda(t) = \exp(x_1 - x_2)$.
- X_1 is uniform(-2,2) and unknown.
- X_2 is the 0/1 treatment variable, known.
- All observations have the same total follow-up.
- We purposely have chosen that the hidden variate have a larger effect than treatment.
- The true coefficients are +1 and -1.

	Number of events							
	0	1	2	3	4	5	6	7
Control	367	312	174	88	39	13	5	2
Treatment	680	250	50	14	6	0	0	0

The AG model treats the data as a Poisson process, which this data is. Therefore, with all the covariates in the model the AG setup should correctly estimate both the coefficient and the variance. More importantly, what happens if the covariate is not included?

The conditional model is also correct if all of the covariates are included.

A parallel analysis is simply wrong, but how far off is it?

		coef	model var	robust var
AG	without covariate			
	β_2	-0.92	0.066	0.084
	with covariate			
	β_2	-0.93	0.066	0.066
	β_1	1.05	0.056	0.056
Cond	without covariate			
	β_2	-0.67	0.070	0.068
	with covariate			
	β_2	-0.91	0.073	0.069
	β_1	1.03	0.065	0.064
WLW	without covariate			
	β_2	-1.23	0.066	0.113
	with covariate			
	β_2	-1.60	0.069	0.117
	β_1	1.82	0.063	0.113

AG model

- The coefficients and variances are correct when the variate is included.
- The robust variance estimate $\widetilde{D}'\widetilde{D}$ agrees with the information matrix variance $\mathcal{I}$ in this case.
- When the model is incorrect, the estimate of β_2 is still very good.
- The usual variance is an underestimate when an important covariate is missing from the model.

Conditional risk sets

- When the covariate is included, the model works well and is just slightly less efficient than the “main effects” or AG model.
- When important covariates are missing the model fails badly. Why?
- There is a progressive loss of balance in the later strata.
  - ⋆ To enter strata 2 a patient must experience an event.
  - ⋆ Because treatment is effective, the treated patients must have, on average, a worse baseline risk in order to have had an event.

	Mean covariate level			
	strata1	strata2	strata3	strata4
Control	.004	.17	.37	.54
Treatment	.005	.23	.43	.61

In a randomized trial, strata 1 will be randomized over the covariates.

Strata 2 will not be a randomized trial.

Long experience has shown that in non-randomized trials, the inclusion of all appropriate covariates is essential (and perhaps impossible).

WLW approach

The WLW model overestimates β_2, and inclusion of the hidden covariate makes the fit even worse.

In this case, the problem is a failure of proportional hazards.

For this particular setup we can work out the details in closed form.

If the true hazard ratio of treatment/control is λ_1/λ_2, then the WLW data set for event k have an initial ratio of $(\lambda_1/\lambda_2)^k$, which slowly dies down to an asymptotic value of λ_1/λ_2.

Models that include a treatment by covariate interaction confirm the prior comments.

	rx1	rx2	rx3	rx4	rx5	rx6	rx7
WLW	-0.99	-1.7	-2.1	-2.3	-4.2	-4.2	-4.2
conditional	-0.99	-0.9	-0.6	-0.3	$-\infty$	NA	NA

The conditional model only has 6 treated subjects remaining in strata 5, none of which had a further event during the time period. Thus, the infinite hazard ratio is simply an unstable estimate due to small sample size. Strata 6 and 7 have no treated subjects at all so the hazard ratio cannot be estimated.

A test for hidden covariates

H_0: All important covariates are in the model.

Test:

- Is it a simulation study?

  1. Yes: Read the paper or ask the author.
  2. No: Reject H_0.

The test has $\alpha = 0$ and power≈ 1, at least for clinical data.

Hidden Covariate Data

- 250 replications of the simulation
  - ⋆ 200 subjects
  - ⋆ Two variables: treatment (binomial) and x (Uniform(-1,1))
  - ⋆ True coefficients of +1, -1
  - ⋆ Exponential time between events, max of 10 events/subject
- Fit models with and without x to each data set
- Tabulate the results

Goals

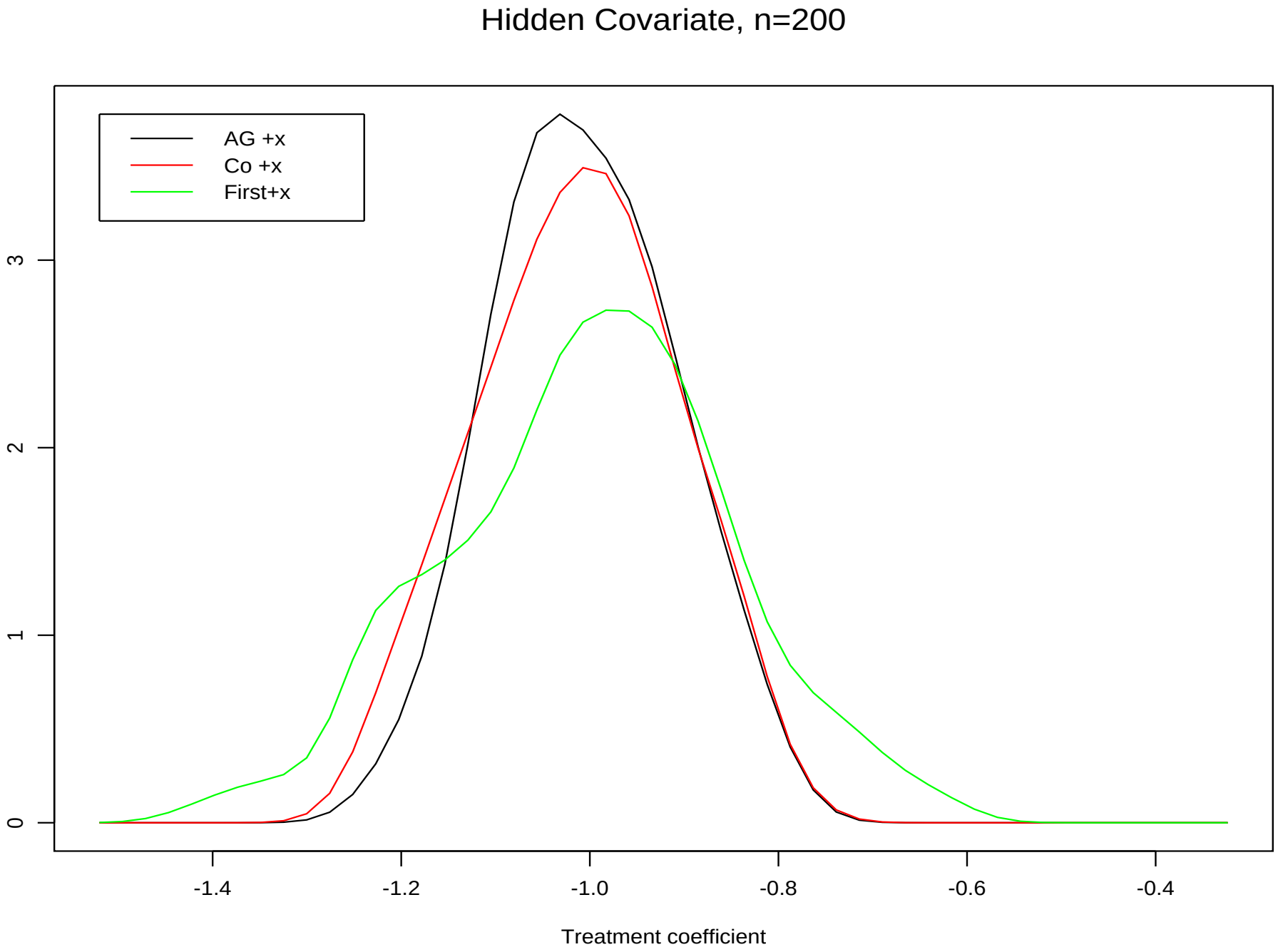
The simulation shows 4 things:

1. The prior results were based on a single realization — we could have been unlucky. This shows that the conclusions are true in general.

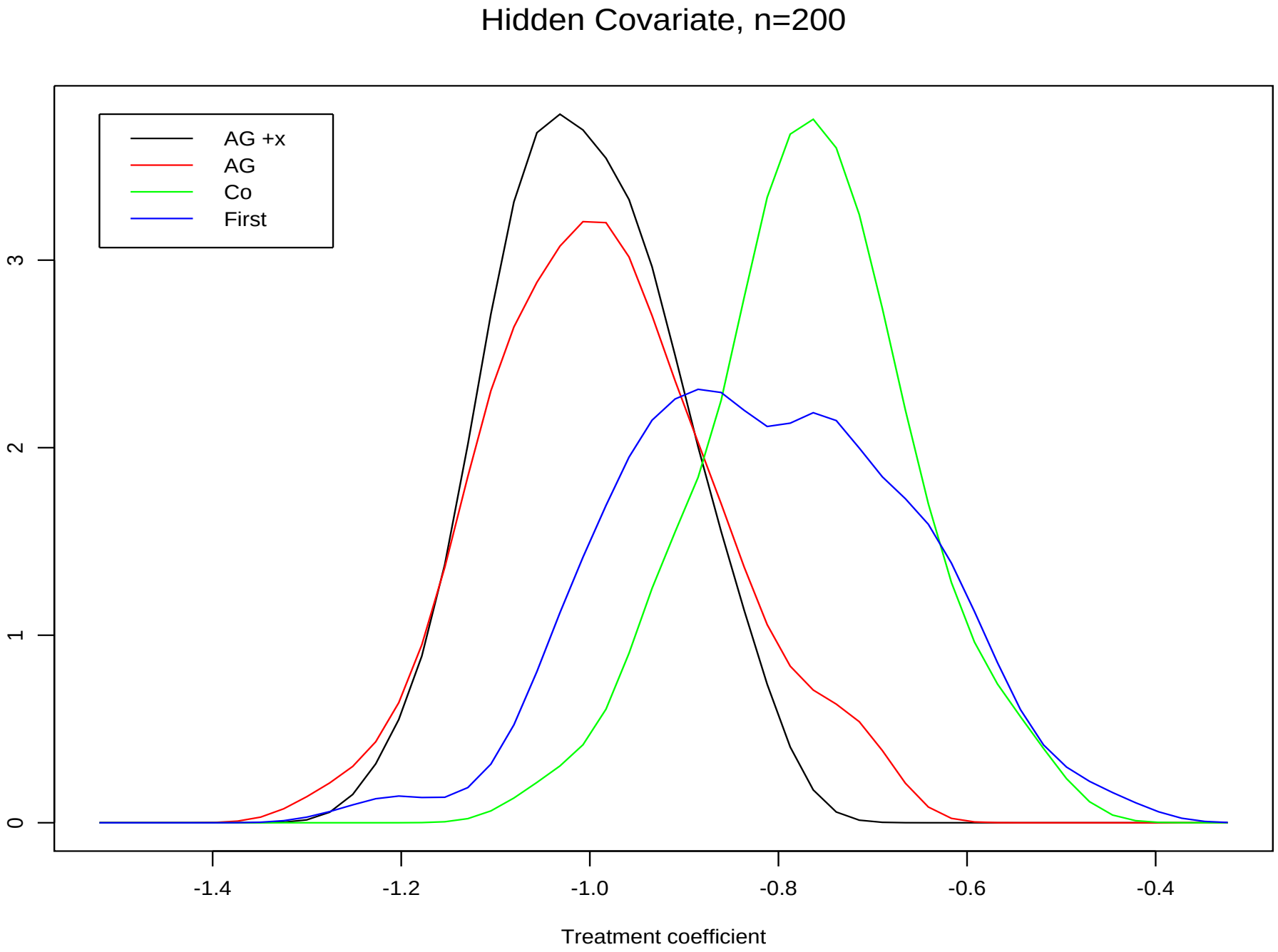
2. The model using only the first event is not unbiased!

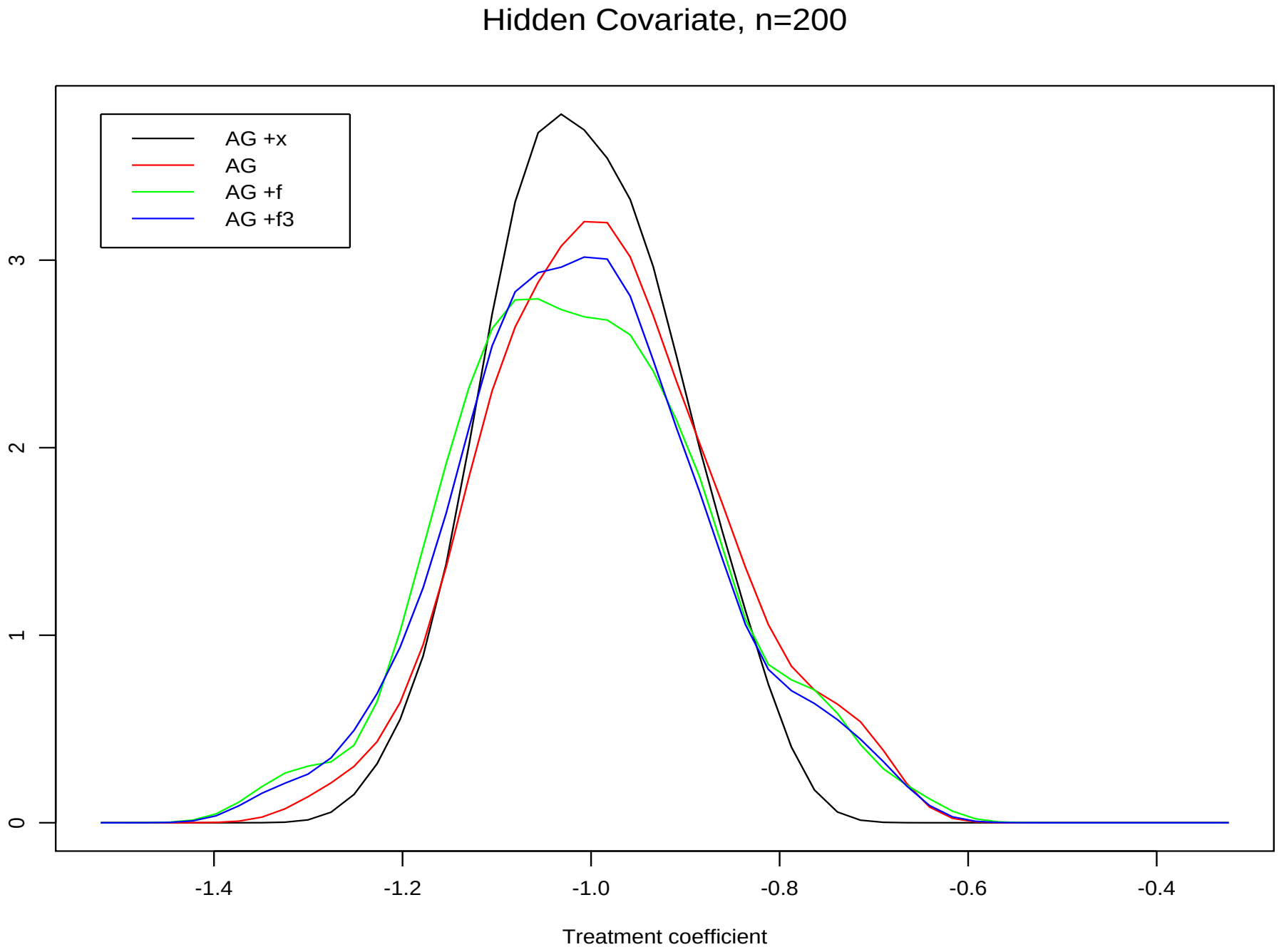
3. The conditional model has a substantial bias if there are any “hidden” covariates.
   - We can think of the hidden covariate as a latent trait, i.e. a random effect.
   - Perhaps a random effects model will repair the damage?
   - Do we need to have the right distribution for b.

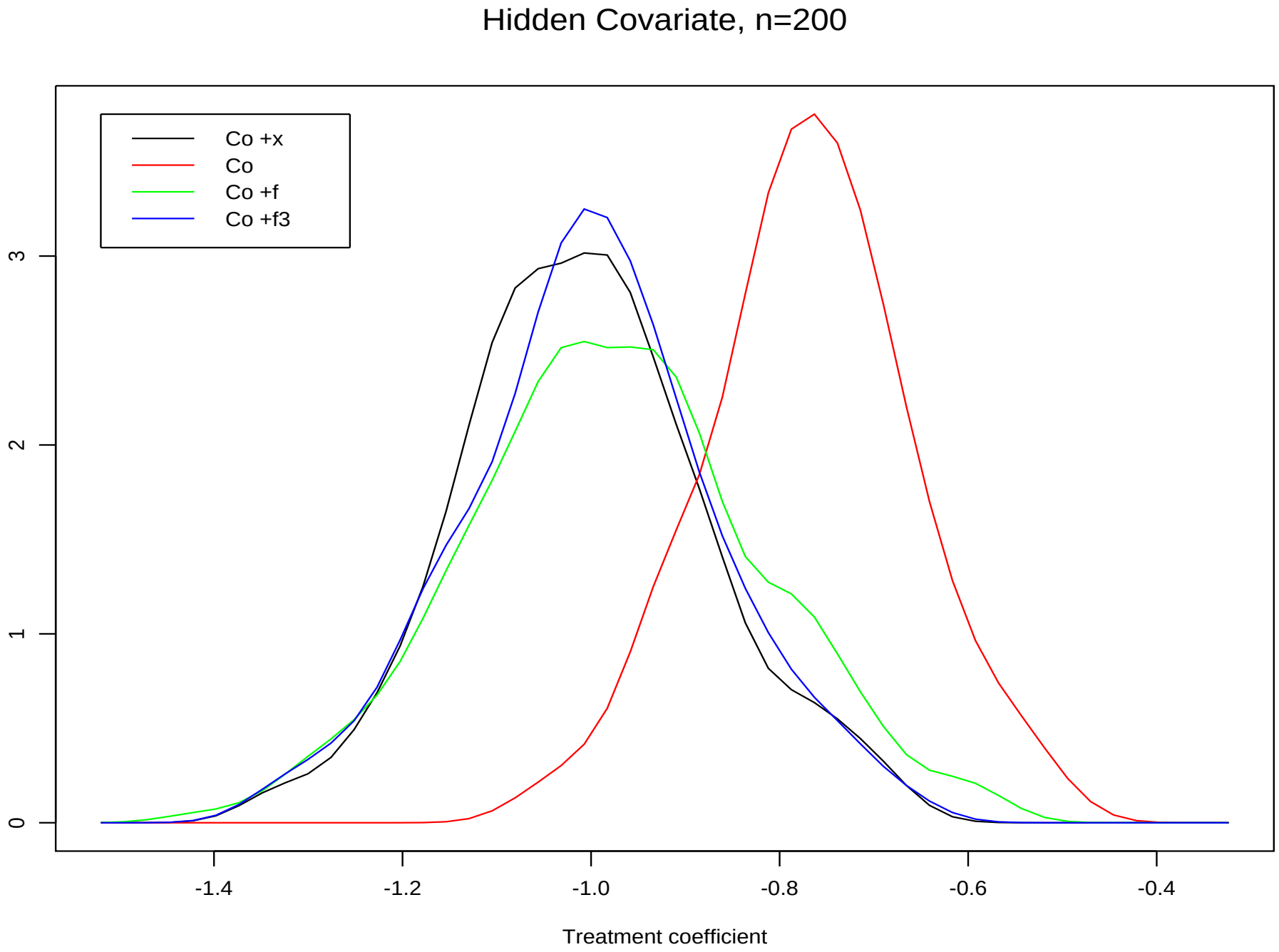
4. Substantial bias is removed from β even if the variance of b is not well estimated.

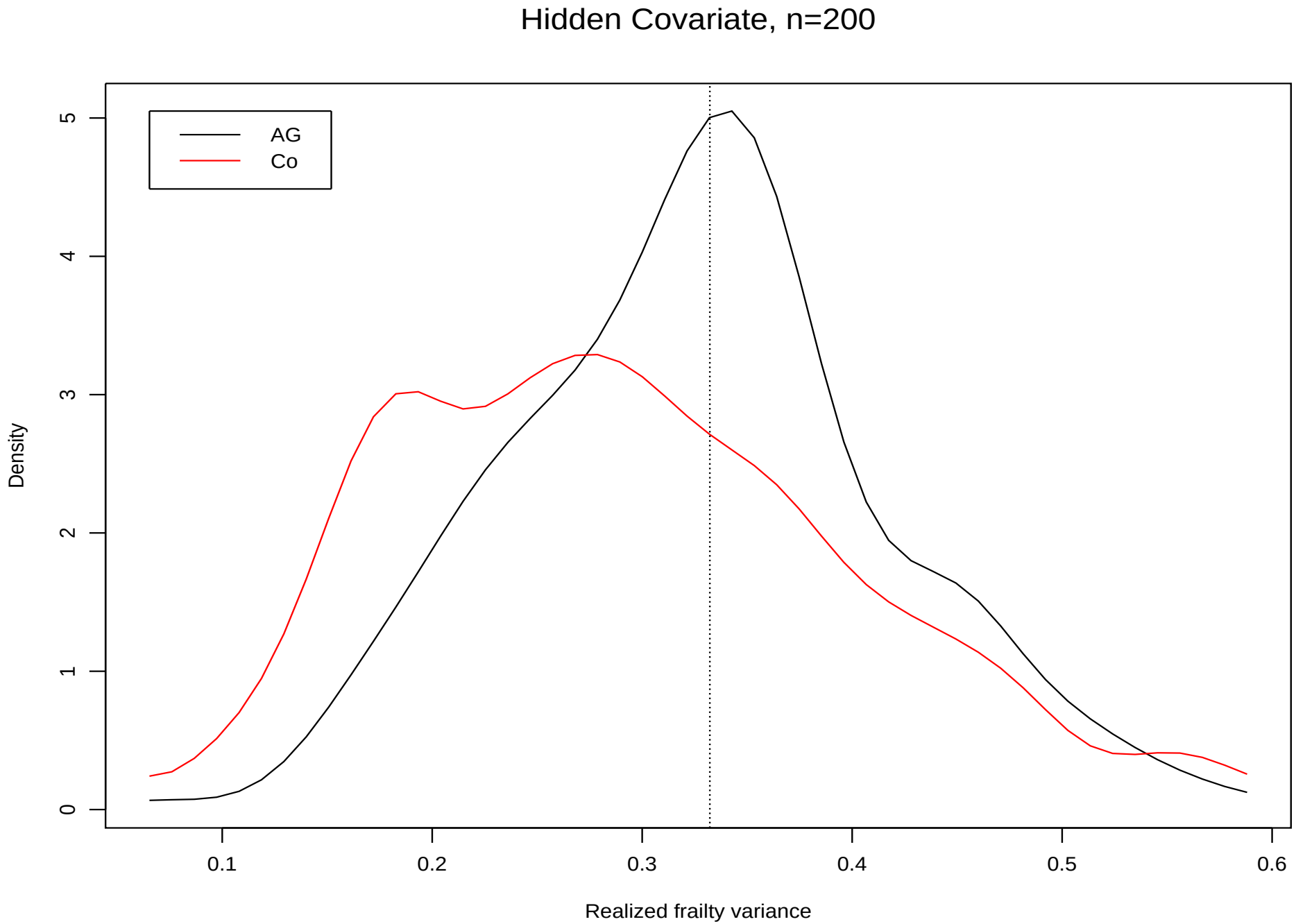

Hidden Covariate, n=200
AG +x
Co +x
First+x
0
1
2
3
-1.4
-1.2
-1.0
-0.8
-0.6
-0.4
Treatment coefficient

- The best model is Andersen-Gill, when the hidden covariate x is known
  - ⋆ It uses all of the data
  - ⋆ It uses the simplest model
  - ⋆ It is unbiased about the true coef -1
- The conditional model pays a small price in efficiency
- The simple Cox model is more variable: there is an average of approx 1.6 events/subject.


Hidden Covariate, n=200
AG +x
AG
Co
First
0
1
2
3
-1.4
-1.2
-1.0
-0.8
-0.6
-0.4
Treatment coefficient


Hidden Covariate, n=200
AG +x
AG
AG +f
AG +f3
0
1
2
3
-1.4
-1.2
-1.0
-0.8
-0.6
-0.4
Treatment coefficient


Hidden Covariate, n=200
Co +x
Co
Co +f
Co +f3
0
1
2
3
-1.4
-1.2
-1.0
-0.8
-0.6
-0.4
Treatment coefficient


Hidden Covariate, n=200
AG
Co
Density
0
1
2
3
4
5
0.1
0.2
0.3
0.4
0.5
0.6
Realized frailty variance

	Andersen–Gill	Conditional
0–.1	9	921
.1–.2	106	14
.2–.25	125	5
.25–.3	190	7
.3–.35	197	10
.35–.4	152	5
.4–.5	160	8
.5–.6	42	11
.6+	9	18

Table 1: Distribution of the estimated frailty variance θ in 1,000 simulations

Bladder Cancer Study

From Wei, Lin, and Weissfeld, JASA 1989.

86 subjects randomized to thiotepa or placebo.

Up to 4 recurrence times are recorded for each patient. There may be follow-up beyond the last occurrence.

	Number of Recurrences				
	0	1	2	3	4
Number of Subjects	39	18	7	8	14
Follow-up after last event	38	17	5	6	12

One subject has no events and 0 months of follow-up. Since he adds nothing to $\hat{\beta}$, the observation has been excluded from the rest of the analysis.

First data set

Input data

rx	futime	number	size	recurrences
1	0	1	1	
1	1	1	3	
1	4	2	1	
1	7	1	1	
1	10	5	1	
1	10	4	1	6
1	14	1	1	
1	18	1	1	
1	18	1	3	5
1	18	1	1	12, 16

The (start, stop] data set is

id	time1	time2	status	rx	number	size	enum
1	0	1	0	1	1	3	1
2	0	4	0	1	2	1	1
3	0	7	0	1	1	1	1
4	0	10	0	1	5	1	1
5	0	6	1	1	4	1	1
5	6	10	0	1	4	1	2
6	0	14	0	1	1	1	1
7	0	18	0	1	1	1	1
8	0	5	1	1	1	3	1
8	5	18	0	1	1	3	2
9	0	12	1	1	1	1	1
9	12	16	1	1	1	1	2
9	16	18	0	1	1	1	3

Second data set

The data set for a WLW-style analysis has $85 * 4 = 340$ observations. (The observations in strata 5 would not matter even if they were added on, since there are no events in strata 5).

Input data

rx	futime	number	size	recurrences
1	0	1	1	
1	1	1	3	
1	4	2	1	
1	7	1	1	
1	10	5	1	
1	10	4	1	6
1	14	1	1	
1	18	1	1	
1	18	1	3	5
1	18	1	1	12, 16

The WLW style data set

id	time	status	rx	number	size	enum
1	1	0	1	1	3	1
1	1	0	1	1	3	2
1	1	0	1	1	3	3
1	1	0	1	1	3	4
5	6	1	1	4	1	1
5	10	0	1	4	1	2
5	10	0	1	4	1	3
5	10	0	1	4	1	4
9	12	1	1	1	1	1
9	16	1	1	1	1	2
9	18	0	1	1	1	3
9	18	0	1	1	1	4

```
*
* Read in the bladder data set
*   drop the one useless subject who has no follow-up time
* ;

libname save 'sasdata';

data temp;
    infile 'data.bladder' missover;
    retain id 0;

    input rx  futime number size  r1-r4;
    if (futime =0) then delete;
    id = id +1;
```

```
* Anderson-Gill style data;

data save.bladder1;
    set temp;
    drop futime r1-r4;

    time1 =0;
    enum  =0;

    if (r1 ne .) then do;
time2 = r1;
status= 1;
enum  = 1;
output;
time1 = r1;
end;

    if (r2 ne .) then do;
time2 = r2;
status= 1;
enum  = 2;
output;
time1 = r2;
end;
```

Ordered multiple events

```
if (r3 ne .) then do;
time2 = r3;
status= 1;
enum  = 3;
output;
time1 = r3;
end;

    if (r4 ne .) then do;
time2 = r4;
status= 1;
enum  = 4;
output;
time1 = r4;
end;

    if (futime > time1) then do;
time2 = futime;
status= 0;
enum  = enum +1;
output;
end;
```

```
* data set for a Wei, Lin, and Weissfeld type anal :
*   # obs = #subjects * 4
* ;
data temp3;
    set save.bladder1;
    by id;
    drop temp time1 time2;
    futime = time2;
    if (enum <5) then output;
    if (last.id =1) then do;
temp = enum +1;
do enum = temp to 4;
    status =0;
    output;
    end;
end;
data save.bladder2; set temp3;
    rx1 = rx * (enum=1);     *special indicator variables for later;
    rx2 = rx * (enum=2);
    rx3 = rx * (enum=3);
    rx4 = rx * (enum=4);

proc print;
    var id futime status rx rx1 rx2 rx3 rx4 size number enum;
```

One “feature” of the data set can be seen by doing a fit with `enum` as a factor.

```
> coxph(Surv(time1, time2, status) ~ rx + size +
    number + factor(enum), data=bladder1)

                   coef exp(coef) se(coef)       z       p
          rx  -0.27994  7.56e-01   0.2058 -1.3605 1.7e-01
        size  -0.00375  9.96e-01   0.0703 -0.0533 9.6e-01
      number   0.14034  1.15e+00   0.0514  2.7293 6.3e-03
factor(enum)2   0.58926  1.80e+00   0.2568  2.2948 2.2e-02
factor(enum)3   1.68045  5.37e+00   0.3024  5.5578 2.7e-08
factor(enum)4   1.33765  3.81e+00   0.3510  3.8108 1.4e-04
factor(enum)5 -12.26606  4.71e-06 239.3184 -0.0513 9.6e-01

Likelihood ratio test=61.7  on 7 df, p=6.98e-11  n= 190
Warning messages:
      Loglik converged before variable  7 ; beta may be infinite...
```

Notice the coefficient for `enum=5`: it is essentially infinite. What is happening?

SAS version of the prior fit:

```
data temp;
    set bladder1;
    enum2 = 1*(enum=2);
    enum3 = 1*(enum=3);
    enum4 = 1*(enum=4);
    enum5 = 1*(enum=5);

proc phreg data=temp;
    model (time1, time2) *status(0) = rx size number
enum2 enum3 enum4 enum5 / ties=efron;

---------------------------------
```

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	928.615	866.938	61.677 with 7 DF (p=0.0001)
Score			69.833 with 7 DF (p=0.0001)
Wald			51.770 with 7 DF (p=0.0001)

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
RX	1	-0.279944	0.20577	1.85096	0.1737	0.756
SIZE	1	-0.003751	0.07032	0.00284	0.9575	0.996
NUMBER	1	0.140337	0.05142	7.44916	0.0063	1.151
ENUM2	1	0.589260	0.25678	5.26605	0.0217	1.803
ENUM3	1	1.680455	0.30236	30.88953	0.0001	5.368
ENUM4	1	1.337645	0.35101	14.52240	0.0001	3.810
ENUM5	1	-16.266064	1768	0.0000846	0.9927	0.000

	Enum				
	1	2	3	4	5
Number of Subjects	85	46	27	20	12
Number of events	47	29	22	14	0

A problem line:

```
rx   futime   number  size    recurrences
  1       30       2     1       3    6    8   12
```

- There is a problem with the *construction* of the data set.

- I used the data in examples for 3+ years before noticing the problem.
  - ⋆ and then only when I fit the above model (one I don't recommend – I was just experimenting),
  - ⋆ and no, I had not read the (start, stop] data set as carefully as I should have.

- Because only the first four recurrences are recorded, patients who have had a fourth recurrence become immortal in the eyes of the model. Presumably these patients are at high risk; any follow-up after the fourth event will bias $\hat{\beta}$ back toward zero.

- The 12 observations with `enum=5` should be removed from the data set.

- WLW made the same mistake in their paper.
- SAS 8 manual says "...the fifth observation is not needed for the WLW model, but it is indispensable to the AG analysis."
- Even the experts get it wrong sometimes.
- Think through the problem.
- Check the data set for sanity.

This can be particularly important for structural issues in the data set, such as the above constraint on ≤ 4 events.

There are now 3 data sets

- bladder1: The counting-process style data set, using all patients (190 obs).
- bladder2: The competing-risks style data set (85*4= 320 obs).
- bladder3: A copy of bladder1, with the false “follow-up” after event 4 removed (178 obs).

Andersen-Gill models

```
> coxph(Surv(time1, time2, status) ~ rx + size + number,
          data=bladder1)

          coef exp(coef) se(coef)      z       p
    rx -0.4116     0.663   0.1999 -2.059 0.03900
  size -0.0411     0.960   0.0703 -0.584 0.56000
number  0.1637     1.178   0.0478  3.427 0.00061

Likelihood ratio test=14.7  on 3 df, p=0.00213  n= 190

> coxph(Surv(time1, time2, status) ~ rx + size + number,
   data=bladder3)

          coef exp(coef) se(coef)      z      p
    rx -0.4647     0.628   0.1997 -2.327 0.0200
  size -0.0437     0.957   0.0691 -0.632 0.5300
number  0.1750     1.191   0.0471  3.717 0.0002

   Likelihood ratio test=17.5  on 3 df, p=0.000553  n= 178
```

As predicted earlier, we see that inclusion of the strata 5 “data” biases the coefficients toward zero. (Data set `bladder1` will not be used again.)

Time to first event

```
> coxph(Surv(time2, status) ~ rx + size + number +
       cluster(id), data=bladder3, subset=(enum==1))

  coef exp(coef) se(coef) robust se      z      p
    rx -0.5259     0.591   0.3158    0.3152 -1.668 0.0950
  size  0.0696     1.072   0.1016    0.0886  0.786 0.4300
number  0.2382     1.269   0.0759    0.0746  3.194 0.0014

Likelihood ratio test=9.92  on 3 df, p=0.0193  n= 85
```

The three additive models

Andersen-Gill

```
> coxph(Surv(time1, time2, status) ~ rx + size + number
                 + cluster(id), data=bladder3)
          coef exp(coef) se(coef) robust se      z      p
    rx -0.4647     0.628   0.1997    0.2656 -1.750 0.0800
  size -0.0437     0.957   0.0691    0.0776 -0.563 0.5700
number  0.1750     1.191   0.0471    0.0630  2.775 0.0055
```

Wei, Lin and Weissfeld

```
> coxph(Surv(futime, status) ~ rx + size + number +
     strata(enum) + cluster(id), bladder2)

          coef exp(coef) se(coef) robust se      z      p
    rx -0.5848     0.557   0.2011    0.3079 -1.899 0.0580
  size -0.0516     0.950   0.0697    0.0946 -0.546 0.5900
number  0.2103     1.234   0.0468    0.0666  3.156 0.0016
```

Conditional risk sets

```
> coxph(Surv(time1, time2, status) ~ rx + size + number +
```

```
     cluster(id) + strata(enum), bladder3)

            coef exp(coef) se(coef) robust se      z    p
    rx -0.33349     0.716   0.2162    0.2048 -1.628 0.10
  size -0.00849     0.992   0.0728    0.0616 -0.138 0.89
number  0.11962     1.127   0.0533    0.0514  2.328 0.02
```

The coefficient ordering of the hidden covariate data set is repeated, although not nearly as strongly as we found there.

$$\text{WLW} > \text{Andersen-Gill} > \text{conditional}$$

A logical step is to examine a random effects model.

```
> coxme(Surv(time1, time2, status) ~ rx + size + number,
           data=bladder3, random=~1|id)

                    NULL Integrated Penalized
Log-likelihood -458.7393  -437.7371  -406.525

  Penalized loglik: chisq= 104.43 on 54.65 degrees of freedom, p= 5.8e-05
 Integrated loglik: chisq= 42 on 4 degrees of freedom, p= 1.7e-08

Fixed effects: Surv(time1, time2, status) ~ rx + size + number
               coef exp(coef)   se(coef)     z      p
    rx -0.56605544 0.5677606 0.32517601 -1.74 0.0820
  size -0.01838367 0.9817843 0.10930777 -0.17 0.8700
number  0.22922839 1.2576292 0.08673812  2.64 0.0082

Random effects:  ~ 1 | id
                id
Variance: 0.9930668
```

How much have we gained?

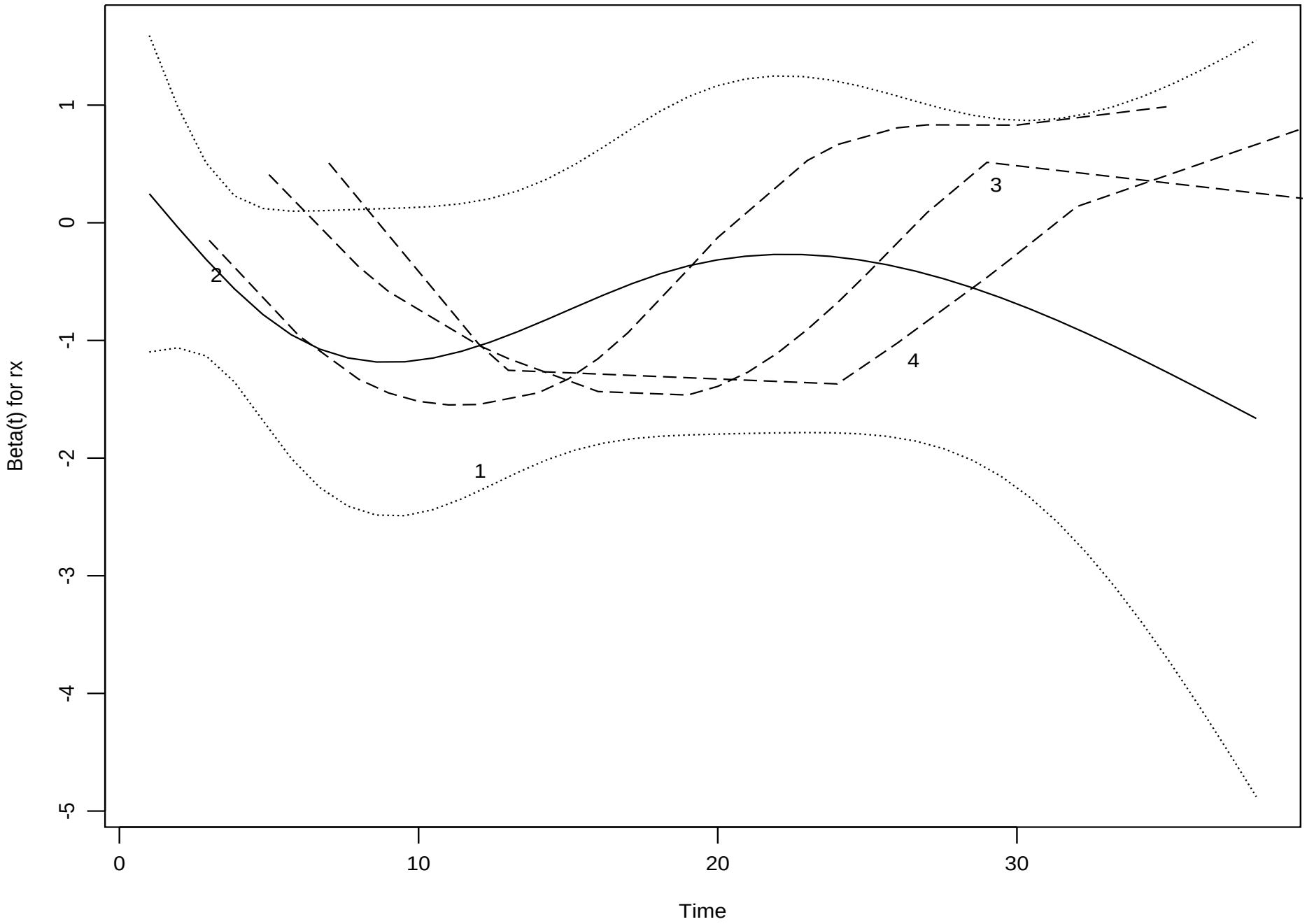
Comparing the A-G model to the first event model –

- There are 47 first events and 112 total events.
- The variance of $\hat{\beta}$ in a Cox model is proportional to 1/(number of events).
- This suggests a potential gain of $\sqrt{47/112} = .65$, or a 35% improvement in the standard error.
- The actual improvement in standard error is $.27/.32 = .84 \approx \sqrt{47/66}$. We 'gained' the reduction in standard error that 19 new (independent) events would have given us; each recurrent event is worth about (66-47)/(112-47)= .29 of a new subject.
- This comparison isn't quite fair since $\hat{\beta}$ differs between the models. (But is a good first approximation).

Testing proportional hazards in the WLW model

```
fit1 <- coxph(Surv(futime, status) ~ rx + size +number,
bladder2, subset=(enum==1))

plot(cox.zph(fit1, transform='identity'), resid=F, var=1)
abline(h= -.58)

for (i in 2:4)
    temp1 <- coxph(Surv(futime, status) ~ rx + size + number,
bladder2, subset=(enum==i))
    temp2 <- cox.zph(temp1, transform='identity')
    lines(smooth.spline(temp2$x, temp2$y[,1], df=4), col=i)


text(locator(4), c("1", "2", "3", "4"))
```

Ordered multiple events



WLW fit, all interactions

This duplicates the results in Wei, Lin and Weissfeld.

```
> coxph(Surv(futime, status) ~ rx1 + rx2 + rx3 + rx4 +
 (size + number)*strata(enum) + cluster(id), bladder2,
  method='breslow')
 coef exp(coef) se(coef) robust se      z      p
  rx1 -0.5176     0.596   0.3158    0.3075 -1.683 0.0920
  rx2 -0.6194     0.538   0.3932    0.3639 -1.702 0.0890
  rx3 -0.6999     0.497   0.4599    0.4152 -1.686 0.0920
  rx4 -0.6504     0.522   0.5774    0.4897 -1.328 0.1800
         size  0.0679     1.070   0.1012    0.0853  0.796 0.4300
       number  0.2360     1.266   0.0761    0.0721  3.274 0.0011
  size:enum=2 -0.1440     0.866   0.1680    0.1119 -1.287 0.2000
  size:enum=3 -0.2792     0.756   0.2086    0.1511 -1.847 0.0650
  size:enum=4 -0.2711     0.763   0.2515    0.1856 -1.460 0.1400
number:enum=2 -0.0984     0.906   0.1193    0.1144 -0.861 0.3900
number:enum=3 -0.0662     0.936   0.1298    0.1167 -0.567 0.5700
number:enum=4  0.0928     1.097   0.1466    0.1175  0.790 0.4300
```

Notice that the treatment effect get larger from strata 1 to 2 to 3.

WLW fit each of the 4 strata separately, then combine the 4 robust variance matrices to get an overall variance. It is easy to verify that we can get the same estimates and variance.

```
> round(fit$var[1:4,1:4], 3)
      1     2     3     4
1 0.095 0.060 0.057 0.044
2 0.060 0.132 0.130 0.116
3 0.057 0.130 0.172 0.159
4 0.044 0.116 0.159 0.240
```

Their overall test for significance of treatment is a 4 degree of freedom quadratic form:

```
> fit$coef[1:4] %*% solve(fit$var[1:4,1:4]) %*% fit$coef[1:4]
        [,1]
[1,] 3.96616
```

which is not significant.

Their overall estimate of treatment effect is $k\hat{\beta}$ where k, the average treatment effect, and its standard error are

```
> k <- solve(fit$var[1:4,1:4], c(1,1,1,1))
> k <- k / sum(k)
> round(k,3)
     1     2      3     4
 0.677 0.257 -0.076 0.142
> sum(k * fit$coef[1:4])
[1] -0.5487979
> sqrt(k %*% fit$var[1:4,1:4] %*% k)
[1,] 0.2852717
```

It is far easier, and I believe more correct, to simply fit a single treatment effect:

```
> coxph(Surv(futime, status) ~ rx + (size + number)*strata(enum)
+ cluster(id), bladder2,  method='breslow')

              coef exp(coef) se(coef) robust se      z       p
       rx -0.5976     0.550   0.2031    0.3029 -1.973 0.04800
     size  0.0695     1.072   0.1009    0.0847  0.821 0.41000
 ...
```

The WLW calculation is actually a 1-step approximate method for backwards stepwise regression, moving from the full model to this one.

Should one fit all interactions?

I think not. In most data sets we cannot afford to spend such a large number of degrees of freedom. The use of all interactions in the WLW paper is partly (mostly?) an artifact of the computational method chosen.

Conditional risk sets, all interactions

```
> coxph(Surv(time1, time2, status) ~ rx1+ rx2 +rx3 + rx4 +
(size + number)*strata(enum) + cluster(id), bladder3)

 coef exp(coef) se(coef) robust se       z      p
  rx1 -0.5259     0.591   0.3158    0.3152 -1.6683 0.0950
  rx2 -0.5038     0.604   0.4062    0.4543 -1.1091 0.2700
  rx3  0.1407     1.151   0.6731    0.4869  0.2889 0.7700
  rx4  0.0503     1.052   0.7917    0.5401  0.0932 0.9300
         size  0.0696     1.072   0.1016    0.0886  0.7854 0.4300
       number  0.2382     1.269   0.0759    0.0746  3.1934 0.0014
  size:enum=2 -0.2303     0.794   0.1591    0.1751 -1.3157 0.1900
  size:enum=3  0.0985     1.103   0.2876    0.1803  0.5461 0.5800
  size:enum=4 -0.0605     0.941   0.3538    0.3764 -0.1608 0.8700
number:enum=2 -0.2628     0.769   0.1176    0.1653 -1.5898 0.1100
number:enum=3 -0.1885     0.828   0.2003    0.1420 -1.3279 0.1800
number:enum=4 -0.0339     0.967   0.2537    0.1935 -0.1752 0.8600
```

In the WLW model the treatment effect grows slightly stronger over time, in the conditional one it changes sign!

SAS fit

```
proc phreg data=save.bladder2 covsandwich(aggregate);
    model futime * status(0) = rx size number
  / ties=efron;
    strata enum;
    id id;

------------------
                      The PHREG Procedure

                       Model Information

              Data Set                     SAVE.BLADDER2
              Dependent Variable           FUTIME
              Censoring Variable           STATUS
              Censoring Value(s)           0
              Ties Handling                EFRON
```

Summary of the Number of Event and Censored Values

Stratum	ENUM	Total	Event	Censored	Percent Censored
1	1	85	47	38	44.71
2	2	85	29	56	65.88
3	3	85	22	63	74.12
4	4	85	14	71	83.53
Total		340	112	228	67.06

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	25.2607	3	<.0001
Score	28.5996	3	<.0001
Modified Score	11.7522	3	0.0083
Wald	15.8977	3	0.0012

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq
RX	1	-0.58477	0.30328	1.508	3.7177	0.0538
SIZE	1	-0.05161	0.09318	1.336	0.3068	0.5797
NUMBER	1	0.21033	0.06560	1.403	10.2815	0.0013

Why do the parameter estimates agree with S-Plus, but the SEs differ slightly?

- SAS does not compute the dfbeta residuals correctly for an Efron approximation.
- This can be verified by computing dfbetas “by hand” for a few data points.
- Educated guess: dfbetas involve the cumulative hazard $\hat{\Lambda}(t)$, and SAS is using the Breslow estimate of that quantity. (It is easier to compute).
- Practically, the difference is trivial (in all examples I have looked at).

CGD data

A clinical trial of gamma interferon versus placebo in children with CGD, a genetic defect that leads to multiple recurrent infections. The data set is found in Fleming and Harrington.

From practical experience, clinical scientists conducting the rIFN-g trial suggested that the risk of recurrent infection remained constant regardless of the number of previous infections. This suggests use of an independent increments or A-G model.

Nevertheless, we will try all three.

The start-stop data set

```
                                      S
                                   I  t
    C    R            H      W   n  e  Fu
    e    a            e      e   h  r
    n    n            i      i   e  o  t
    t    d     S  A   g      g   r  i  i
I   e    o   R e  g   h      h   i  d  m
D   r    m   X x  e   t      t   t  s  e  ---- Infections ----------

1 204 082888 1 2 12 147.0  62.0 2 2 414 219 373
2 204 082888 0 1 15 159.0  47.5 2 2 439   8  26 152 241 249 322 350
3 204 082988 1 1 19 171.0  72.7 1 2 382
4 204 091388 1 1 12 142.0  34.0 1 2 388
5 238 092888 0 1 17 162.5  52.7 1 2 383 246 253
6 245 093088 1 2 44 153.3  45.0 2 2 364
-----------------------------------------
```

Ordered multiple events

ID	CENTER	TSTART	TSTOP	STATUS	ENUM	RX	AGE	...
1	204	0	219	1	1	1	12	
1	204	219	373	1	2	1	12	
1	204	373	414	0	3	1	12	
2	204	0	8	1	1	0	15	
2	204	8	26	1	2	0	15	
2	204	26	152	1	3	0	15	
2	204	152	241	1	4	0	15	
2	204	241	249	1	5	0	15	
2	204	249	322	1	6	0	15	
2	204	322	350	1	7	0	15	
2	204	350	439	0	8	0	15	
3	204	0	382	0	1	1	19	
4	204	0	388	0	1	1	12	

Ordered multiple events

```
data gamma;
    infile '../../../data/cgd.data' missover;
    input id 1-3 center 5-7 +1 random mmddyy6. rx 2. sex 2. age 3.
  height 6.1 weight 6.1 inherit 2. steroids 2. propylac 2.
  hos_cat 2. futime 4.  (event1-event7) (7*4.);

data cgd1; set gamma;   *counting process data set;
    drop event1-event7 futime;

    tstart = 0;
    enum  = 1;

    if (event1 NE .) then do;
tstop = event1;
status = 1;
output;
tstart = tstop;
enum = enum +1;
end;
```

```
    if (event2 NE .) then do;
tstop = event2;
status =1;
output;
tstart = tstop;
enum = enum +1;
end;
```

```
    .
    .
    .

    if (event7 NE .) then do;
tstop = event7;
status =1;
output;
tstart = tstop;
enum = enum +1;
end;

    if (futime > tstart) then do;
tstop = futime;
status =0;
output;
end;

proc print;
```

The `cgd1` data set has 203 observations.

A printout of the data set (or some portion of it) is critical. Errors in creating the data set are easy to make, and can lead to some bizarre analyses.

Creating the data set in S

```
cgd0 <- read.table('data.cgd', na.strings='.',
   col.names=c('id', 'center', 'rand.dt', 'rx', 'sex',
               'age', 'height', 'weight', 'inherit',
               'steroids', 'propylac', 'hos.cat',
               'futime', paste('event', 1:7, sep='')))

#Turn the randomization date into a date object (as it should be)
#  I have to insert "/" marks, since the timeDate function isn't
#  smart enough to do without them
temp <- cgd0$rand.dt
tempm<- floor(temp/10000)
tempd<- floor(temp/100)%%100
tempy<- 1900 + temp%%100
cgd0$rand.dt <- timeDate(charvec=paste(tempm, tempd, tempy, sep='/'),
 format="%d%b%C")
```

```
#Find the max event time for each subject, setting it to 0 for
#  those with no events
n <- nrow(cgd0)
etemp <- as.matrix(cgd0[,14:20])
maxtime <- apply(cbind(0,etemp), 1, max, na.rm=T)

#
# Create the WLW style data set
#  everyone gets 7 rows
#
tstop <- ifelse(is.na(etemp), cgd0$futime, etemp)
tstat <- ifelse(is.na(etemp), 0, 1)
cgd2 <- data.frame(cgd0[rep(1:n,7), 1:12],
   time = c(tstop),
   status=c(tstat),
   enum = rep(1:7, rep(n,7))
   )
```

A key idea in the above is duplicated subscripts,
`cgd0[c(1,1,1,1,2,2,4,4,4,4),]` is a data set with 4 copies of row 1 of cgd0,
2 copies of row 2, etc.

```
# Now, create the data Andersen-Gill style data set
# First, pretend that everyone had all 7 events + more follow-up
#  Then thin things out
tstart <- c(t(cbind(0, ifelse(is.na(etemp), maxtime, etemp))))
tstop  <- c(t(cbind(etemp, cgd0$futime)))
tstat  <- rep(c(1,1,1,1,1,1,1,0), n)
keep <- (!is.na(tstop) & tstop > tstart)  #which rows to keep
nrow <- apply(matrix(keep, nrow=8), 2, sum) #how many rows remain for each
enum <- apply(matrix(keep, nrow=8), 2, cumsum)

indx <- rep(1:n, nrow)
cgd1 <- data.frame(cgd0[indx, 1:12],
   tstart=tstart[keep],
   tstop =tstop [keep],
   status=tstat[keep],
   enum = enum[keep], row.names=NULL)

rm(tstart, tstop, tstat, keep, nrow, indx)
rm(etemp, maxtime, n, temp)
```

The competing-risks (WLW) data set

```
                                      S
                                   I  t
    C    R            H      W   n  e  Fu
    e    a            e      e   h  r
    n    n            i      i   e  o  t
    t    d     S  A   g      g   r  i  i
I   e    o   R e  g   h      h   i  d  m
D   r    m   X x  e   t      t   t  s  e  ---- Infections ----------

1 204 082888 1 2 12 147.0  62.0 2 2 414 219 373
2 204 082888 0 1 15 159.0  47.5 2 2 439   8  26 152 241 249 322 350
3 204 082988 1 1 19 171.0  72.7 1 2 382
4 204 091388 1 1 12 142.0  34.0 1 2 388
-----------------------------------------
```

ID	CENTER	TIME	STATUS	ENUM	RX	AGE	...
1	204	219	1	1	1	12	
1	204	373	1	2	1	12	
1	204	414	0	3	1	12	
1	204	414	0	4	1	12	
1	204	414	0	5	1	12	
1	204	414	0	6	1	12	
1	204	414	0	7	1	12	
2	204	8	1	1	0	15	
2	204	26	1	2	0	15	
2	204	152	1	3	0	15	
2	204	241	1	4	0	15	
2	204	249	1	5	0	15	
2	204	322	1	6	0	15	
2	204	350	1	7	0	15	

The `cgd2` data set has 896 observations.

Treatment only

```
cfit1 <- coxph(Surv(tstop, status) ~ rx + cluster(id),
   data=cgd1, subset=(enum==1))
cfita <- coxph(Surv(tstart, tstop, status) ~ rx + cluster(id),
           data =cgd1)
cfitc <- coxph(Surv(tstart, tstop, status) ~ rx + cluster(id)
          + strata(enum), data=cgd1)
cfitc2<- coxph(Surv(tstop-tstart, status) ~ rx + cluster(id)
          + strata(enum), data=cgd1)
cfitw <- coxph(Surv(time, status) ~ rx + cluster(id)
           + strata(enum), data=cgd2)
```

	β	se(β)	robust se
Time to first event	-1.09	0.34	0.34
A-G	-1.10	0.26	0.31
conditional	-0.86	0.28	0.29
cond (gap time)	-0.88	0.28	0.28
WLW	-1.34	0.27	0.36

Table 2: Fits for the CGD data

- The results of an Anderson-Gill, WLW, and conditional analysis are surprisingly similar to the pattern of the hidden covariate data set.

- The ordinary standard errors are too small.

- There are 44 first infections and 76 total.
  - ⋆ Potential gain of $\sqrt{44/76} = .76$ or 24%, exactly the gain for the 'naive' variance: $.76 * .34 = .26$.
  - ⋆ Actual gain of $.31/.34 = .91$ ($\approx \sqrt{44/53}$) or 9%
  - ⋆ each extra event was worth about 1/3 of an original event: "9" extra patients rather than 32.

We might expect that the independent increment and conditional models would give closer results if the model were to include significant covariates.

The most important factors, other than treatment, are age, steroids and enrollment center. Inheritance was a stratification factor.

```
> fita <- coxph(Surv(tstart, tstop, status) ~ rx + cluster(id) +
              age + inherit + steroids, data=cgd1)

             coef exp(coef) se(coef) robust se     z       p
      rx -1.0998     0.333    0.262    0.3093 -3.56 0.00038
     age -0.0397     0.961    0.014    0.0149 -2.67 0.00760
 inherit  0.3825     1.466    0.246    0.3208  1.19 0.23000
steroids -1.0615     0.346    0.527    0.5906 -1.80 0.07200

...
```

Below are the coefficients for the 4 models.

	rx	age	inherit	steroids
Time to first event	-1.16	-0.03	0.26	-0.91
A-G	-1.10	-0.04	0.38	-1.06
conditional	-0.90	-0.03	0.22	-0.78
WLW	-1.41	-0.05	0.48	-1.20

And here are those for models that stratify on center (coxph can have two `strata` terms), and drop inheritance.

	rx	age	steroids
Time to first event	-1.22	-0.02	-0.99
A-G	-1.26	-0.02	-1.05
conditional	-1.20	-0.03	-1.07
WLW	-1.59	-0.02	-1.25

The similarity to the hidden covariate data set is even stronger!

- Center is a very powerful effect.
- The conditional and A-G models grow more similar; with the larger change in the conditional fit.
- The WLW model's coefficient becomes more extreme.

Similar changes occur if center is added as a factor (class) variable.

Frailty models

These use a Gamma random effects term.

```
> coxme(Surv(tstart, tstop, status) ~ rx, cgd1,
       random= ~1|id)

Fixed effects: Surv(tstart, tstop, status) ~ rx
        coef exp(coef)  se(coef)    z       p
rx -1.008511 0.3647615 0.3058774 -3.3 0.00098

Random effects:  ~ 1 | id
               id
Variance: 0.63929

> coxme(Surv(tstart, tstop, status) ~ rx + strata(enum),
        cgd1,  random= ~1|id)

        coef exp(coef)  se(coef)     z       p
rx -1.128267 0.3235934 0.3174406 -3.55 0.00038

Random effects:  ~ 1 | id
                 id
Variance: 0.5804364
```

The random effects term seems to have “reconstructed” the hidden covariates

sufficiently well to synchronize the A-G and conditional solutions.

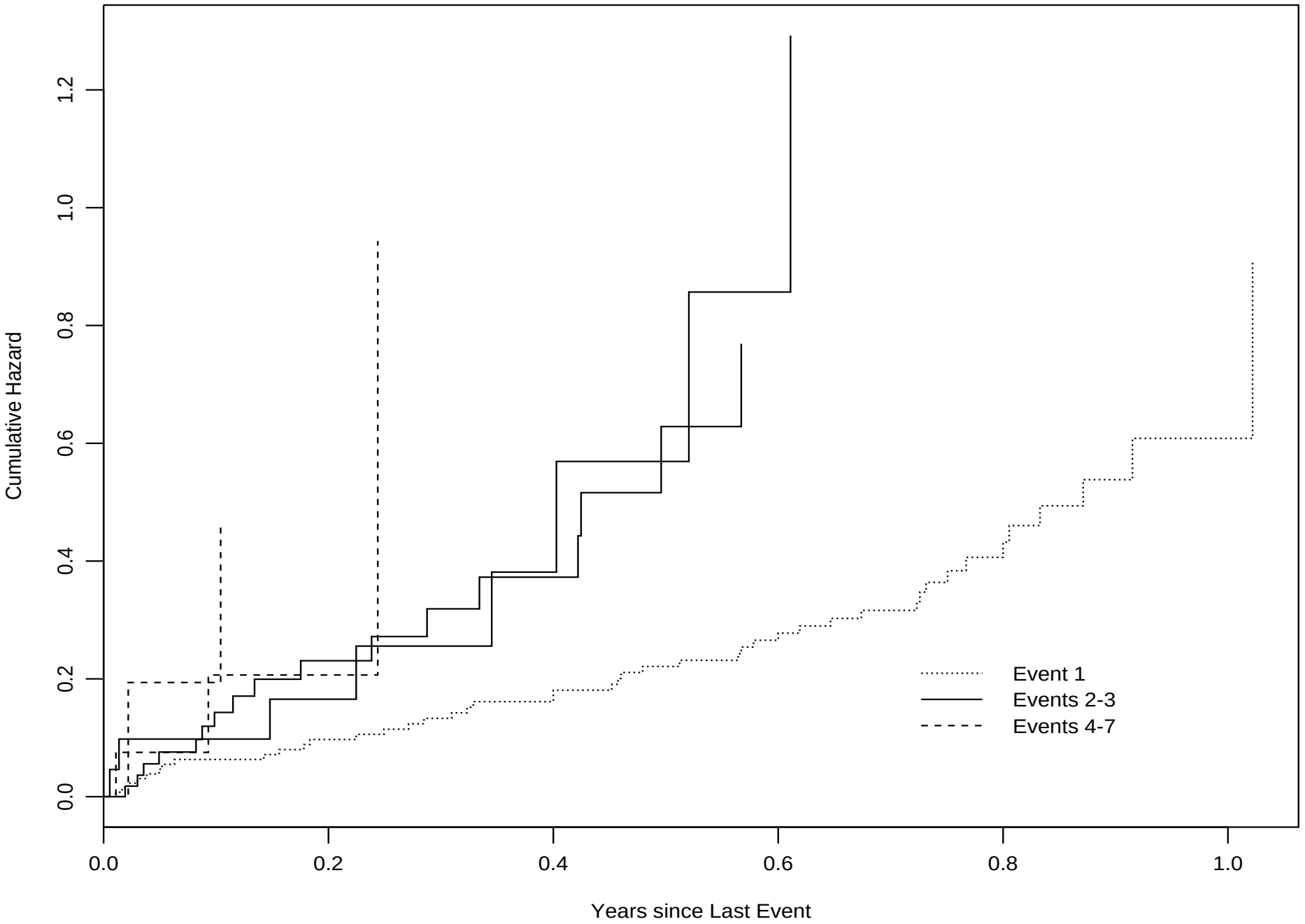
The predicted hazard curves based on the gap time model are interesting.

```
> curve <- survfit(cfitc2)
> plot(curve, lty=c(2,1,1,3,3,3,3), fun='cumhaz', xscale=365)
> legend(30, .2, lty=c(2,1,3), c("Event 1", "Events 2-3", "Events 4-7"),
       bty='n')
```

The hazards for all but the first strata seem to be quite similar. Subjects seem to have a longer 'wait' to first event.

We should not be surprised: enrollment criteria often select for patients in a state of relative health, and there is the issue of length biased sampling.

Ordered multiple events



Perhaps use one of the ‘hybrid’ models:

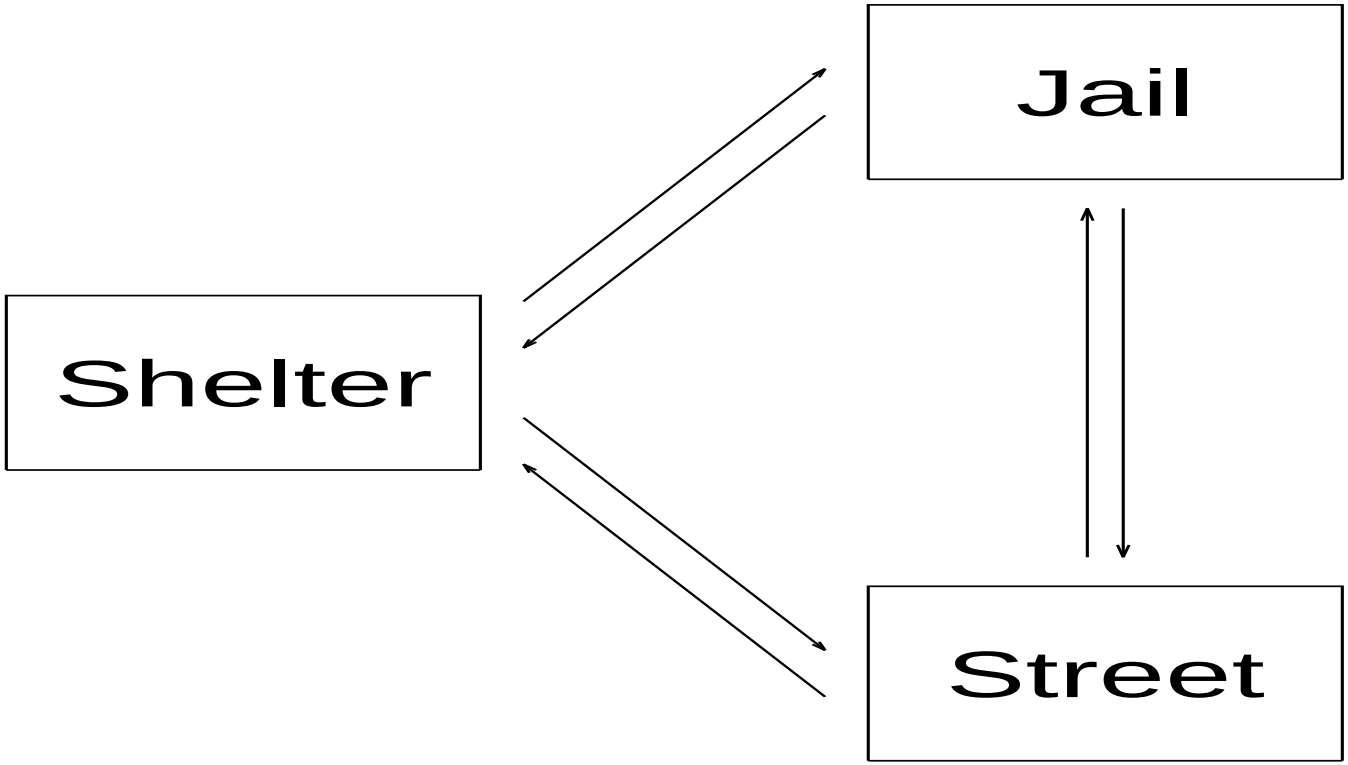
```
data temp; set cgd1;
    gtime = tstop - tstart;
    if (enum >1) then enum2=2;
    else              enum2=1;


proc phreg data=temp covsandwich(aggregate);
    model gtime * status(0) = rx age steroids/ ties=efron;
    strata enum2;
    id id;
```

Suggestion: we can use the first, and test for proportional hazards.

“Absence of proof is not proof of absence.”

Multiple State Models

Modeling the Homeless



Modeling the Homeless

There will be 6 strata, one for each possible transition

Assume a subject “Jones” enters our study on day 29 at a shelter, moves out of the shelter on day 40, reenters the shelter on day 58, is jailed on day 60 and lost to our followup on day 80. The data set would be

Start	Stop	Status	Strata	Covariates
29	40	1	Shelter $\Rightarrow$ Street	. . .
29	40	0	Shelter $\Rightarrow$ Jail	
40	58	0	Street $\Rightarrow$ Jail	
40	58	1	Street $\Rightarrow$ Shelter	
58	60	0	Shelter $\Rightarrow$ Street	
58	60	1	Shelter $\Rightarrow$ Jail	
60	80	0	Jail $\Rightarrow$ Shelter	
60	80	0	Jail $\Rightarrow$ Street	

Multi-state disease models are another example of this type of structure.

One important part of the setup is deciding which variables should have strata * covariate interactions, and which should not.

- The effect of age might well be different for shelter to street and shelter to jail transitions.
- It might be similar, however, for release from jail.
- One may be forced to assume fewer interactions, to limit the total number of coefficients in the model.

Crohn's Disease (IBD)

This is a recurrent inflammatory disease of the gut, of uncertain origin but possibly immune related.

- frequently leads to intestinal obstruction and abscess formation
- a high rate of recurrence
- flare-ups are treated with strong anti-inflammatory regimens (steroids, immunosuppression) but often lead to surgical removal of the inflamed section of the bowel followed by a quiescent period
- the waxing and waning nature of Crohn's disease can make it difficult to describe long-term outcomes

The study consists of the cohort of 174 Olmsted County, Minnesota, residents with a diagnosis of Crohn's disease from January 1, 1970, through December 31, 1993.

- From 2 months to 26 years of follow-up (median 61 months)
- 0–32 state changes/subject observed

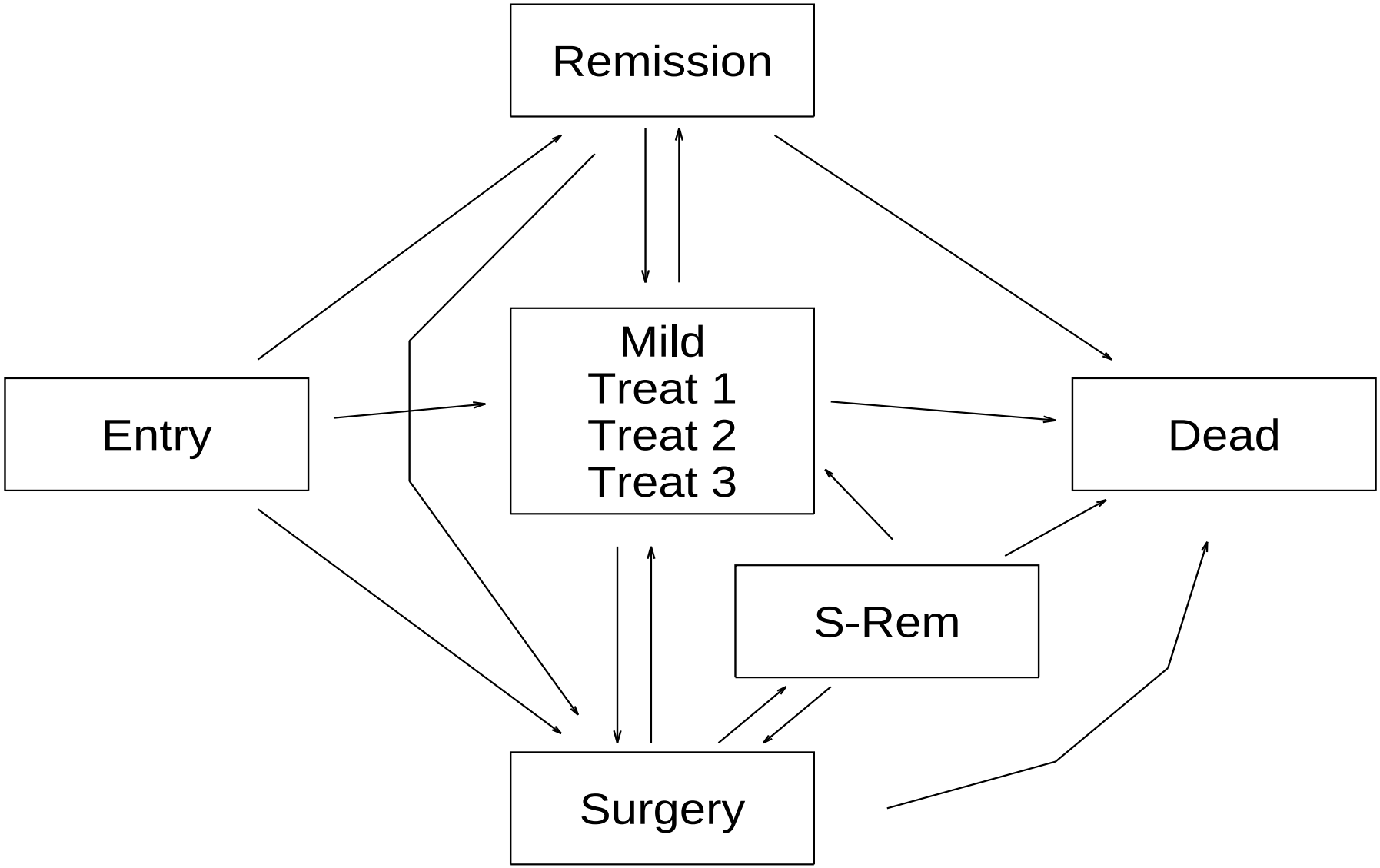
Over their time course, subjects were classified into one of 8 possible states

- Remission: Quiescent disease, no medication.
- Mild disease: Moderate medications such as antibiotic, sulfasalazine, or topical corticosteroids.
- Severe disease: Patients are being treated with oral corticosteroids or immunosuppressive therapy. There are 3 subcategories of drug-responsive, drug-dependent and drug-refractory; in what follows, these three states are labeled as ‘Treat1’, ‘Treat2’, and ‘Treat3’.

- Surgery: Inpatient surgical procedures.
- Surgical remission: A quiescent phase that often follows surgical intervention, no medication or treatment.
- Death from any cause.

State	No. patients ever in state	Entries to State					
		1	2	3	4–5	6–9	10+
Remission	146	40	49	28	16	10	3
Mild	138	50	32	20	23	11	2
Severe							
responsive	64	44	12	2	4	2	0
dependent	42	29	5	6	2	0	0
refractory	45	29	13	1	2	0	0
Surgery	100	64	20	7	7	1	1
Postsx remission	85	55	22	6	1	1	0

Inflammatory Bowel Disease



The starting data set has one observation for each transition, containing the prior state, current state, next state, time to transition, and covariates.

Consider a subject who enters on 1Mar78 into the mild state, treatment escalates on 28Mar78, remission is achieved on 16May78, and the patient is then followed forward to 28Sept78. The data would look like

id	sex	dx	age	pstate	current	nstate	dur
167	0	1Mar78	27	.	1	2	27
167	0	1Mar78	27	1	2	0	49
167	0	1Mar78	27	2	0	.	136

If the final observation does not end in a transition, e.g., a subject enters the ‘mild’ state midway through observation and is still in it at the last follow-up, then the nstate variable is equal to missing on the last line.

Internally, the state codes for the data set are 0=Remission, 1–4=Treatment, 5=Surgery, 6=Surgical remission and 7=Dead.

There are 1314 observations in the starting data set.

Creating a data set

- From each state, there are 6 possible transitions.
- The starting data set has 1314 transitions (not counting “Entry $\rightarrow$”).
- The analysis data set will have 7884 observations in 42 different strata.

```
data ibd2; set ibd1;
    if (0<= current <= 5) then do;
do i = 0 to 7;
    if (nstate =i) then status =1;
                   else status =0;
    strat = current*10 + i;
    if (i ne current and i ne 6) then output;
    end;
end;
```

```
    else do;
do i = 1 to 7;
    if (nstate =i) then status =1;
                   else status =0;
    strat = current*10 + i;
    if (i ne current) then output;
    end;
end;
```

Initial analysis

Question: is the process Markov?

That is, do the outcome and risk for a subject depend only on the current state?

Solution: treat prior state as a CLASS variable, and test it.

```
data temp1; set save.crohn2;
    entry   = (pstate= .);
    ps_rem  = (pstate= 0);
    ps_mild = (pstate= 1);
    ps_trt1 = (pstate= 2);
    ps_trt2 = (pstate= 3);
    ps_trt3 = (pstate= 4);
    ps_surg = (pstate= 5);
    ps_srem = (pstate= 6);

    if (stratum=17 or stratum=57) then delete; *SAS bug;

proc phreg data=temp1 covsandwich(aggregate);
    model time * status(0) = male age prev_dur ps_rem
        ps_mild ps_trt1 ps_trt2 ps_trt3 ps_surg ps_srem
           / ties=efron;
    strata stratum;
```

```
test ps_rem, ps_mild, ps_trt1, ps_trt2, ps_trt3,
      ps_surg, ps_srem;
id id;
```

Variable	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq
MALE	1	-0.21169	0.06479	1.052	10.6746	0.0011
AGE	1	-0.0004887	0.00248	1.195	0.0388	0.8438
PREV_DUR	1	-0.0000350	0.0000492	0.976	0.5056	0.4770
PS_REM	1	-0.46393	0.13266	1.222	12.2308	0.0005
PS_MILD	1	-0.65367	0.12862	1.226	25.8271	<.0001
PS_TRT1	1	-0.58529	0.14469	1.071	16.3628	<.0001
PS_TRT2	1	-0.55834	0.20338	1.184	7.5365	0.0060
PS_TRT3	1	-0.45987	0.16811	1.044	7.4827	0.0062
PS_SURG	1	-0.87276	0.21317	1.002	16.7618	<.0001
PS_SREM	1	-0.49334	0.18063	1.179	7.4596	0.0063

Label	Wald Chi-Square	DF	Pr > ChiSq
Test 1	33.1388	7	<.0001

(The actual output is spread over 3 pages).

- Clearly, the prior state has an influence on the rate with which we leave the current state.

- But notice that the coefficients, which are all contrasts with the prior state of “New patient”, are all about the same size.

- It would be useful to rerun, using Remission as the baseline state.

```
proc phreg data=temp1 covsandwich(aggregate) nosummary;
    model time * status(0) = male age prev_dur entry
        ps_mild ps_trt1 ps_trt2 ps_trt3 ps_surg ps_srem
            / ties=efron;
    strata stratum;
    id id;
    test ps_mild, ps_trt1, ps_trt2, ps_trt3, ps_surg, ps_srem;
-----------------------------------------
```

Variable	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq
MALE	1	-0.21169	0.06479	1.052	10.6746	0.0011
AGE	1	-0.0004887	0.00248	1.195	0.0388	0.8438
PREV_DUR	1	-0.0000350	0.0000492	0.976	0.5056	0.4770
ENTRY	1	0.46393	0.13266	1.222	12.2308	0.0005
PS_MILD	1	-0.18973	0.13613	1.274	1.9426	0.1634
PS_TRT1	1	-0.12136	0.13046	0.993	0.8653	0.3523
PS_TRT2	1	-0.09441	0.17430	1.024	0.2934	0.5881
PS_TRT3	1	0.00406	0.15864	1.024	0.0007	0.9796
PS_SURG	1	-0.40883	0.18656	0.921	4.8025	0.0284
PS_SREM	1	-0.02941	0.16913	1.245	0.0302	0.8619

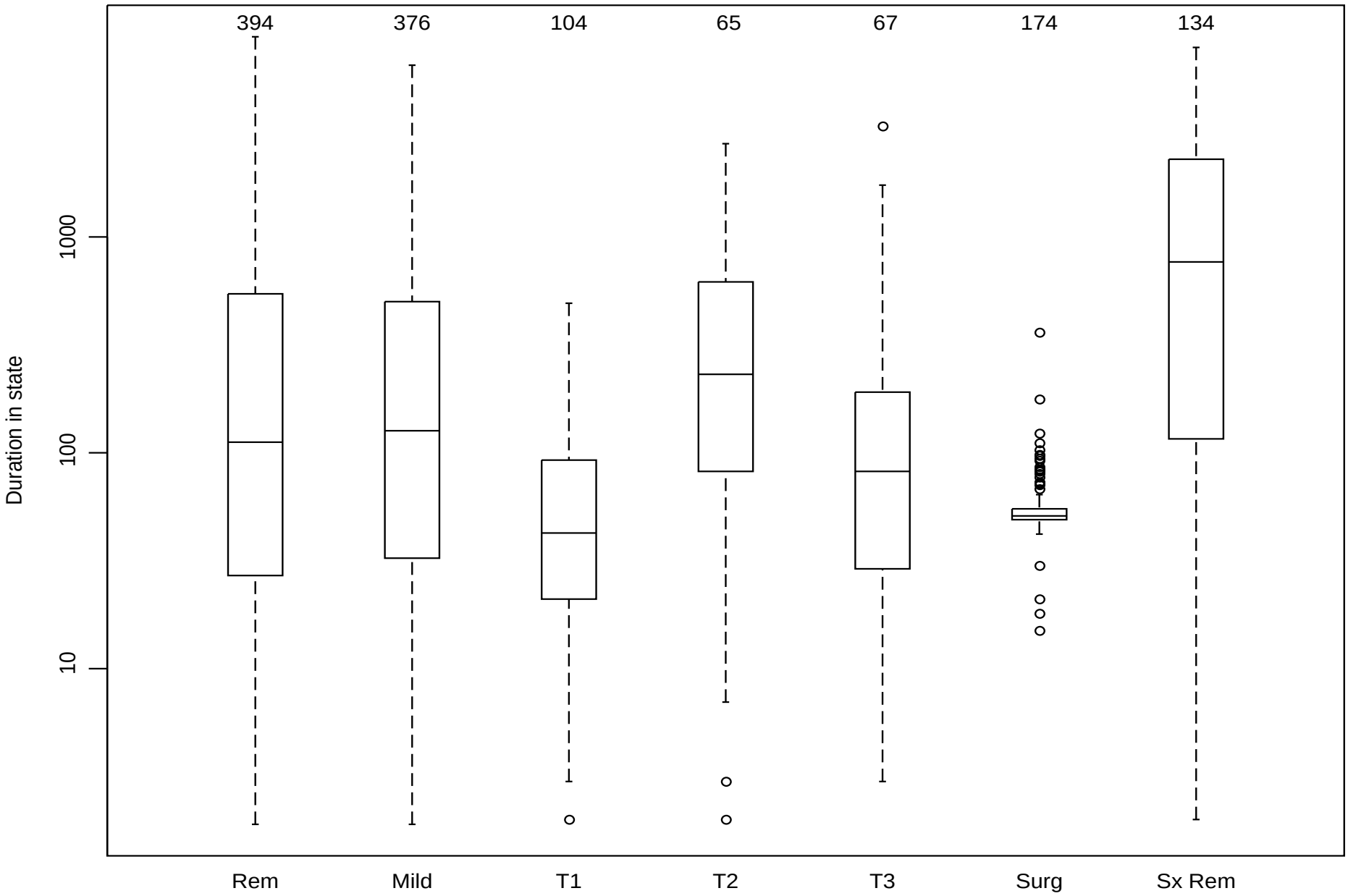
	Chi-Square	DF	Pr > ChiSq
Test 1	6.4872	6	0.3709

- If a subject is new, they have a higher rate, exp(.46)=1.6, of transition. The initially assigned state doesn't 'last' as long.

- Most subjects go from the surgery state to surgical remission (134), but a few directly to one of the 4 treatment states (36). The duration of these treatment states is longer than usual. (Surgery is the definitive treatment.)

- Notice that this is not a reflection of longer duration in the surgical remission state, since surgery is the only gateway into that state.

- The analysis so far is only for “time in state”, not distinguishing states
- Do variables affect some transition times more than others?
- We first note that, almost by definition, nothing can affect time in the surgical state.


394
376
104
65
67
174
134
Duration in state
1000
100
10
Rem
Mild
T1
T2
T3
Surg
Sx Rem

```
data temp2; set temp1;
   if (current=5);

proc phreg data=temp2;
    model time*status(0) = male age prev_dur;
    strata stratum;
```

```
----------------------------------------
```

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
MALE	1	0.05042	0.16010	0.0992	0.7528	1.052
AGE	1	-0.00539	0.00534	1.0188	0.3128	0.995
PREV_DUR	1	-0.0000597	0.0001171	0.2594	0.6105	1.000

A second fit (not shown) shows that the time is also not significantly related to which state preceded surgery.

The outlier in the surgery time period (361 days) is at first worrisome, but it actually does not have a major effect on the estimates, helped by the fact that a Cox fit is essentially based on the *ranks* of the exit times.

Take home message: look at your data before starting the analysis.

Based on the above, we fit a summary model.

First, create two new covariates

- `entry`=1 if this is a subject's first state after entry into the study
- `prior_sx`=1 for the first state after a surgery.

Time in the surgery state is left out of the model, since it is clearly not related to the covariates being tested (this is, in one sense, a particular state by covariate interaction).

```
> entry   <- 1*(is.na(crohn2$previous))
> priorsx <- 1*(crohn2$previous ==  5)

> fit4 <- coxph(Surv(time, status) ~ entry + priorsx + male +
                              strata(stratum) + cluster(id),
              data=crohn2, subset=(current != 5))
> fit4
          coef exp(coef) se(coef) robust se     z       p
  entry  0.588     1.801   0.0963    0.1231  4.78 1.8e-06
priorsx -0.324     0.723   0.1969    0.1762 -1.84 6.6e-02
   male -0.267     0.766   0.0671    0.0757 -3.53 4.2e-04

Likelihood ratio test=53.3  on 3 df, p=1.55e-11  n= 6840
```

The coefficient for being in a prior surgical state is (barely) not significant in the reduced model.

The first state after entry has a reduced duration (1.8 fold increased rate of departure from the state), but beyond that first transition a Markovian model might be entertained as neither the prior state nor the duration in that state has significant predictive power.

We have not looked at more complicated patterns, of course, such as the pattern formed by the prior two states, prior three states, etc., but this could be explored by creation of the proper interaction variables (given sufficient sample size).

The analysis so far has only looked for a covariate's overall effect on "time in state", not distinguishing the states.

Does gender, for instance, affect some transition times more than others?

```
data temp3; set temp1;
    if (current =5) then delete;

    malerem = 1*(male=1 and current=0);
    malemild= 1*(male=1 and current=1);
    maletrt1= 1*(male=1 and current=2);
    maletrt2= 1*(male=1 and current=3);
    maletrt3= 1*(male=1 and current=4);
    malesrem= 1*(male=1 and current=6);

proc phreg data=temp3 covsandwich(aggregate);
    model time * status(0) = age entry malerem malemild maletrt1
         maletrt2 maletrt3 malesrem / ties=efron;
    strata stratum;
    id id;
```

Variable	DF	Parameter Estimate	Standard Error	StdErr Ratio	Chi-Square	Pr > ChiSq
AGE	1	0.00169	0.00304	1.352	0.3069	0.5796
ENTRY	1	0.59257	0.12289	1.273	23.2512	<.0001
MALEREM	1	-0.18781	0.12161	1.075	2.3849	0.1225
MALEMILD	1	-0.25068	0.12994	1.147	3.7217	0.0537
MALETRT1	1	-0.74524	0.20973	0.905	12.6263	0.0004
MALETRT2	1	-0.41741	0.31800	1.055	1.7229	0.1893
MALETRT3	1	-0.10254	0.29137	1.121	0.1239	0.7249
MALESREM	1	-0.27288	0.24776	1.148	1.2131	0.2707

It appears that the gender effect is largest for transitions out of the *mild* and *early treatment* states.

This is not inconsistent with the social hypothesis.

Further questions

- Which prior state information to use
- Gender interactions?
- Duration in prior state interactions?
- Hazard curves

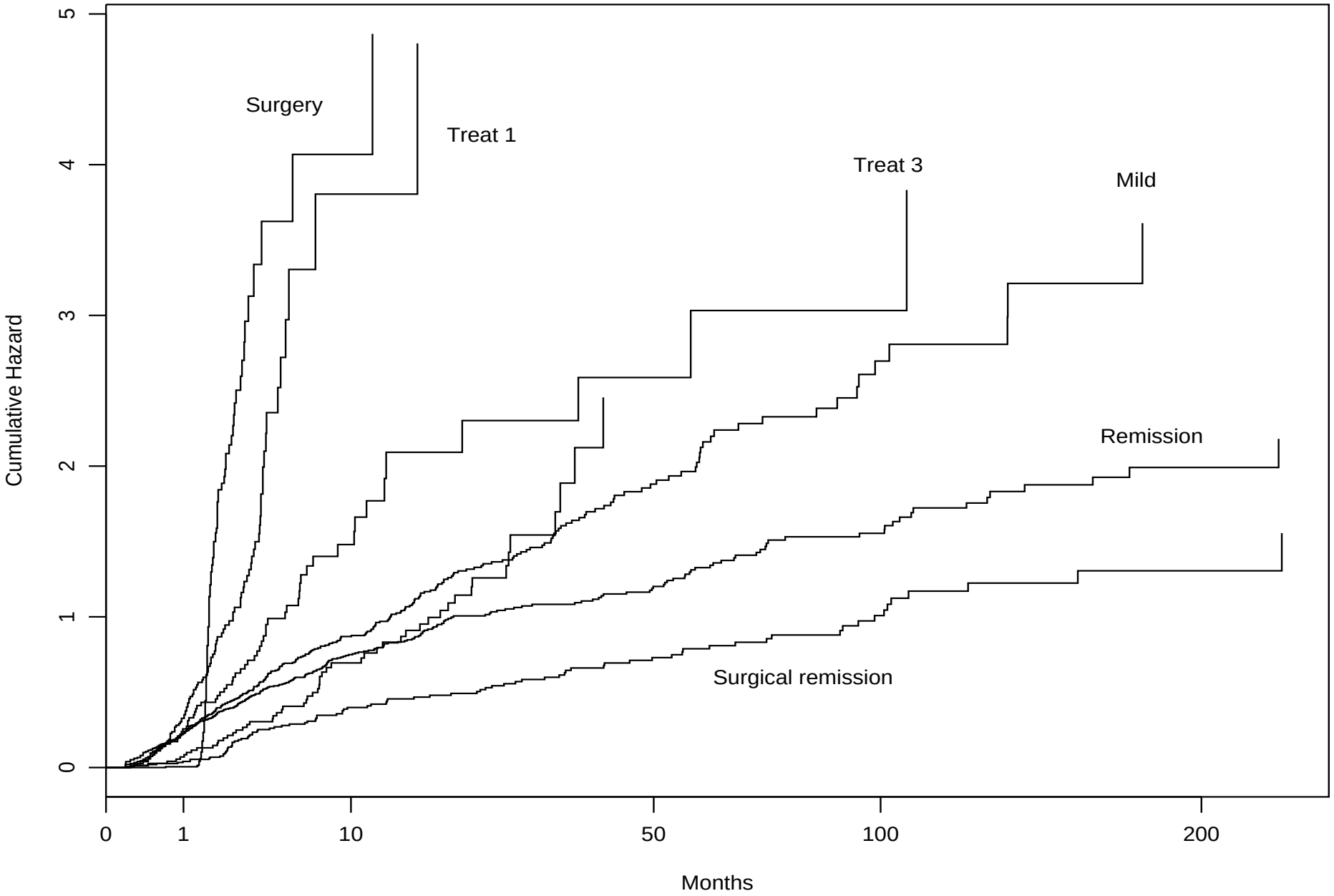
Hazard functions

```
> fit7 <- coxph(Surv(time, status) ~ male + ent.rem + ent.trt1 +
                              strata(stratum),
         data=crohn1)
> tdata <- data.frame(male=1, ent.rem=0, ent.trt1=0)
> fsurv <- survfit(fit7, newdata=tdata)

> fsurv
Call: survfit.coxph(object = fit7, newdata = tdata)

            n events mean se(mean) median 0.95LCL 0.95UCL
current=0 394    339 5.17   0.3554   2.85    2.43    3.77
current=1 376    340 3.65   0.2063   2.41    2.06    2.89
current=2 103    100 1.43   0.0806   1.36    1.16    1.59
current=3  64     55 3.50   0.2655   2.91    2.73    4.32
current=4  67     63 2.48   0.3726   1.83    1.53    2.31
current=5 173    170 1.40   0.0345   1.31    1.29    1.32
current=6 134     90 7.73   0.5468   6.44    5.14    9.51
```

```
> fsurv$time <- sqrt(fsurv$time/30.5)
> plot(fsurv, fun='cumhaz', lty=1, xaxt='n',
       xlab='Months', ylab='Cumulative Hazard')
> temp <- c(0, 1, 10, 50, 100, 200)
> axis(1, sqrt(temp), format(temp))
```

These fits are based on the data set *crohn1*, which has one observation per state.

They ask: how long do you stay in a state, ignoring which state is the destination.


Surgery
Treat 1
Treat 3
Mild
Remission
Surgical remission
Cumulative Hazard
0
1
2
3
4
5
0
1
10
50
100
200
Months

Hazard functions

```
data tdata;
   male=1; ent_rem=0; ent_trt1=0;
   output;

proc phreg data=crohn1;
   model time*status(0) = male ent_rem ent_trt1;
   strata stratum;
   baseline covariates=tdata out=osurv
  u=usurv l=lsurv/nomean;

proc plot;
   plot sqrt(time) * surv;   /* not legal SAS */
```

Frailty Models

Software Maturity

- Exists but you wouldn’t want it
- Worked application(s) in a paper
  software shared with friends
- General program/ S function
  - ⋆ available from author and/or internet
  - ⋆ advertised
- Available in a package, or R ’recommended’
- In SAS

Random Effects and Cox

- Simple effects
  - Stata
    - Built in
  - S/R
    - `coxph (Surv(time, status)~ frailty(grp))`
  - SAS
    - Macro from J. Klein
    - Parametric weibull + frailty, NLMIXED examples
- Complex
  - S/R: coxme function

Data

There are exciting challenges with real data sets –

- Multiple events per subject
  - ⋆ Recurrent infections, competing risks, multi-state models
  - ⋆ First line analysis: marginal models
  - ⋆ What can random effects models add?
  - ⋆ Can we “squeeze out” more information?
- Familial and genetic data
  - ⋆ Known correlation structure
  - ⋆ Random effects are interesting of themselves
  - ⋆ Varying amounts of data per family
  - ⋆ Increasing interest in time to event models

- Large covariate sets
  - Is “per patient” prediction possible?
  - How much can be gained from redundant covariates?
  - Can we gain flexibility and still control the degrees of freedom?

“Medical research at Mayo is like standing in front of a fire hydrant with your mouth open.” – John Blinks

In the last several years there has been significant and active research concerning the addition of random effects to survival models.

In this setting, a random effect is a continuous variable which describes excess risk or *frailty* for distinct categories, such as individuals or families.

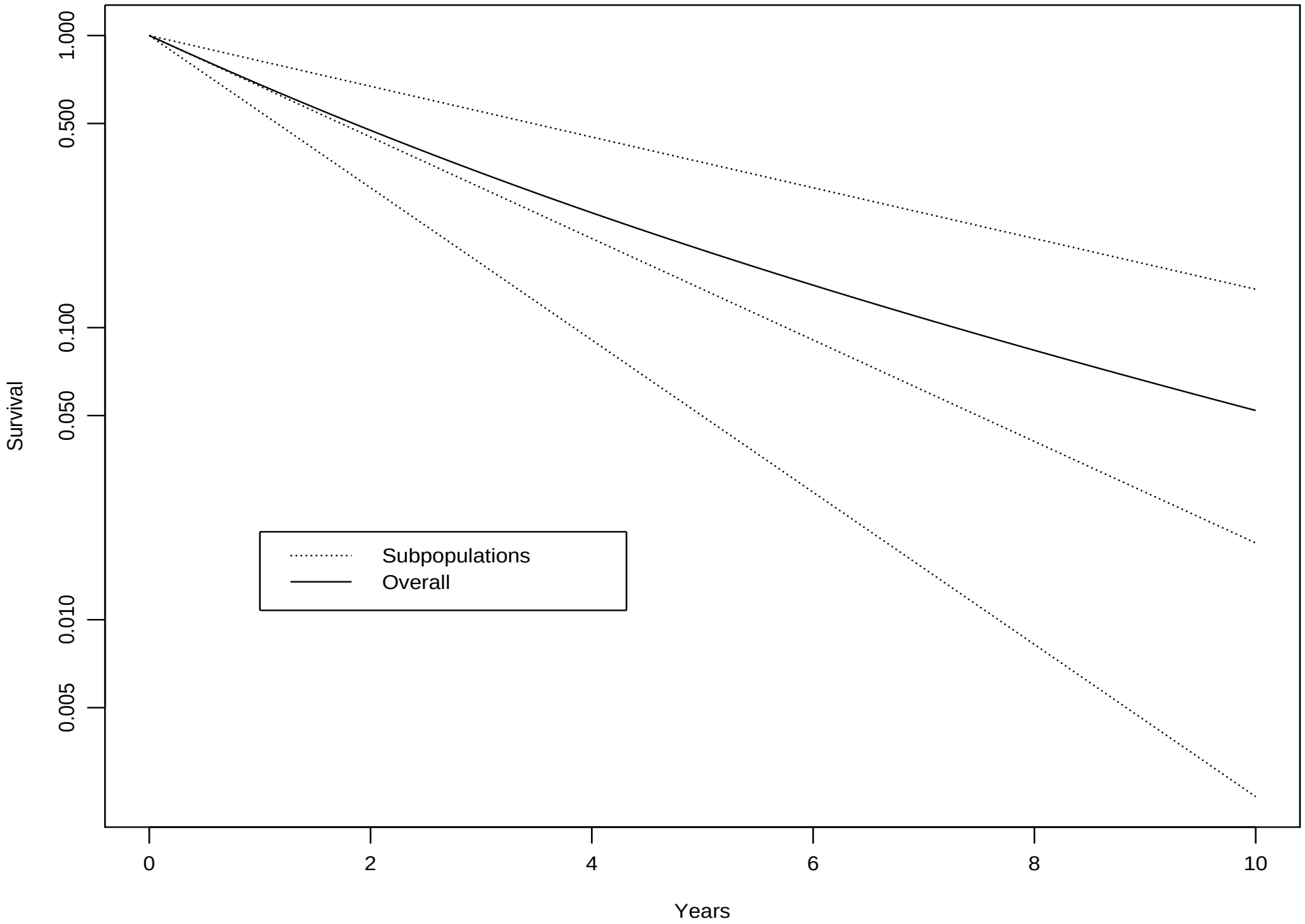
The idea is that individuals have different frailties and that those who are most frail will die earlier than the others.

Aalen (Stat Med 88) provides theoretical and practical motivation for frailty models by discussing the impact of heterogeneity on analyses and by illustrating how random effects can deal with it.

He states

> "It is a basic observation of medical statistics that individuals are dissimilar. . . . Still, there is a tendency to regard this variation as a nuisance, and not as something to be considered seriously in its own right. Statisticians are often accused of being more interested in averages, and there is some truth to this."

One issue is the selection problem: Assume 3 groups of subjects with hazards of 20, 40, and 60 events/100 per year. What does the overall survival curve look like?


Survival
1.000
0.500
0.100
0.050
0.010
0.005
0
2
4
6
8
10
Years
Subpopulations
Overall

We are all familiar with this.

Related to several other “individual patient” vs “public health effect” problems.

But, in survival analysis:

- This has an effect on our estimates and can bias them toward 0. Omori and Johnson, Biometrika 1993: the rate of bias acquisition = var$(X|t)$, where X is the frailty with an initial mean of 1.
- The effect can be profound on multiple event studies.
- Aalen shows that with a sufficiently variable frailty the population relative risk can go from r to $1/r$ over time, even though the true relative risk stays at r.

As an example of the possibility of reversing risks, assume that we have a treatment for multiple recurrent infections. Unknown to us, the population is stratified as follows:

Subset	Proportion	rate	Drug effect
A	.4	1/year	.5
B	.4	2/year	.5
C	.2	10/year	.2

The drug is very effective for the majority of patients, reducing the rate of infections by half.

There is a subset of relatively intractable patients with severe disease, however, for which it is less effective but still quite worthwhile, reducing the infection rate by 20%.

A study is planned with 1 year of follow-up per patient, 1000 subjects per arm. Assuming that infections are a simple Poisson process with the above rates, we can compute the expected number of subjects who will finish the year with 0, 1, . . . events directly:

	Number of infections					
	0	1	2	3	4	5+
Placebo	201	256	182	98	46	217
Treatment	390	269	106	35	18	182

For a patient with 2 infections already, what is the chance of a third? In the placebo arm it is $361/543 = 66\%$ and for the treatment arm the chance of another infection within the year is $235/341 = 69\%$, higher than the placebo group!

The problem is that the treatment arm's '2 events or more' group is dominated by the severe disease patients, 58%, while the placebo group is 37% severe disease patients at that point.

This wouldn't happen for an analysis stratified by subset, but we don't know the subsets.

Aalen shows that with a sufficiently important frailty, the population relative risk can go from r to $1/r$ over time, even though the true relative risk stays at r.

Linear mixed effects model

$$y = X\beta + Zb + \epsilon$$

- y = response
- β = fixed effects
- b = random effects $\sim N(0, \tau^2)$
- ϵ = random error $\sim N(0, \sigma^2)$

What if you fit $y = X\beta$

- β is still unbiased.
- The estimated error is larger.

Preview

Consider a random effects Cox model

$$\lambda_i(t) = \lambda_0(t) e^{X_i\beta + Z_i b}$$

- $\beta_1, \ldots, \beta_p$ are the fixed effects
- $b_1, \ldots, b_q$ are the random effects
- usual linear models notation

Four computing approaches

1. Bayes: write down a prior for b, hyperprior, hyper-hyper prior, . . . , use MCMC, and explore the solution.

2. EM: for a small subset of cases explicit solutions can be found.

3. Penalized methods:

   - Agree with EM for several cases
   - Extensible to a larger variety of cases
   - Fast, flexible

4. Laplace approximation

   - Particular to the case of a Gaussian random effect
   - Closely related to the penalized methods
   - The basis for *coxme*

Choosing the distribution

$$\lambda_i(t) = \lambda_0(t) e^{X_i\beta + Z_i b}$$

with $b \sim G(0, \sigma^2)$.

What is the "proper" distribution for the random effect b?

- Easy manipulation of the likelihood
  - ⋆ Linear model: Gaussian
  - ⋆ Cox model: Gamma (special cases)
- Closure: models with and without $Z_i b$ are from the same family.
  - ⋆ Linear model: Gaussian
  - ⋆ Cox model: Positive stable

- Correlated random effects
  - ⋆ Linear model: Gaussian
  - ⋆ Cox model: Gaussian (bivariate: gamma)

- Biological: a distribution that is plausible for the “hidden” covariates.
  - ⋆ continuous vs multinomial b for genetic traits
  - ⋆ skewed vs symmetric lab values
  - ⋆ Linear model: very little discussion
  - ⋆ Cox model: almost no discussion
  - ⋆ It is very hard to discern the distribution of a variable you cannot see

Simple Frailty

$$\lambda(t) = \lambda_0(t)e^{X\beta+Zb}$$

- Assume that there are j = 1, 2, . . . k groups (family, enrolling institution, . . .)
- $b_1, b_2, \ldots, b_k$ are the random effects
- Z has the structure of a 1-way anova
  - ⋆ $Z_{ij} = 1$ if and only if subject i is a member of group j

- In linear models, this is often written as
  - $y_{ij} = X_i\beta + b_i$
  - i = groups
  - j = subject within groups
  - X_i = data matrix for group i

Equivalent

$$\begin{aligned}\lambda_{i(j)}(t) &= \lambda_0(t)e^{X_i\beta}e^{b_j} \\ &= \lambda_0(t)e^{X_i\beta}\varpi_j\end{aligned}$$

- $\lambda_{i(j)}$ = hazard for subject i (who is a member of group j)

- $\varpi \sim G(1, \sigma^2)$
  - G is a positive distribution
    - gamma
    - positive stable
  - wlog we can assume that G has mean 1

Expectation-Maximization Algorithm (E-M)

$$LPL = \sum_{i=1}^{n} \int_0^{\infty} \frac{e^{X_i\beta + Z_i b}}{\log\left(\sum_{j=1}^{n} Y_j(t) e^{X_j\beta + Z_j b}\right)} dN_i(t)$$

This is a function of β and b where the latter is random: $\ln(b) \sim G(1, \sigma^2)$

Iterate

1. E-step $Q(\beta, \sigma) =$ E (LPL $(\hat{\beta}, \hat{\sigma})$)
2. M-step $(\hat{\beta}, \hat{\sigma}) = \max Q(\beta, \sigma)$

Guaranteed to converge (eventually)

The EM can be excrutiatingly slow

- Newton-Raphson iteration, once close, doubles the number of correct digits at each step
- Even binomial search gets one more (binary) digit per step
- The EM *is* guaranteed to get better at each step.

Formal algebra

Let $\phi(s)$ be the Laplace transform of ϖ, and $\phi^{(n)}(s)$ be its nth derivative.

Parner (96) shows that the EM algorithm for solving the shared frailty model has the "E" step

$$\varpi_j = -\frac{\phi^{(d_j+1)}(A_j)}{\phi^{(d_j)}(A_j)},$$

where

- d_j = number of events in family j
- A_j = expected number of events, based on $\hat{\beta}$

The "M" step is an ordinary Cox model, treating the b_j as known.

He also shows that the observed-data log-likelihood, i.e., the result after integrating out the frailty, is

$$L_g = \sum_{i=1}^{n} \delta_i \log(\lambda_i e^{X_i\beta}) + \sum_{j=1}^{q} \log[(-1)^{d_j}\phi^{(d_j)}(A_j)]$$

The derivation depends critically on separation of the PL into separate ‘family’ terms. (Only 1 non-zero element of Z per row).

If b has a gamma distribution with mean 1, then

$$\phi(s) = (1 + 1/\nu)^{-\nu}$$

where $1/\nu$ = variance.

The derivatives of $\phi(s)$ are

$$\phi^{(d)}(s) = \left(-\frac{1}{\nu}\right)^d \left(1 + \frac{s}{\nu}\right)^{-(\nu+d)} \prod_{i=0}^{d-1} (\nu + i)\,,$$

and the E-step reduces to

$$e^{b_j} = \frac{d_j + \nu}{A_j + \nu}\,.$$

This has a simple interpretation as a shrinkage estimate.

$$\varpi_j = \frac{d_j + 1/\theta}{A_j + 1/\theta}$$

If the variance of the random effect (θ) is infinite (no restriction on the size of the random effect), then the family effect is the simple Poisson estimate (events/ total time).

As the variance goes to 0, the familial effect is forced to 1.

Other suggested distributions are the inverse Gaussian

$$\phi(s) = e^{2\theta} e^{-2\sqrt{\theta^2+\theta s}}$$

or the positive stable

$$\begin{aligned} \phi(s) &= e^{-s^p} \\ \phi'(s) &= -ps^{p-1}e^{-s^p} \\ \phi''(s) &= pe^{-s^p}[ps^{p-1} - (p-1)s^{p-2}] \end{aligned}$$

The Parner algorithm does not simplify in these cases.

The EM approach is a dead-end, for practical computation.

Positive Stable Frailty

Hougaard recommends using the positive stable distribution for $\ln(b)$.

Laplace transform

$$L(t) = e^{-t^{\alpha}} \quad 0 < \alpha \leq 1$$

Parner shows

$$
\begin{aligned}
L^{(n)}(t) &= \sum_{i=0}^{n} -1 \left(\frac{n-i}{i} \right) L^{(i)}(t) h(n-i) t^{\alpha+i-n} \\
h(x) &= -\alpha(\alpha-1)(\alpha-2)\ldots(\alpha-x+1)
\end{aligned}
$$

so a simple computer program is possible.

Other patterns

For sibling pairs, Li worked out a frailty using

$$b = \gamma_1 + \gamma_2$$

where γ_1 and γ_2 are Gamma random effects, one of which is shared between the siblings and one is not.

Others have extended this idea to more complex situations.

xxx has worked out a nested model with gamma frailty.

Connection to penalized models

The frailty model can be re-written as

$$\lambda_0(t)\,\varpi_j e^{X_i\beta} = \lambda_0(t) e^{X_i\beta + Z_i b}$$

with

- Z a matrix of appropriate indicator variables
- $b_j \equiv \log(\varpi_j)$
- b is from a log-Gamma distribution

For the gamma frailty model and *fixed* value for σ^2

- frailty solution = penalized solution with $p(b, \sigma) = \sum(b_j - e^{b_j})/\sigma^2$
- $L_g = \mathrm{PPL} + \sum_j c(d_j, \sigma)$, where d_j is the total number of events in group j. This does not depend on β or b.
- The outer loop or "control" function is just
  - ⋆ Add the correction term (giving a marginal likelihood)
  - ⋆ Try various values of θ, to get a plot of the *profile likelihood*
  - ⋆ Find the maximum of the plot

Thus, the gamma frailty model is easily fit using `coxph`.

$$c(d, \sigma) = \sigma^2 + \log \Gamma(\sigma^2 + d) - (\sigma^2 + d) \log(\sigma^2 + d)$$

Rat data

- Female rats from Mantel et. al.
- 3 rats/litter, one treated + two placebo
- 50 litters

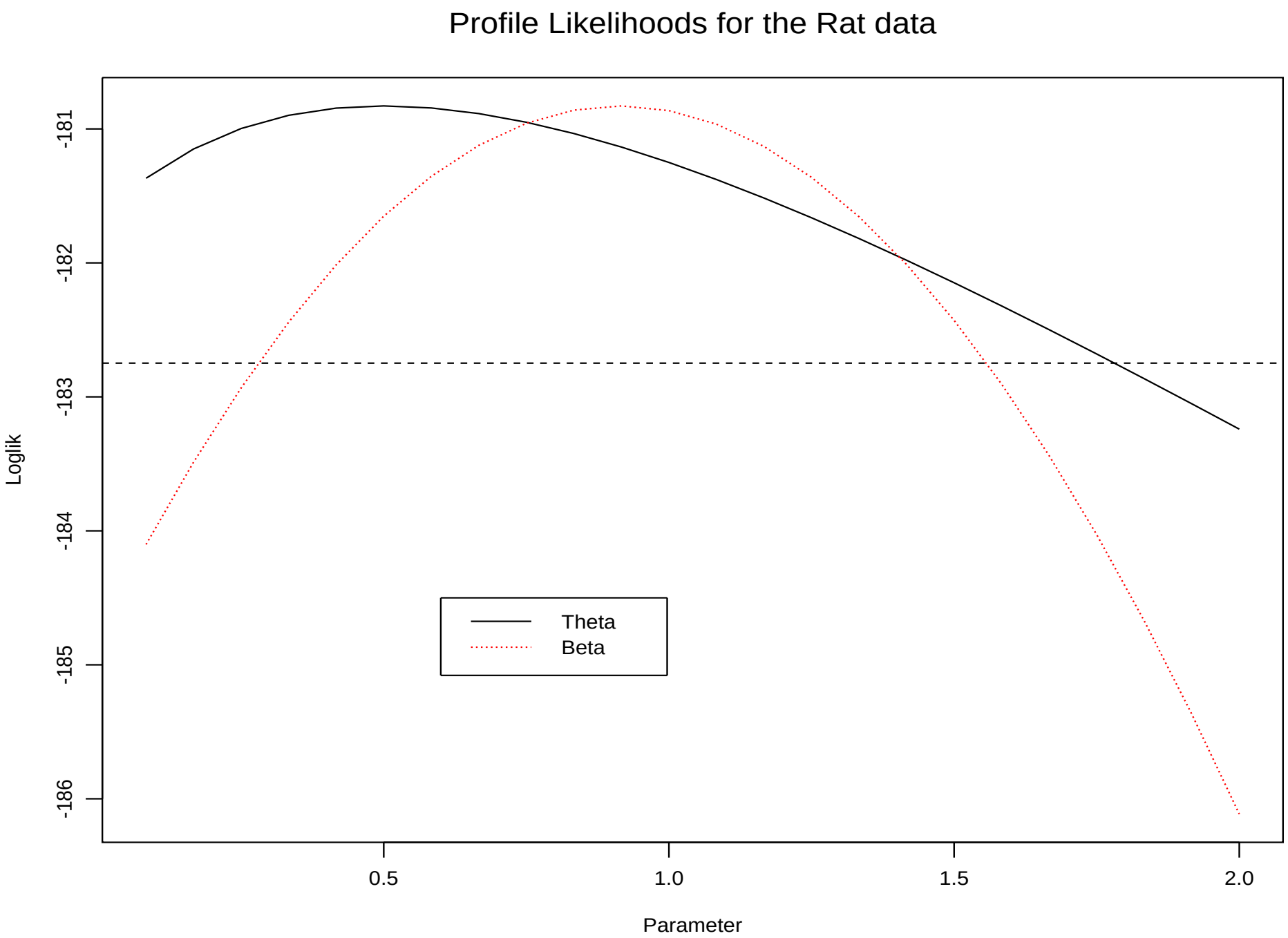
WLW marginal analysis

```
> coxph(Surv(time, status) ~ rx + cluster(litter), rats)

     coef   exp(coef) se(coef)  robust se    z       p
rx  0.898      2.46    0.317        0.3     2.99 0.0028

Likelihood ratio test=7.87  on 1 df, p=0.00503  n= 150
```

Gamma frailty

```
> coxph(Surv(time, status) ~ rx + frailty(litter), rats)
                 coef se(coef)   se2 Chisq   DF      p
```

```
             rx 0.914 0.323    0.319  8.01  1.0 0.0046
frailty(litter)                       17.69 14.4 0.2400

Iterations: 6 outer, 19 Newton-Raphson
     Variance of random effect= 0.499   EM likelihood = -180.8
Degrees of freedom for terms=  1.0 14.4
Likelihood ratio test=37.6  on 15.38 df, p=0.00124  n= 150
```


Profile Likelihoods for the Rat data
Loglik
-181
-182
-183
-184
-185
-186
0.5
1.0
1.5
2.0
Parameter
Theta
Beta

What is actually being fit?

- An indicator variable for each litter is added to the model (like a `class` statement) = 51 coefficients.
- The 50 litter coefficients are shrunk toward 0, using the penalty.
- The computation uses a special sparse computation trick to avoid having a 51x51 second derivative matrix.

- The 50 “litter” coefficients and their variance is suppressed from the printout. (`print(fit$frail)`).

Gaussian frailty

The hazard function for the sample is defined as

$$\lambda(t) = \lambda_0(t)e^{X\beta+Zb} \text{ or} \quad (1)$$

$$\lambda_i(t) = \lambda_0(t)e^{X_i\beta+Z_ib} \quad (2)$$

Consider the log partial likelihood $PL(\beta, b)$, treating β and b as covariates.

Assume $b \sim N(0, \Sigma)$. Integrating the random effect out of the partial likelihood gives an integrated log-partial likelihood $\mathcal{L}$ of

$$\begin{aligned} e^{\mathcal{L}} &= \frac{1}{\sqrt{2\pi|\Sigma|}} \int e^{\mathrm{PL}(\beta,b)} e^{-b'\Sigma^{-1}b/2} db \\ &= \frac{1}{\sqrt{2\pi|\Sigma|}} \int e^{\mathrm{PPL}(\beta,b)} db \\ &\approx \frac{1}{\sqrt{2\pi|\Sigma|}} e^{\mathrm{PPL}(\beta,\hat{b})} \int e^{-(b-\hat{b})'H_{\hat{b}\hat{b}}(b-\hat{b})/2} db \end{aligned}$$

$$\mathcal{L} \quad = \quad \mathrm{PPL}(\beta, \hat{b}) - - \left[\frac{1}{2} \log |\Sigma| + \log |H_{\hat{b}\hat{b}}| \right]$$

Laplace approximation

$$L = \int e^{PL(\beta,b)} e^{-b'Ab/2} db$$

Let

$$g(b) = PL(\beta, b) - b'Ab/2$$

and $\hat{b}$ = the value that minimizes $g(b)$

Then

$$\begin{aligned} g(b) &\approx g(\hat{b}) + (b-\hat{b})\frac{\partial g(b)}{\partial b}_{\hat{b}} + (b-\hat{b})'\frac{\partial^2 g(b)}{\partial b^2}_{\hat{b}}(b-\hat{b})/2 \\ &= g(\hat{b}) + 0 + (b-\hat{b})' H_{bb}(b-\hat{b})/2 \end{aligned}$$

But

$$\frac{1}{2\pi|H|}\int e^{(b-\hat{b})'H(b-\hat{b})/2} db = 1$$

So

$$L = \int e^{PL(\beta,b)} e^{-b'Ab/2} db \approx e^{f(\hat{\beta},\hat{b})} + 2\pi|H|$$

- For likelihoods, integrating out a random effect is still a likelihood
- For a Cox partial likelihood, does integrating out a random effect still give a valid PL?
  - ⋆ Ripatti and Palmgren, JASA, show that it does
- Laplace approximations don't always work well
  - ⋆ The Cox PL is *extremely* quadratic
  - ⋆ Newton-Raphson iteration never fails
  - ⋆ Recent tests show excellent agreement

How good is the Laplace?

- The coxme functions allows iterative refinement of the solution, for any fixed value of θ,
- It uses monte carlo to estimate (integrand - approx integrand)
- For the diabetes data
  - -851.11369 : approximate integrated PL (Laplace)
  - -851.11361 : corrected
- For an admittedly small set of examples, this is the largest difference I have yet found.

- There are two ways to do the Taylor series
  - b around $\hat{b}$, treating $\hat{\beta}$ as fixed
  - b around $(\hat{b}, \hat{\beta})$

- By analogy with linear mixed models, Yau and McGilchrist define
  - "ML" estimate: $-\log|(H_{bb})^{-1}|$
  - "REML" estimate: $-\log|(H^{-1})_{bb})|$

- There is some early evidence from Cortinas (PhD thesis) that the 'REML' is not as stable.

- The `coxme` function only does the ML estimate.

```
> coxph(Surv(time, status) ~ rx + frailty(litter, dist='gauss'),
           data=rats)
                           coef se(coef)   se2 Chisq   DF      p
                     rx 0.913 0.323    0.319  8.01  1.0 0.0046
frailty(litter, dist = "g                     15.57 11.9 0.2100

Iterations: 6 outer, 16 Newton-Raphson
     Variance of random effect= 0.412
Degrees of freedom for terms=  1.0 11.9
Likelihood ratio test=35.3  on 12.87 df, p=0.000712  n= 150

> coxme(Surv(time, status) ~ rx, data=rats,
                random = ~1 | litter)
  Data: rats
  n= 150
  Iterations= 6 73
                     NULL Integrated Penalized
Log-likelihood -185.6556   -180.849  -173.774

  Penalized loglik: chisq= 23.76 on 13.17 degrees of freedom, p= 0.036
 Integrated loglik: chisq= 9.61 on 2 degrees of freedom, p= 0.0082

Fixed effects: Surv(time, status) ~ rx
       coef exp(coef)  se(coef)    z      p
rx 0.913261  2.492437 0.3226852 2.83 0.0047

Random effects:  ~ 1 | litter
             litter
```

```
Variance: 0.4255484
```

Diabetic Retinopathy Data

Diabetic retinopathy is a complication associated with diabetes mellitus; which can cause abnormalities in the microvasculature of the retina which in turn can lead to macular edema and visual loss.

It is leading cause of blindness in patients under 60 years of age in the United States.

Between 1972 and 1975 seventeen hundred forty-two patients were enrolled in a multi-center study to evaluate the efficacy of photocoagulation treatment for proliferative diabetic retinopathy; photocoagulation was randomly assigned to one eye of each study patient, with the other eye serving as an untreated control. A major goal was to assess whether treatment significantly delayed the onset of severe visual loss.

We will use a subset 197 patients, which is a 50% sample of the high-risk patients as defined by the study. The data set has 394 observations.

The four models below are a marginal, stratified, gamma frailty and Gaussian frailty model for the diabetic retinopathy data set.

```
> coxph(Surv(time, status) ~ trt + adult + cluster(id), diabetes)
          coef exp(coef) se(coef) robust se      z       p
   trt -0.7789     0.459    0.169     0.149 -5.245 1.6e-07
 adult  0.0539     1.055    0.162     0.179  0.302 7.6e-01

Likelihood ratio test=22.5  on 2 df, p=1.31e-05  n= 394

> coxph(Surv(time, status) ~ trt +adult + strata(id), diabetes)
         coef exp(coef) se(coef)     z       p
   trt -0.962     0.382    0.202 -4.77 1.8e-06
 adult     NA        NA    0.000    NA      NA

Likelihood ratio test=25.5  on 1 df, p=4.45e-07  n= 394
```

Diabetic Retinopathy Data

```
> coxph(Surv(time, status) ~ trt +  adult + frailty(id), diabetes)

               coef se(coef)   se2  Chisq DF       p
        trt -0.911 0.174    0.171  27.31  1 1.7e-07
      adult  0.041 0.221    0.166   0.03  1 8.5e-01
frailty(id)                        113.79 84 1.7e-02

Iterations: 6 outer, 24 Newton-Raphson
     Variance of random effect= 0.851   I-likelihood = -850.8
Degrees of freedom for terms=  1.0  0.6 84.0
Likelihood ratio test=201  on 85.56 df, p=2.77e-11  n= 394

> coxme(Surv(time, status) ~ trt +  adult, diabetes,
          random= ~ 1 |id)
                    NULL Integrated Penalized
Log-likelihood -867.9858  -851.1117 -805.2744

  Penalized loglik: chisq= 125.42 on 73.68 degrees of freedom, p= 0.00016
 Integrated loglik: chisq= 33.75 on 3 degrees of freedom, p= 2.2e-07

Fixed effects: Surv(time, status) ~ trt + adult
              coef exp(coef)  se(coef)     z       p
  trt -0.89963021   0.40672 0.1742047 -5.16 2.4e-07
adult  0.06028536   1.06214 0.2095425  0.29 7.7e-01

Random effects:  ~ 1 | id
                 id
Variance: 0.7784043
```

The Gaussian and gamma frailty models both estimate a substantial random effect for subject, of size 0.78 and 0.85, respectively. (Gaussian random effect with REML was .77).

The estimated size of the treatment coefficient rises with the variance of the random effect, as is expected theoretically (Henderson JRSSB 99), but only by a small amount.

The stratified model, which treats each adult as a separate strata, is the most aggressive correction for possible correlation, and has the largest treatment effect but also the largest variance.

The adult/juvenile covariate is not estimable in the stratified model, since the covariate value is constant within a given stratum.

Interpreting the variance

The Gaussian random effects model had a substantial variance of 0.78 for the random effect.

What did I mean by *substantial*?

$$\lambda_i(t) = \lambda_0(t)e^{X_i\beta+b_i}$$

- The component b_i is the per-subject random effect
- $\sim N(0, .78)$, standard deviation of .88.
- An 'average' person has a risk which may be $\exp(.88) = 2.4$ fold higher or lower than the population as a whole.
- In a linear mixed-effects model, the variance of the random effect needs to be compared to the error variance
  - ⋆ if y were multiplied by 10, variances all change by 100

- In a Cox model, the values are absolute.

Survival of kidney catheters

The following data set is presented in McGilchrist and Aisbett (Biometrics 91), and has been used by several authors to illustrate frailty.

Each observation is the time to infection, at the point of insertion of the catheter, for kidney patients using portable dialysis equipment.

Catheters may be removed for reasons other than infection, in which case the observation is censored.

There are 38 patients, each with exactly 2 observations. Variables are the subject id, age, sex (1=male, 2=female), disease type (glomerulo nephritis, acute nephritis, polycystic kidney disease, and other), and the time to infection or censoring for each insertion.

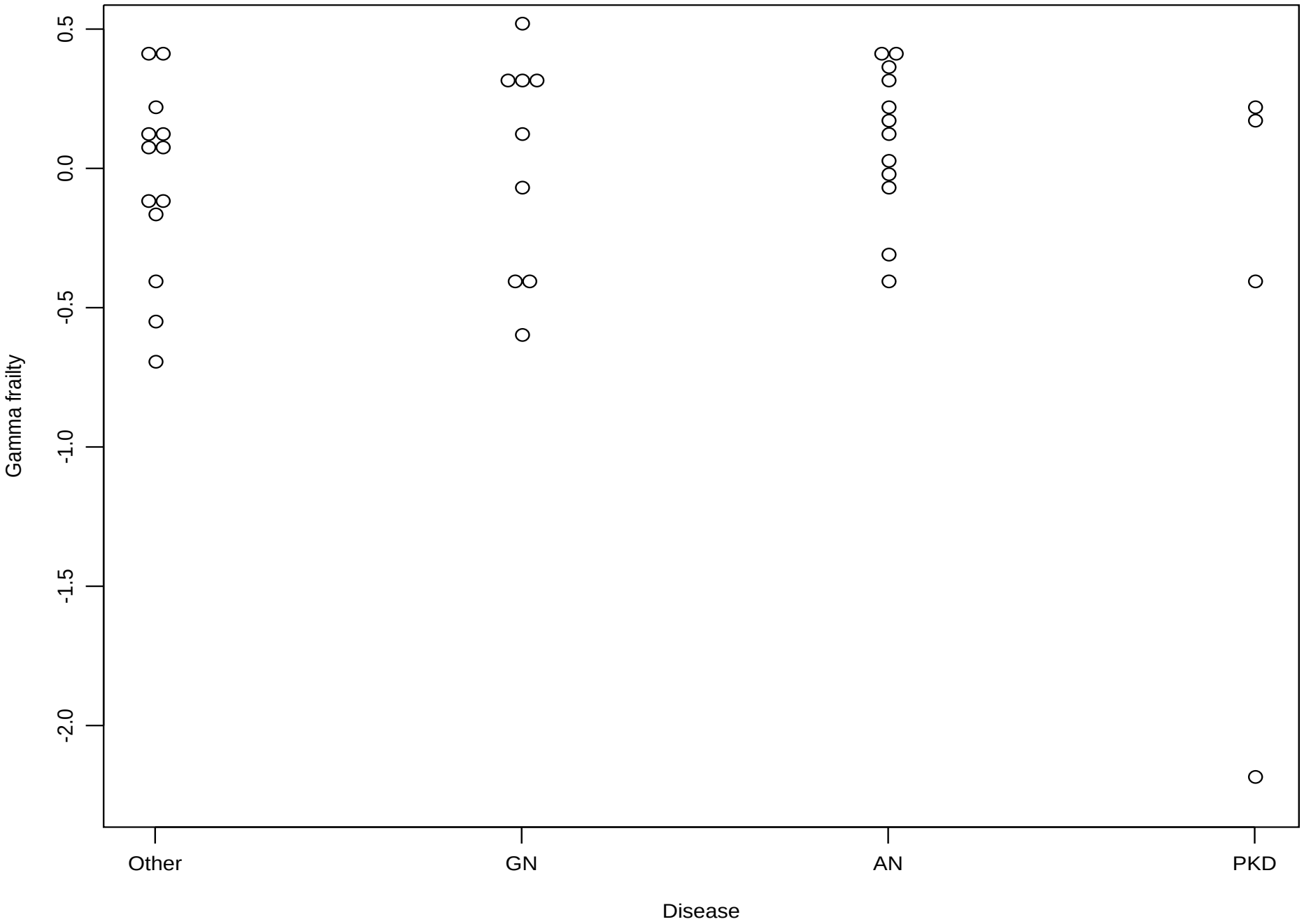
```
> kfit1 <- coxph(Surv(time, status) ~ age + sex, data=kidney)
> kfit2 <- coxph(Surv(time, status) ~ age + sex + disease,
                 data=kidney)
> kfit3 <-  coxph(Surv(time, status) ~ age + sex + disease +
                                      frailty(id), data=kidney)
> kfit3
                coef se(coef)    se2 Chisq DF       p
       age   0.00318 0.0111   0.0111  0.08 1  7.8e-01
       sex  -1.48314 0.3582   0.3582 17.14 1  3.5e-05
 diseaseGN   0.08796 0.4064   0.4064  0.05 1  8.3e-01
 diseaseAN   0.35079 0.3997   0.3997  0.77 1  3.8e-01
diseasePKD  -1.43111 0.6311   0.6311  5.14 1  2.3e-02
frailty(id)                           0.00 0  9.5e-01

Iterations: 6 outer, 29 Newton-Raphson
Penalized terms:
     Variance of random effect= 1.47e-07   M-likelihood = -179.1
Degrees of freedom for terms= 1 1 3 0
Likelihood ratio test=17.6  on 5 df, p=0.00342  n= 76
```

- PL for kfit1= -184.3, kfit2 = -179.1, with 2 and 5 degrees of freedom, respectively.
- Disease type is a significant addition.
- In kfit3, $\hat{\theta}$ is essentially 0.

When the disease variable is left out of the random effects model, however, we get a quite different result.

```
> kfit4 <- coxph(Surv(time, status) ~ age + sex + frailty(id),
        data=kidney)
> kfit4
                 coef se(coef)    se2 Chisq   DF       p
        age  0.00522 0.0119   0.0088  0.19  1.0 0.66000
        sex -1.58335 0.4594   0.3515 11.88  1.0 0.00057
frailty(id)                           22.97 12.9 0.04100

Iterations: 7 outer, 49 Newton-Raphson
     Variance of random effect= 0.408   M-likelihood = -181.6
Degrees of freedom for terms=  0.6  0.6 12.9
Likelihood ratio test=46.6  on 14.06 df, p=2.36e-05  n= 76
```

- Both the approximate Wald test and the likelihood show significance for θ.
- LR test $= 2(184.3 - 181.6) = 5.4$ on 1 df.


Gamma frailty
0.5
0.0
-0.5
-1.0
-1.5
-2.0
Other
GN
AN
PKD
Disease

There is one large outlier

- Subject 21, a 46 year old male, the median for the study is 45.5.
- 10 males, most had early failures
  - 2 observations were censored at 4 and 8 days, respectively
  - the remaining 16 male kidneys had a median time to infection of 19 days.
- Subject 21 had failures at 154 and 562 days.
- In a model with only age and gender, this subject has a martingale residual of -7.4; the next smallest is -1.8.

With this subject removed, neither the disease (p=0.53) nor the frailty ($p>0.9$) are important.

With this subject in the model, it is a toss-up whether the disease or the frailty term will be credited with ‘significance’.

Different frailty distributions do give different estimates. Unfortunately, some papers have used this to proclaim that the distribution matters — now if they had only *looked* at the data.

Mn Breast Ca Family Study

Original

- original case-control study of Elving Anderson in 1944-52
- 544 sequential breast cancer cases seen at the U of Minnesota
- information gathered on parents, siblings, offspring, aunts/uncles, and grandparents

Rebirth

- data exhumed in 1991 by Thomas Sellers
- exclude 58 prevalent BRCA and 19 with < 2 living relatives
- 462 families remain: 10 no living members, 23 lost, 8 refused
- 426 participating families

Current

- 26050 in the pedigrees
  - ⋆ 13351 male
  - ⋆ 12699 female (5183 marry-ins)

- 1063 breast cancers, 426/376/188

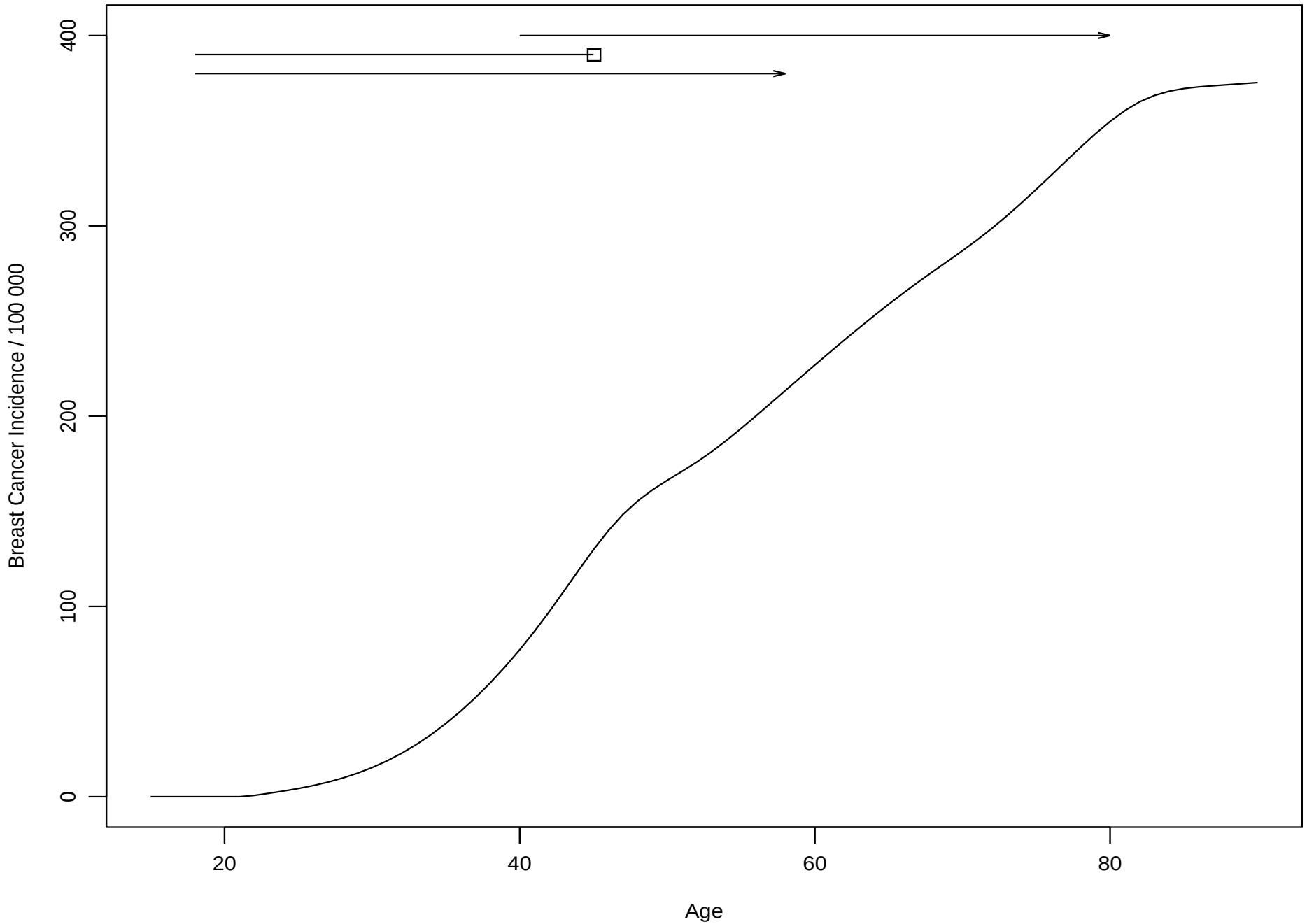
family size	1	4–20	21–50	51–100	> 100
count	8191	72	228	115	11

Questions

- BRCA1 and BRCA2
- Early life exposures
- Breast density
- Pharmacogenomics

Statistical Questions

1. What to do with the probands?
   - Exclude: wastes 426/1063 = 40% of our endpoints
   - Include: biased
2. What about the marry-ins?
   - kind of tied in and kind of not
   - treat those with no issue differently?
3. How to use age of onset?
   - some traditional approaches are lacking
   - survival models?
4. How to allow for birth cohort effects
5. Missing data
6. Which model?


Breast Cancer Incidence / 100 000
0
100
200
300
400
20
40
60
80
Age

- Do we need to account for within-family correlation, for assessment of other variables?
- How to best account for correlation?
- Is a within-family “random effect” of interest in itself?
- What model is best — proportional hazards or accelerated failure time?

```
coxph(Surv(startage, endage, ibrca) ~ parity + cluster(famid),
              data= cohort)

          coef   exp(coef) se(coef) robust se     z      p
parity -0.343      0.71     0.12     0.126     -2.71 0.0066
```

The overall correction for ‘famid’ is rather crude.

Excess risk

Create a model that incorporates important covariates (parity, OC use, age, . . .)

Look at the excess risk

$$\frac{\text{number of observed events}}{\text{expected, given the model}} = \frac{O}{E}$$

Excellent way to examine subgroups *if they are large enough*

Strategies for small subgroups

- Don't look at small subgroups
- Informal shrinkage (O + 1/2) / (E +1)
- Formal shrinkage
  - ⋆ Random effects
  - ⋆ Bayesian

Shared Random Effects Model

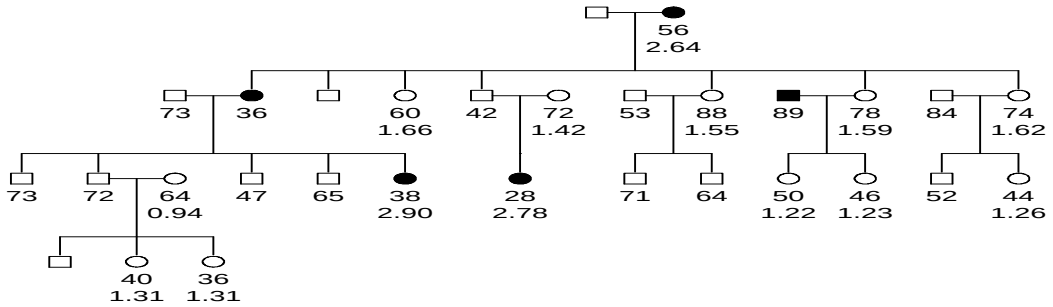
We can't genotype everyone. Can we make any progress with *latent* genetic models.

Assume that the risk for a given subject i from family j is

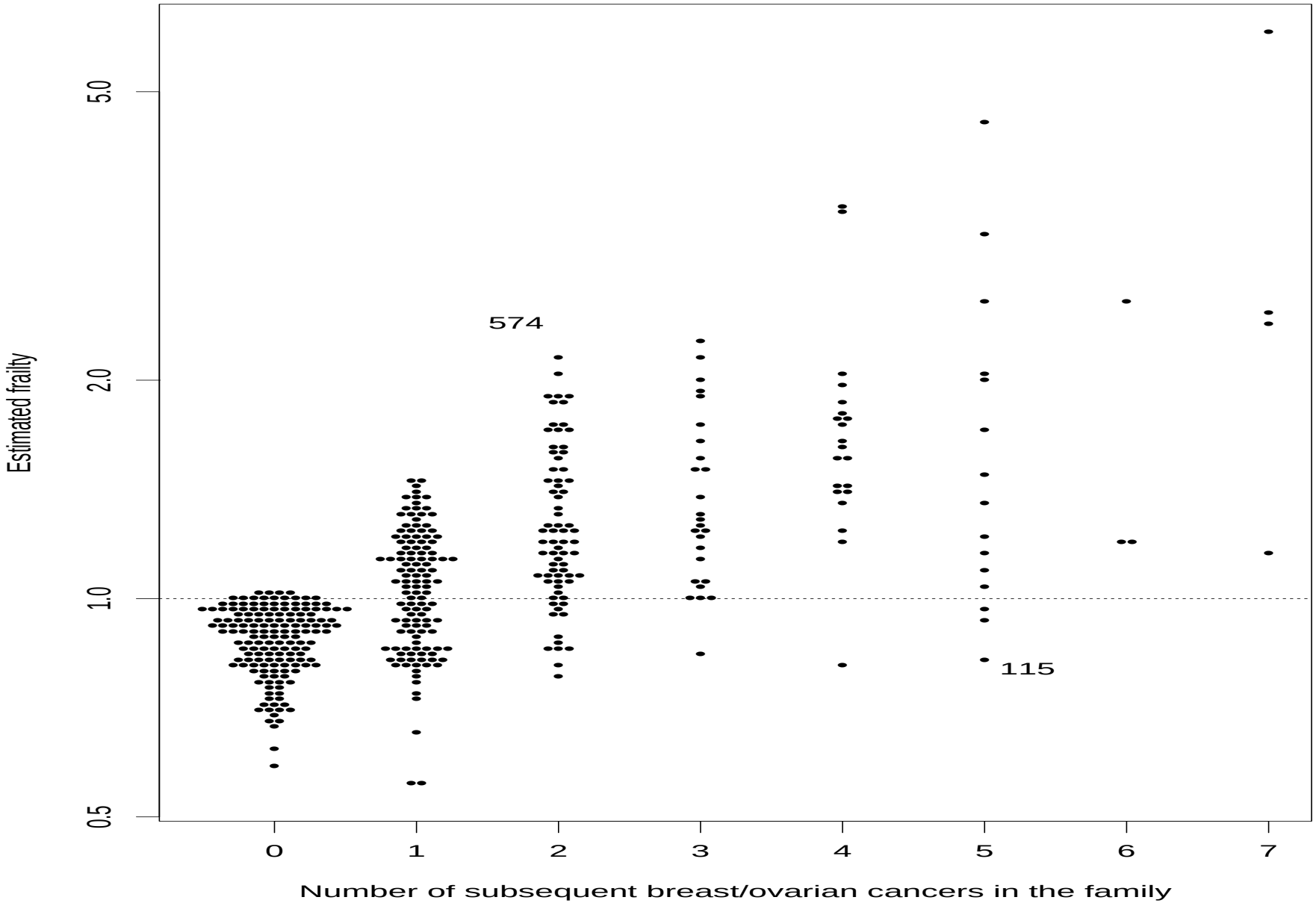
$$\begin{aligned} \lambda_i(a) &= \lambda_0(a)e^{X_i\beta+f_j} \\ f &\sim N(0,\sigma^2) \end{aligned}$$

where

- X are measured covariates (parity, estrogen use, etc)
- f_j is an excess risk or *frailty* that captures shared genetic or other effects
- a is age.


56
2.64
73
36
60
1.66
42
72
1.42
53
88
1.55
89
78
1.59
84
74
1.62
73
72
64
0.94
47
65
38
2.90
28
2.78
71
64
50
1.22
46
1.23
52
44
1.26
40
1.31
36
1.31

- parity is 1 for those women with 1 or more offspring and 0 for nulliparous women
- It is a powerful effect, reducing risk by nearly a third.
- The frailty coefficients under a Gaussian model are constrained to have mean 0, the average value for the marry-in cases was -0.005.
- There have been 5 identified breast/ovarian cancers in family 115 among 33 blood relatives (not counting the initial index case), and the overall risk estimate for the family is below the average of the marry-in subjects.
- Family 574 is estimated to have high risk; there are 2 breast cancers and only 3 blood relatives, and one of the cancers occurred before the age of 35.


Estimated frailty
0.5
1.0
2.0
5.0
574
115
0
1
2
3
4
5
6
7
Number of subsequent breast/ovarian cancers in the family

Goal: for screening studies, select those families with the greatest likelihood of inherited risk.

- Total number of affecteds per family was a first suggestion for the selection protocol
- it does not take into account the total family size, the age-intervals of follow-up, or the ages at which cancers occurred
- The plot more clearly shows the high risk families
- 4383 frailty coefficients, 3961 marry-in.
- Useful tool even if the model isn't completely correct (we don't necessarily believe a Gaussian frailty).

Actual Splus code

```
> fit1 <- coxph(Surv(startage, endage, cancer) ~ parity0,  breast,
subset=(sex=='F' & proband==0) )
> fit1
          coef exp(coef) se(coef)     z      p
parity0 -0.303     0.739    0.116 -2.62 0.0088

Likelihood ratio test=6.35  on 1 df, p=0.0117
  n=9399 (2876 observations deleted due to missing values)
>fit1$loglik
 -5186.994 -5183.817
```

Several subjects are missing information on either parity (1452), follow-up time/status (2705) or both (1276).

Many of these were advanced in age when the study began and were not available for the follow-up questionnaire in 1991.

```
> coxme(Surv(startage, endage, cancer) ~ parity0, breast,
random= ~1|famid, subset=(sex=='F'& proband==0))

                    NULL Integrated Penalized
Log-likelihood -5186.994  -5174.865 -5121.984

  Penalized loglik: chisq= 130.02 on 90.59 degrees of freedom, p= 0.0042
 Integrated loglik: chisq= 24.26 on 2 degrees of freedom, p= 5.4e-06

              coef exp(coef)  se(coef)     z      p
parity0 -0.3021981 0.7391916 0.1172174 -2.58 0.0099

Random effects:  ~ 1 | famid
              famid
Variance: 0.2090332
```

Penalized Models

The model can be viewed in 2 ways

- Random effects model
  - ⋆ Loglik of 2*(5187 - 5175.9) = 24.3 on 2 degrees of freedom
  - ⋆ Parameters: β = parity0, σ^2 = variance of b
- Penalized fixed effects model
  - ⋆ Loglik of 2*(5187 - 5123) = 130 on 90.6 degrees of freedom
  - ⋆ Parameters: β = parity0, b = 426 family effects
  - ⋆ Shrinkage of the family effects towards 0 means 89.6 ‘effective’ degrees of freedom, rather than 425.
  - ⋆ Preferably choose the amount of shrinkage by AIC or BIC.
  - ⋆ AIC = PL - 2*df, maximize.
- Both views are valid.

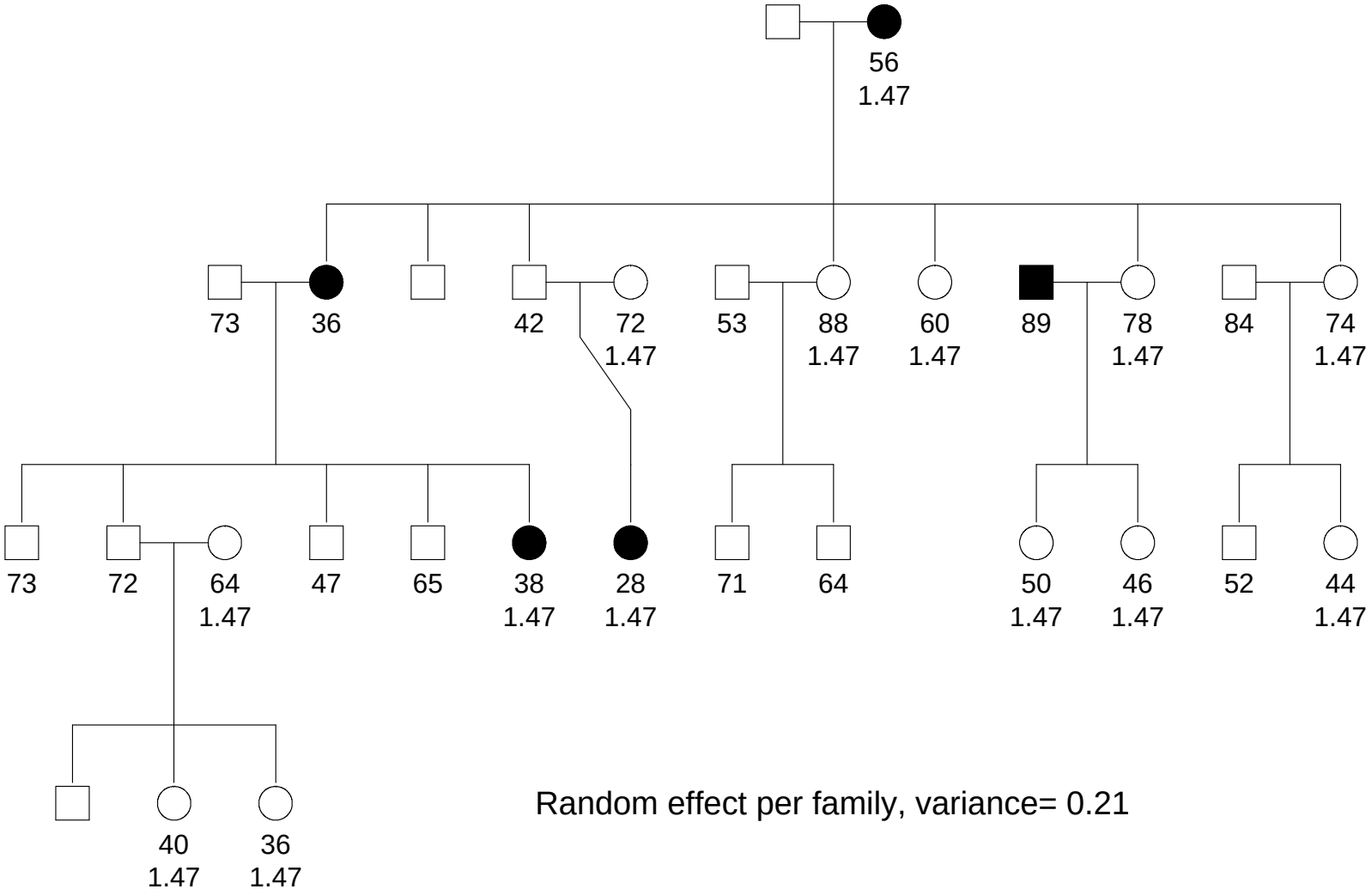
There is a substantial family effect, with a variance of .209.

$$\sqrt{.209} = .46,\ e^{.46} = 1.6$$

Particular families commonly have 60% higher (or lower) risk than others.

However, it is applied in a way that is not entirely satisfying.

Family 8



Random effect per family, variance= 0.21

Modified family effect

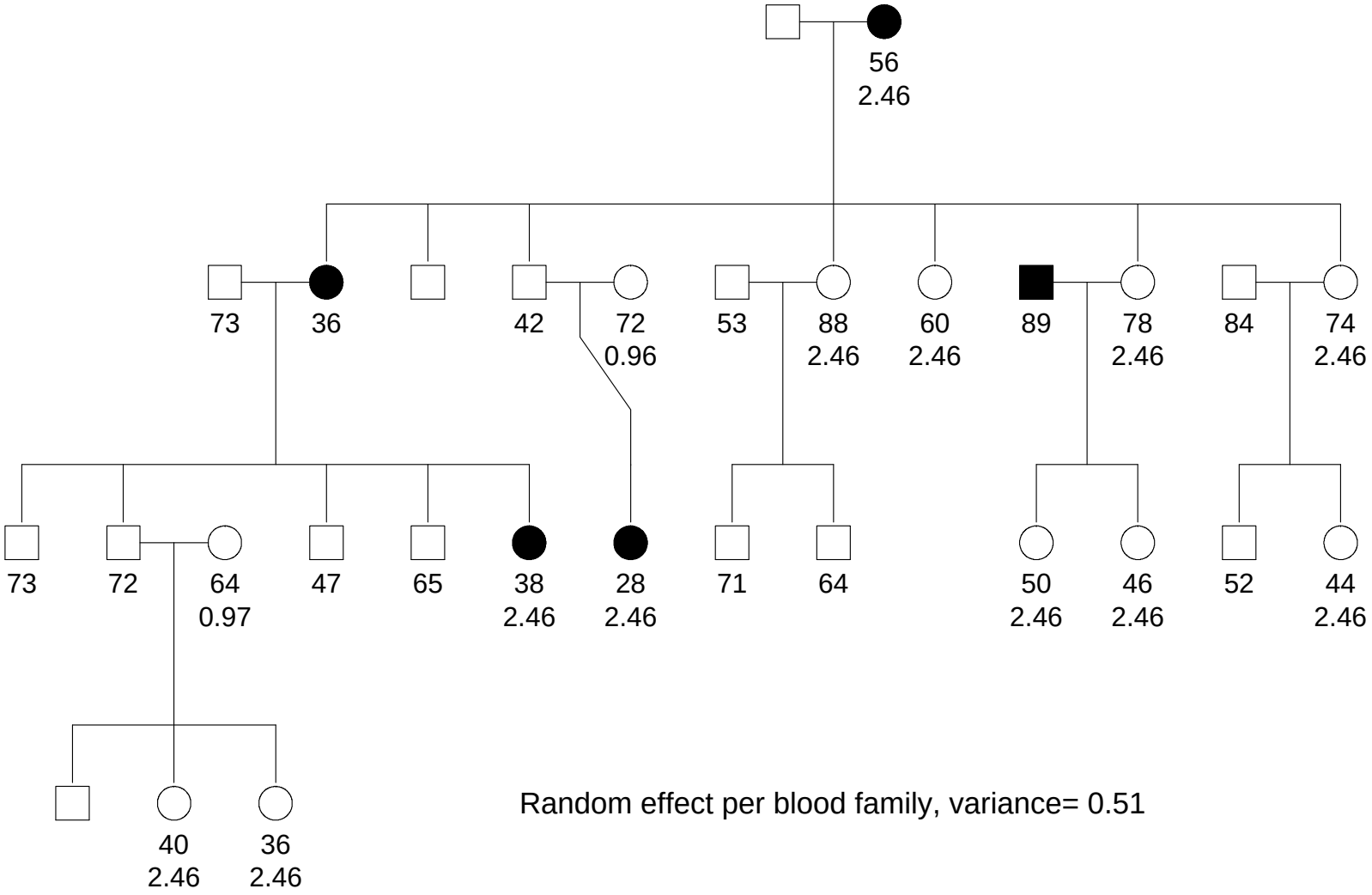
- It is clearly incorrect for a marry-in to “inherit” the familial breast cancer risk of their engrafted family

- Idea 1: Each marry-in has her own risk
  - ⋆ Solves the problem, partly (marry-in with children)
  - ⋆ Lots of degrees of freedom, in particular n=1 groups
  - ⋆ Groups of this size are not identifiable

- Idea 2: All marry-ins are at backround risk
  - ⋆ Marry-ins are a part of family 1000: “Minnesota”
  - ⋆ Solves the df problem

```
> tempid1 <- ifelse(breast$bloodrel, breast$famid, 1000 + 1:nrow(breast))
> tempid2 <- ifelse(breast$bloodrel, breast$famid, 1000)
> fit2 <- coxme(Surv(startage, endage, cancer) ~ parity0 , breast,
               random= ~1|tempid1, subset=(sex=='F' & proband==0))
> fit3 <- coxme(Surv(startage, endage, cancer) ~ parity0 , breast,
               random= ~1|tempid2, subset=(sex=='F' & proband==0))
> fit2

  n=9399 (2876 observations deleted due to missing values)
  Iterations= 4 50
                     NULL Integrated Penalized
Log-likelihood -5186.994  -5163.226 -5031.745

  Penalized loglik: chisq= 310.5 on 232.65 degrees of freedom, p= 0.00048
 Integrated loglik: chisq= 47.54 on 2 degrees of freedom, p= 4.8e-11

Fixed effects: Surv(startage, endage, cancer) ~ parity0
              coef exp(coef) se(coef)     z     p
parity0 -0.2935372 0.7456215 0.119258 -2.46 0.014

Random effects:  ~ 1 | tempid1
            tempid1
Variance: 0.5063702
```

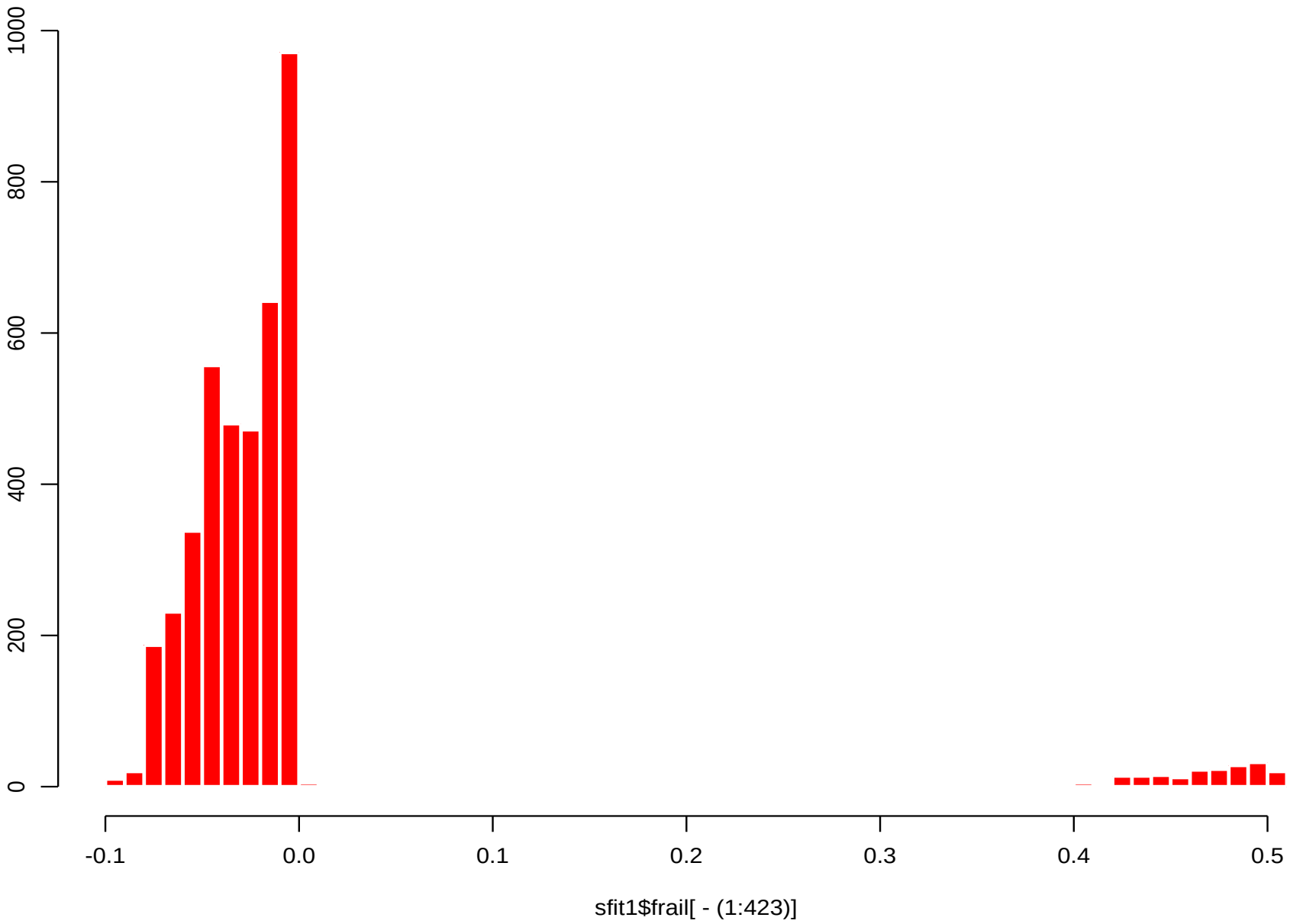
Family 8



Random effect per blood family, variance= 0.51

```
> length(fit2$frail)
[1] 424

        600         601         603         605        1000
-0.2381351 -0.0428534 -0.2052715 -0.3235427 -0.5020491
```

- Marry-ins are now more realistic
- But not connected to offspring
- Larger random effect .51 vs .21


1000
800
600
400
200
0
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
sfit1$frail[- (1:423)]

```
> fit2
Cox mixed-effects model fit by maximum likelihood
  Data: main
  Subset: (sex == "F" & proband == 0)
  n=9399 (2876 observations deleted due to missing values)
  Iterations= 7 82
                      NULL Integrated Penalized
Log-likelihood -5186.994  -5147.241 -5064.799

  Penalized loglik: chisq= 244.39 on 127.84 degrees
          of freedom, p= 2.5e-09
 Integrated loglik: chisq= 79.51 on 2 degrees of freedom,
          p= 0

Fixed effects: Surv(startage, endage, cancer) ~ parity0
               coef exp(coef)  se(coef)     z     p
parity0 -0.2283946 0.7958102 0.1193444 -1.91 0.056
```

```
Random effects:  ~ 1 | tempid2
            tempid2
Variance: 0.5141345

> length(fit2$frail)
[1] 424

> sfit$frail[420:424]
        600         601         603         605        1000
 -0.2381351 -0.0428534 -0.2052715 -0.3235427 -0.5020491
```

Sparse computation

- In fit2 above there are `length(fit2$frail)` = 4527 random effects and 1 fixed effect.
- For sanity, the program can't deal with a 4528 $\times$ 4528 Hessian matrix.
- By default, any random coefficient corresponding to a group that represents 2% or less of the data is kept as a sparse term. Cross-product terms between it and other sparse terms are not stored or computed.

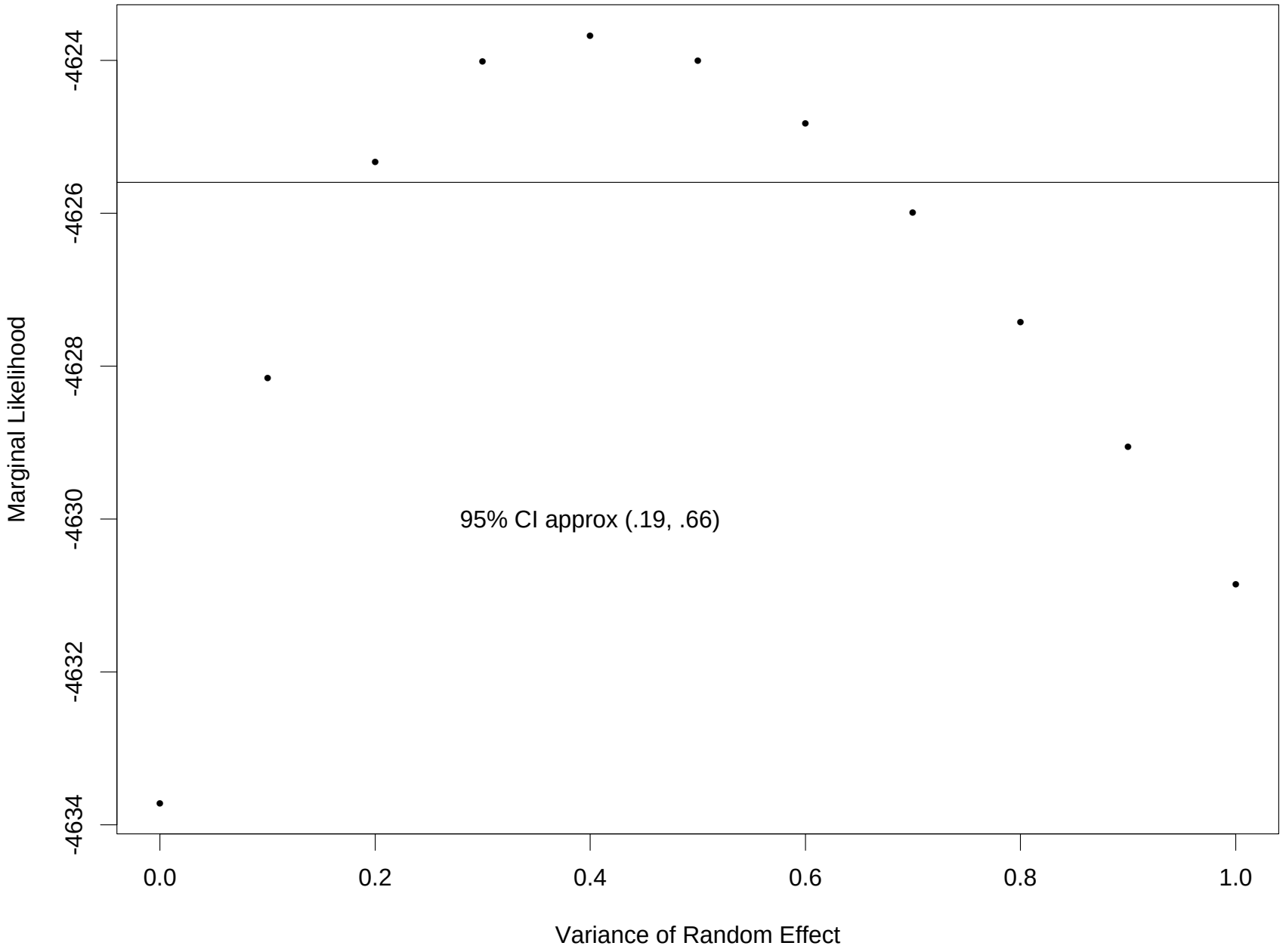
- Let p be the vector of proportions for a group variable.
  - ⋆ Diagonal elements of the Cox model's second derivative matrix H are of order $p_i(1-p_i)$
  - ⋆ Off diagonal elements of H are of order $p_i p_j$
  - ⋆ For the 2% default, we are ignoring terms that are .02 of the diagonal; for this data, terms are approx 1/4000 of the diagonal.
  - ⋆ The NR steps are nearly as good as with the full H matrix.
  - ⋆ There is a similar small effect on the Laplace approximation.

- If there are multiple random effects, only one is allowed to be sparse.

Profile likelihood for the frailty term

```
> theta <- seq(0, 1, .05)
> profile <- rep(0, length(theta))
> for (i in 1:length(theta)) {
    fit <- coxme(Surv(startage, endage, cancer) ~ parity0,
          data=breast, subset=(sex=='F' & proband==0),
                  random = ~1|tempid1, variance=theta[i]))
    profile[i] <- fit$loglik[2]
    }
> plot(theta, profile)
```

- The profile likelihood is often very asymmetric.
- The formulas for se(sigma) are nasty.
- The program does not even provide the estimate, in order to remove temptation.


Marginal Likelihood
-4624
-4626
-4628
-4630
-4632
-4634
95% CI approx (.19, .66)
0.0
0.2
0.4
0.6
0.8
1.0
Variance of Random Effect

Individual frailty

```
>  coxme(Surv(startage, endage, cancer)  ~parity0, main,
          random=~1|gid, subset=(sex=='F' & proband==0))

  n=9399 (2876 observations deleted due to missing values)
  Iterations= 7 77
                     NULL Integrated Penalized
Log-likelihood -5186.994  -5183.815 -5165.126

  Penalized loglik: chisq= 43.74 on 38.25 degrees of freedom, p= 0.25
 Integrated loglik: chisq= 6.36 on 2 degrees of freedom, p= 0.042

Fixed effects: Surv(startage, endage, cancer) ~ parity0
             coef exp(coef)  se(coef)     z     p
parity0 -0.303778 0.7380247 0.1162148 -2.61 0.009

Random effects:  ~ 1 | gid
                gid
Variance: 0.0611838
```

- Makes no use of relatedness information at all
- Not identifiable

Problems with the shared frailty fit:

- Some families are very large (max=130). Can one parameter capture an entire family?
- A genetic effect might be passed onto a single branch.
- What about children of marry-in subjects?

Correlated Frailty

Assume that the risk for subject i is

$$\begin{aligned} \lambda_i(a) &= \lambda_0(a)e^{X_i\beta+b_i} \\ b &\sim N(0, \sigma^2 K) \end{aligned}$$

where K is a *kinship* matrix. The ij element describes the relatedness of subjects i and j:

$$\begin{aligned} 1 &= \text{self} \\ 1/2 &= \text{mother/daughter, sisters, . . .} \\ 1/4 &= \text{uncle/niece, . . .} \\ 0 &= \text{different families} \end{aligned}$$

- To create a kinship matrix for the females, you need *all* the people.
  - ⋆ If men are left out, my sister and my daughter will appear unrelated.
- Data set of family id, my id, father id, mother id
  - ⋆ Ancillary information: twins
- The *subset* of coxme operates on the K matrix.

```
> newfam <- makefamid(main$gid, main$dadid, main$momid)
> kmat   <- makekinship(newfam, main$gid, main$dadid, main$mo

> kfit1 <- coxme(Surv(startage, endage, cancer) ~ parity0, ma
                 random= ~1|gid, varlist=kmat,
subset=(sex=='F' & proband==0))

>kfit1
  n=9399 (2876 observations deleted due to missing values)
  Iterations= 3 38
                     NULL Integrated Penalized
Log-likelihood -5186.994  -5170.335 -4910.234

  Penalized loglik: chisq= 553.52 on 489.55 degrees of freedo
 Integrated loglik: chisq= 33.32 on 2 degrees of freedom, p=

Fixed effects: Surv(startage, endage, cancer) ~ parity0
              coef exp(coef)  se(coef)     z       p
parity0 -0.321735 0.7248902 0.1224286 -2.63 0.0086
```

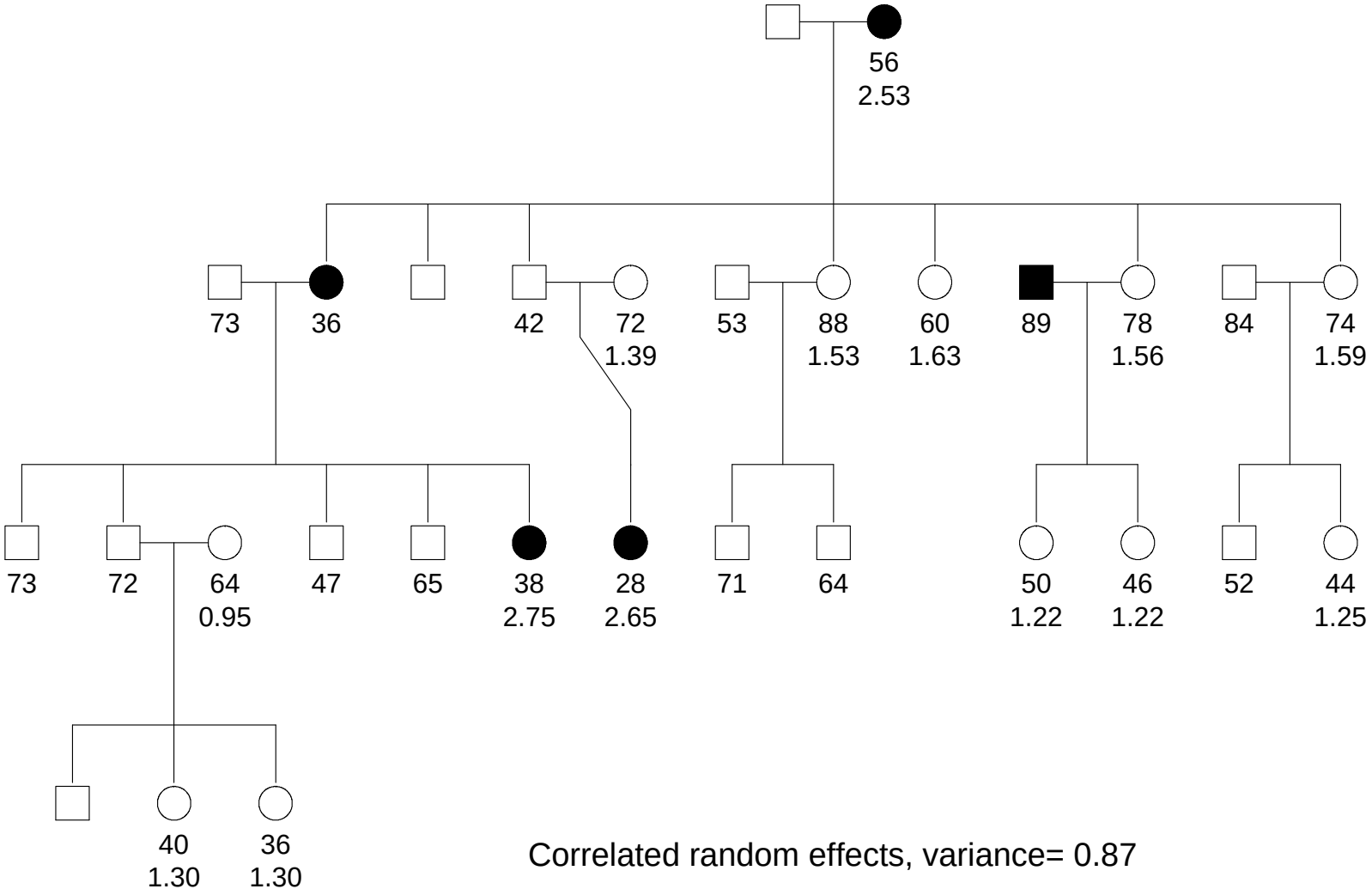
```
Random effects:  ~ 1 | gid
 Variance list: kmat
                gid
Variance: 0.9093012

> length(kfit1$frail)
[1] 9399
```

- Make famid: construct a *family id* that marks disjoint families.

- Make kinship: create the kinship matrix
  - Separately for each family
  - Stored as a bds matrix object

- The kinship matrix is block-diagonal
  - Relationship between families = 0

Family 8



Correlated random effects, variance= 0.87

- The variance of the random effect is much larger in the model 0.91
- There are 9399 random effects b_i, one per person
- But only 490 ‘effective’ degrees of freedom, due to the rich correlation structure

Connections between breast and prostate cancer

Question: is there evidence for common genetic risk factors?

A sub-study was carried out, by obtaining PCA information from 141 families.

- 60 high risk: at least 4 breast cancers
- 81 low risk: no breast cancers beyond the original proband

Male relatives over the age of 40 were screened for prostate cancer via a survey.

Three models

1. Common genes: each person's risk of cancer depends on that of both male and female relatives

2. Separate genes: a female's risk of cancer is linked to the risk of her female relatives, a male's is linked to that of his male relatives, but there is no interdependency

3. Combined: some risk of each type exists

Tools: Create variants of the kinship matrix which have no male/female correlation.

$$\lambda_i(t) = \lambda_{0,1}(t)e^{X\beta+b_i}$$

λ_0 = baseline risk for breast cancer, females

λ_1 = baseline risk for prostate cancer, males

β = covariates = parity

For simplicity assume a single family with parents, daughter, grandson, and granddaughter.

The kinship matrix K is

	Father	Mother	Daughter	Grandson	Granddauther
Father	1	0	1/2	1/4	1/4
Mother	0	1	1/2	1/4	1/4
Daughter	1/2	1/2	1	1/2	1/2
Grandson	1/4	1/4	1/2	1	1/2
Granddaughter	1/4	1/4	1/2	1/2	1

The correlated frailty models fit $b \sim N(0, \sigma^2 K)$

We want to fit a model with

	Father	Mother	Daughter	Grandson	Granddauther
Father	a	0	c(1/2)	a(1/4)	c(1/4)
Mother	0	b	b(1/2)	c(1/4)	b(1/4)
Daughter	c(1/2)	b(1/2)	b	c(1/2)	b(1/2)
Grandson	a(1/4)	c(1/4)	c(1/2)	a	c(1/2)
Granddaughter	c(1/4)	b(1/4)	b(1/2)	c(1/2)	b

where $a = \sigma_1^2$, $b = \sigma_2^2$ and $c = \sigma_3^2$ are three components of variance for M/M, F/F, and M/F reactions.

How do we do this?

The coxme routine allows a random effect to have variance

$$\sigma_1^2 A_1 + \sigma_2^2 A_2 + \sigma_3^2 A_3 + \ldots$$

where A_1, A_2 and A_3 are known matrices. It then estimates σ_1^2, σ_2^2, and σ_3^2.

Let

$$K = \begin{bmatrix} K_1 & K_3 \\ K_3 & K_2 \end{bmatrix}$$

- Data set has been ordered as males, then females
- K_1 = male-male relationships
- K_2 = female-female relationships
- K_3 = male-female relationships

$$\text{kmat.m} = \begin{pmatrix} K_1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\text{kmat.f} = \begin{pmatrix} 0 & 0 \\ 0 & K_2 \end{pmatrix}$$

$$\text{kmat.mf} = \begin{pmatrix} 0 & K_3 \\ K_3 & 0 \end{pmatrix}$$

In reality, sparseness requires K not be reordered. Families stay together.

kmat.m = K, with all m/f and f/f elements set to 0.

kmat.f = K, with all m/f and m/m elements set to 0.

kmat.mf = K, with all m/m and f/f elements set to 0.

lmat.m + kmat.f + kmat.mf = K

Common variance model

- Knowing my sister's breast cancer history is informative about my prostate cancer risk
- It is as informative as knowing my male relatives' prostate history
- Hypothesis: common genes

$$b \sim N(0, \sigma^2 K)$$

```
#
# Fit the overall model
#
nfit1 <- coxme(Surv(startage, endage, cancer) ~ parity0 +
                 strata(sex),
                 main, random= ~1|gid,
                 subset=(proband==0), varlist=kmat)

  n=13165 (12459 observations deleted due to missing values)
  Iterations= 7 77
                       NULL Integrated Penalized
Log-likelihood -6145.714  -6129.606 -5862.736

  Penalized loglik: chisq= 565.96 on 504.11 degrees of
                 freedom, p= 0.029
 Integrated loglik: chisq= 32.22 on 2 degrees of freedom,
                 p= 1e-07

Fixed effects: Surv(startage, endage, cancer) ~ parity0
                 + strata(sex)coef exp(coef)
```

```
                se(coef)z      p
parity0 -0.3253943 0.7222425 0.1212548 -2.68 0.0073

Random effects:  ~ 1 | gid
 Variance list: kmat
                gid
Variance: 0.7585038
```

Note, the majority of the 12,459 deleted due to missing values are men who were not a part of the substudy.

Separate Model

- Male relatives share prostate risk
- Female relative share breast risk
- Knowledge of the males is not informative for the females' risk and vice-versa
- Hypothesis: no common genes

$$b \sim N(0, \sigma_1^2, K_1 + \sigma_2^2 K_2)$$

```
nfit2 <- coxme(Surv(startage, endage, cancer) ~ parity0 +
                          strata(sex),
                 main, random= ~1|gid,
              subset=(proband==0), varlist=list
                  (kmat.m, kmat.f))

  n=13165 (12459 observations deleted due to missing values)
  Iterations= 7 117
                      NULL Integrated Penalized
Log-likelihood -6145.714  -6127.533 -5810.144

 Penalized loglik: chisq= 671.14 on 597.61 degrees
                    of freedom, p= 0.019
Integrated loglik: chisq= 36.36 on 3 degrees of freedom,
                    p= 6.3e-08
Fixed effects: Surv(startage, endage, cancer) ~ parity0
                   + strata(sex)coef exp(coef)  se(coef)     z
parity0 -0.3219113 0.7247625 0.1224996 -2.63 0.0086

Random effects:  ~ 1 | gid
```

```
 Variance list: list(kmat.m, kmat.f)
                    gid1      gid2
Variance: 0.8374137 0.9190065
```

It is equivalent to fitting the two models separately

```
> nfit2a <- coxme(Surv(startage, endage, cancer) ~ parity0,
                  data = main,
                  random = ~1| gid, varlist =kmat,
 subset=(sex=='F' & proband==0))
> nfit2b <- coxme(Surv(startage, endage, cancer) ~ 1,
                main, random= ~1|gid,
                subset=(proband==0 & sex=='M'),
                varlist=kmat)

> nfit2a$loglik[2] + nfit2b$loglik[2]
 Integrated
  -6127.533

> nfit2a$coef

> sfit3$coef
    parity0         gid
-0.3219134    0.918989
```

Combined Model

- Males are informative about males with coef σ_1^2
- Females are informative about females with coef σ_2^2
- Male history gives *some* information about female risk, and vice-versa, σ_3^2

$$b \sim N(0, \sigma^2 K_1 + \sigma_2^2 K_2 + \sigma_3^2 K_3)$$

```
> nfit3 <- coxme(Surv(startage, endage, cancer)
                  ~ parity0 + strata(sex),
                 main, random= ~1|gid,
                 subset=(proband==0), rescale=F, pdcheck=F,
                 varlist=list(kmat.m, kmat.f, kmat.i))

  n=13165 (12459 observations deleted due to missing values)
  Iterations= 10 213
                       NULL Integrated Penalized
Log-likelihood -6145.714  -6127.091 -5806.212

  Penalized loglik: chisq= 679 on 603.42 degrees of
                 freedom, p= 0.017
 Integrated loglik: chisq= 37.25 on 4 degrees of
                 freedom, p= 1.6e-07
```

```
Fixed effects: Surv(startage, endage, cancer) ~ parity0
               + strata(sex)coef exp(coef)  se(coef)     z
parity0 -0.3242092  0.723099 0.1225147 -2.65 0.0081

Random effects:  ~ 1 | gid
 Variance list: list(kmat.m, kmat.f, kmat.i)
                gid1      gid2      gid3
Variance: 0.8806324 0.9232614 0.2713364
```

Results

	Variance			
	M/F	F/F	M/M	$2\mathcal{L}$
Common	0.76	0.76	0.76	64.4
Separate	0	0.92	0.84	72.7
Combined	0.27	0.92	0.88	74.5

- The separate effects model fits better than the common effect model
  - ⋆ Chisquared of 8.3 on 1 df
  - ⋆ Female relatives are more informative about breast cancer than male relatives are about prostate cancer
  - ⋆ Both are very useful, however.
- The combined model shows some evidence for crossover
  - ⋆ The estimate of .27 is much smaller then the M/M and F/F estimates
  - ⋆ Chisquared of 1.8 on 1 df, p= .18
  - ⋆ The confidence interval crosses 0

Random Center Effects

Thesis work of Jose Cortinas Abrahantes, Limburgs Universitair Centrum.

- Simulation based on clinical trail of bladder cancer
- 37 centers
- 2323 patients
- Enrollments of 20–29 (5), 30–39 (11), 40–59 (8), 60–89 (6), 91, 104, 116, 120, 155, 183, 247.
- 40 or 60% censoring
- Random institution effect + random treatment effect.

Conclusions

Moderate censoring, $\beta = 0.7$

σ^2	Estimates	
	β	σ^2
.04	.715 (.119, .094)	.038
.08	.702 (.083, .095)	.076
.4	.691 (.122, .165)	.383

The Ripatti method does quite well.

Heavy censoring, $\beta = -.182$

σ^2	Estimates	
	β	σ^2
.04	-.166 (.092, .130)	.044
.08	-.176 (.100, .162)	.079
.4	-.156 (.143, .160)	.392

Ocassional convergence problems.

Bootstrapping the fit

- Choose a set of *families* from the data
- Sample those families, with replacement
- Relabel the new family id
- Fit
- Repeat

Shrinkage

Look at the excess risk for each family:

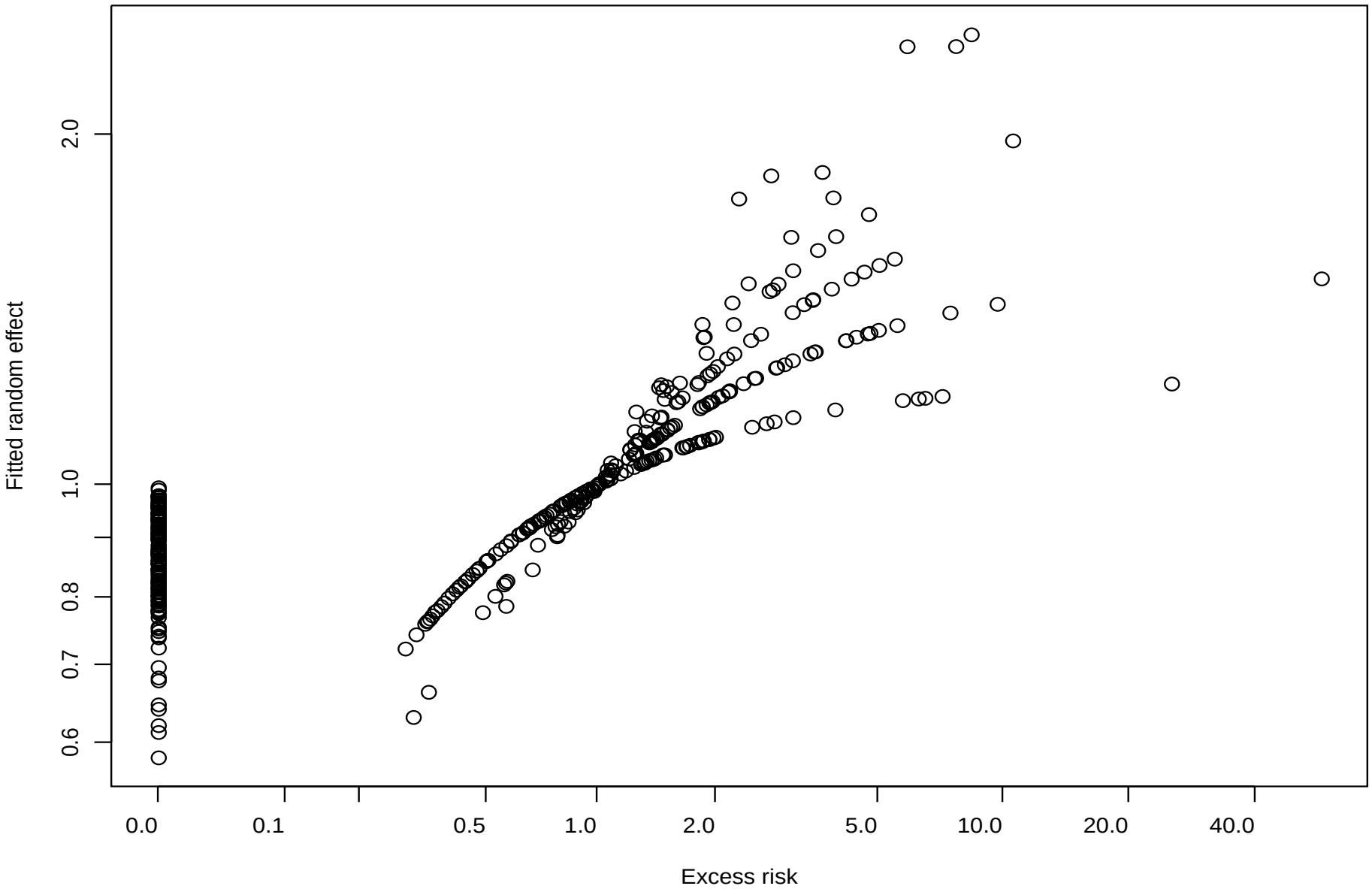
$$\frac{\sum O_i}{\sum E_i}$$

where E_i is the expected from the simple Cox model (fit1).

Values range from 0 (immortals) to 40 (very high risk).

The random effects are smaller

```
> range(bfit1$frail)
[1] -0.5411219  0.8903696

> var(bfit1$frail)
[1] 0.0438545
```


Fitted random effect
0.6
0.7
0.8
1.0
2.0
0.0
0.1
0.5
1.0
2.0
5.0
10.0
20.0
40.0
Excess risk

Shrinkage

Remember the formula for a gamma random effect

$$b_i = \ln\left(\frac{O_i + 1\sigma^2}{E_i + 1\sigma^2}\right)$$

With $1/\sigma^2 \approx 23$

For small families (O and $E < 2$), the estimated random effect for the family is close to 0.

1. The estimated variance term from the model is for the *true* family information.
   - What I would have with full knowledge (1000's of members)
2. The estimated b_i for each family is a best estimate, after the fact, with small information, for that family

Code questions

What it does it does well.

There is a lot it doesn't do.

- Random slopes
- Nested random effects within a strata
- Balanced random effect, given randomization
- Biggest issue: control language and syntax.

Current flaws:

- Residuals not formalized.
- The tail of L_g can be very flat $\rightarrow$ unreasonable solutions.

- When σ^2 is very large, the NR iteration may fail.
- Wider testing.

Statistical questions

- How much information do we need per ‘random’ effect to get stable solutions?
- What is the best way to estimate the variance of the random effect: MLE, REML, AIC, . . . ?
- Are the Wald and LR tests reliable – or even consistent – when the number of coefficients grows with n?
- What is the correct n for a BIC estimate?
- How good is the variance estimate for the fixed effects?
- How important is the correct χ^2 distribution to the tests?

GEE models

- A major shortcoming of the marginal models is a lack of correlation structure.
- Equivalent to GEE models, restricted to *working independence*
- An extension to other correlation stuctures is possible
  - ⋆ A Zhang thesis
- Not yet added to any packages

Random Effects Models

- Growing in popularity
- Require moderate to large data sets
- Significant advantages over the marginal (GEE) approach
  - ⋆ IF you are correct about the correlation structure
  - ⋆ If there is enough data
- Multiple choices for the form of the random effect
  - ⋆ Gaussian, log-gamma, positive stable
  - ⋆ Does not appear to make a reproducable difference in the fit
- Very complicated models have been suggested
- An area of active research

Expected Survival

The calculation of an expected survival (based on some reference population) for a cohort of patients under study has a long history.

- Methods for a census based reference population are most familiar
  rate table = United States population by age and sex

- Recently, these ideas have been rediscovered and applied to the proportional hazards model
  rate table = a prior Cox model

- An important distinction is
  - *individual* expected survival
  - *cohort* expected survival

Individual survival, population based

Expected survival of a

- 45-year-old US male
- over the next 10 years
- beginning on July 4, 1967

The code below shows that the chance of reaching a 55th birthday is 0.911.

```
> tdata <- data.frame(age= chron('7/4/67') - chron('3/10/22'),
                      sex='male', year= chron('7/4/67'))
> fit <- survexp(~1, data=tdata, ratetable=survexp.us,
                       times=(1:5)*730.5)
> fit

   Time n.risk survival
    730      1    0.987
   1461      1    0.971
   2192      1    0.953
   2922      1    0.933
   3652      1    0.911
```

Rate Tables

S-Plus has several built in rate tables:

- United States, 1940–2000, male and female
- United States, 1940–2000, male and female, by race
- Minnesota, Florida, and Arizona
- West North Central region of the US

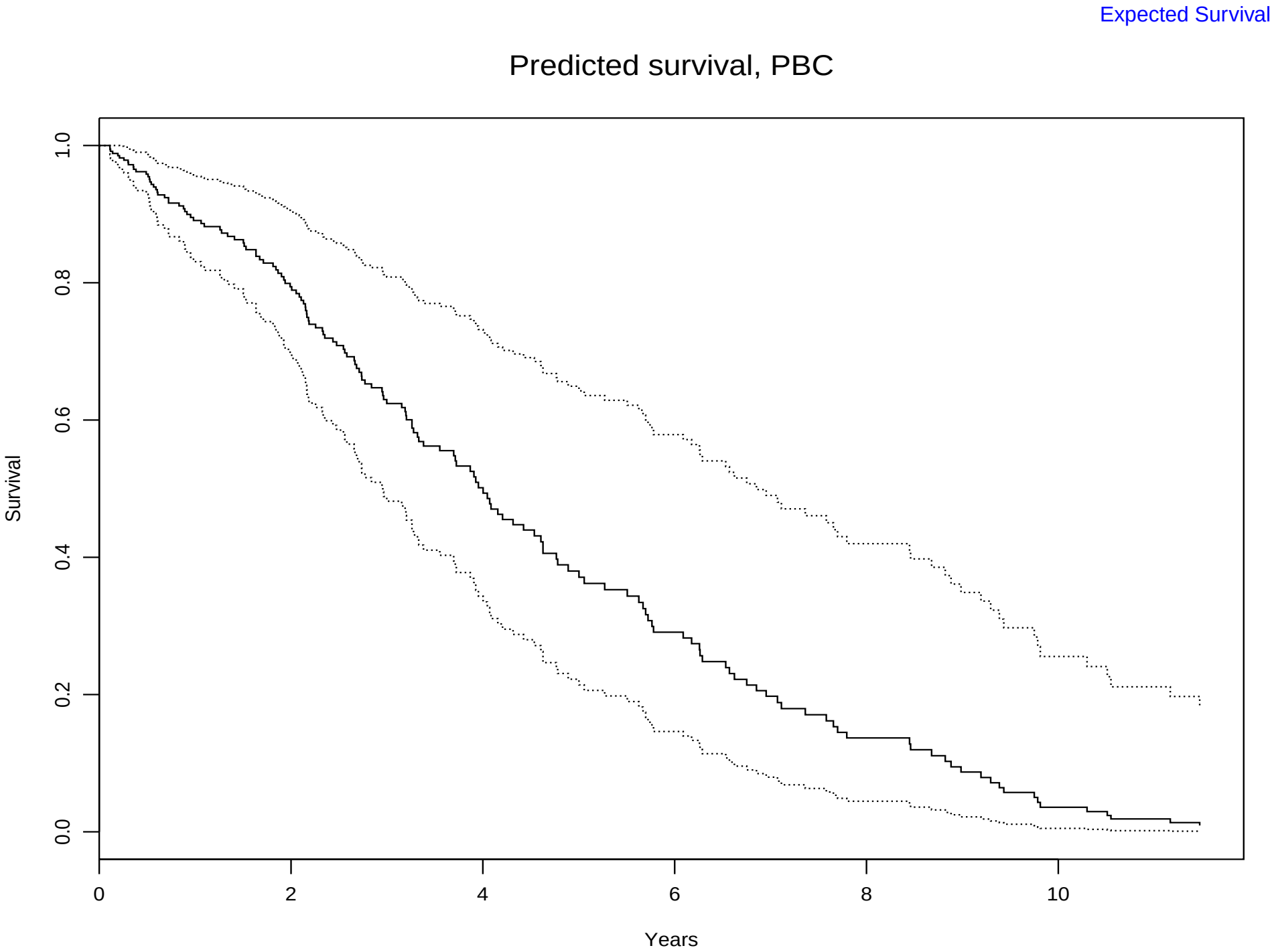
It is quite easy to build your own, for example cancer incidence rates by age, sex, calendar year and tumor type.

See the report on the Mayo Biostatistics web site.

Individual survival, Cox model

- A prior Cox model now acts as the “rate table”
- Covariates in the Cox model take the place of age, sex, etc.
- Example: Natural history of PBC
  - patient-specific survival curves are an important output
  - that is, what is the expected future or “natural history” survival for this patient
  - assume that the patient in question is 53 years old, has no edema, and has bilirubin, protime, and albumin values of 2, 12, and 2.

```
> # Fit a Cox model to the original data set
> pbcfit <- coxph(Surv(futime, status==2) ~ age + log(bili) +
                        log(protime) + log(albumin) + edema, pbc)

> # Create a data set corresponding to the hypothetical subject
> temp <- data.frame(age=53, edema=0, bili=2, protime=12,
                                              albumin=2)
> # Obtain and plot the expected curve
> sfit <- survfit(pbcfit, newdata=temp)
> plot(sfit, xscale=365.24, xlab="Years",ylab="Expected Survival")
```


Expected Survival
Predicted survival, PBC
Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
Years

If there are multiple observations in the `temp2` data set, then the result will be multiple curves.

If there are multiple strata in the Cox model, this also gives multiple curves.

```
> pfit2 <- coxph(Surv(futime, status==2) ~ age + log(bili) + log(albumin)
               + log(protime) + strata(edema), data=pbc)

> temp2 <- data.frame(age=c(50,65), bili=c(2,2), albumin=c(3.5,3.5),
                      protime=c(10,10))

> curves <- survfit(pfit2, newdata=temp2)
> curves
            n events mean se(mean) median 0.95LCL 0.95UCL
  edema=0 352    115 3120     82.0   3574    3222      NA
  edema=0 352    115 2556     57.4   2583    2286    3244
edema=0.5  44     26 2780    347.3     NA    3282      NA
edema=0.5  44     26 2293    266.6   2071    1616      NA
  edema=1  20     19 2167   1025.9   1434    1434      NA
  edema=1  20     19 1600    623.7   1434    1217      NA

> plot(curves[1:3,1], xscale=365)
```

- Model stratified on `edema`, which has 3 levels. Two observations in the `temp2` data set, one is age 50 the other age 65.

- 6 curves result, a 50 year old with edema=0, 60 year old with edema= 0, 50 year old with edema=0.5, . . .

- In S-Plus the curves are a standard “survival curves” object (print and plot them like any other survival curve).

An nice advantage of the S-Plus form is that the `curve` object is recognized as a survival curve, so all of the usual plot options work. In particular, it is easily superimposed on a Kaplan-Meier.

```
data temp;  set save.pbc;
    lbili = log(bili);
    lalb  = log(albumin);
    lpro  = log(protime);

data temp2;
    age=50; lbili=log(2); lalb=log(3.5); lpro=log(10); output;
    age=65; lbili=log(2); lalb=log(3.5); lpro=log(10); output;

proc phreg data=temp;
    model futime * fustat(0) = age  lbili lalb lpro;
    strata edema;
    baseline out=temp3 covariates=temp2 survival=surv
           u=usurv l=lsurv/nomean;
```

In SAS the final data set has the curves one after the other, sorted by strata.

“Mean” survival

If there is no `newdata` argument in S-Plus, or the `mean` option in SAS, then the predicted curve will be for a subject with ‘average’ covariates.

- With discrete variables, what does the curve mean? Gender=1.6, treatment=.5, edema=.568 do not correspond to any real patient.
- An alternate suggestion is to use the most prevalent group.
- This “mean” curve is NOT the predicted survival for the group of subjects.

The default curve is rarely useful.

Comparison of the default curve to the Kaplan-Meier is not IN ANY SENSE a valid check for the “goodness of fit” of the Cox model (more on this later).

Time-dependent covariates

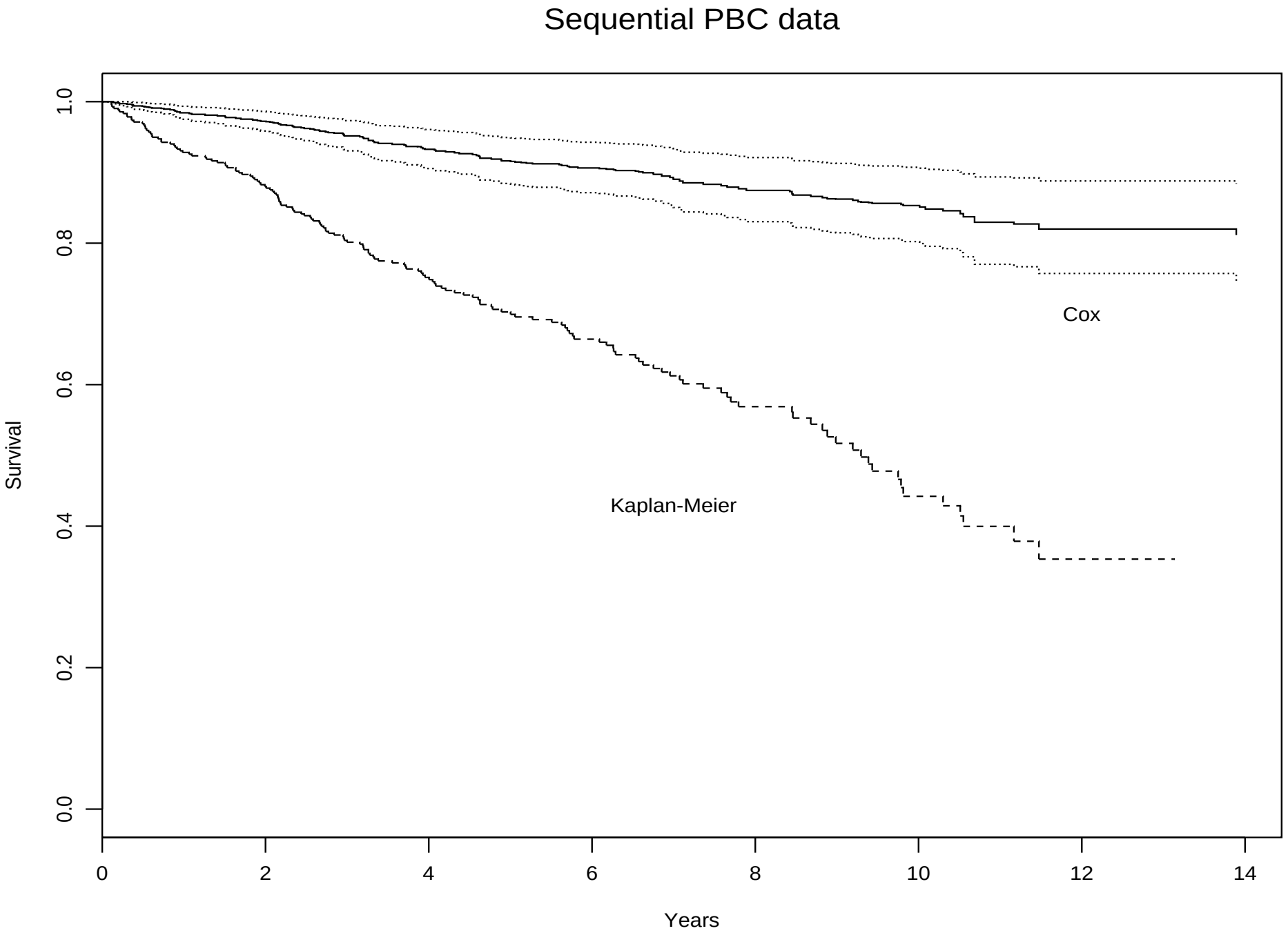
When the model contains time-dependent covariates baseline survival estimates can still be produced, but the results can be quite surprising.

The data set `pbcseq` contains sequential laboratory measurements on the 312 protocol patients of the PBC study. Patients were scheduled to return at 6 months, 12 months, and yearly thereafter; most patients have these visits and many have one or two “extra” in addition. A time dependent covariate model is fit to the data as

```
> pbcsfit <- coxph(Surv(start, stop, event==2) ~ age + log(bili) +
                   log(protime) + log(albumin) + edema,
                   data=pbcseq)
> pbcsfit
                coef exp(coef) se(coef)     z       p
         age  0.046    1.0471  0.00891  5.17 2.4e-07
   log(bili)  1.085    2.9592  0.11112  9.76 0.0e+00
log(protime)  2.848   17.2604  0.63166  4.51 6.5e-06
log(albumin) -3.719    0.0243  0.49528 -7.51 6.0e-14
       edema  0.806    2.2387  0.23270  3.46 5.3e-04

Likelihood ratio test=474  on 5 df, p=0  n= 1945

> rbind(pbcfit$coef, pbcsfit$coef)
```

```
        age log(bili) log(protime) log(albumin) edema
(old) 0.040     0.864        2.387       -2.507 0.896
(new) 0.046     1.085        2.848       -3.719 0.806

> ssurv <- survfit(pbcsfit)
> plot(ssurv, xscale=365.24, xlab='Years', ylab='Survival')
> sfit <- survfit(Surv(futime, status==2)~1, pbc)
> lines(sfit, lty=3, xscale=365.24, mark.time=F)
```

Expected Survival



Why the difference?

- the normal course is for liver function tests to slowly worsen over several years, then progress rapidly as the sclerosis and damage near a critical threshold, followed shortly thereafter by liver failure.

- The model captures this fact, that failure is preceded by large values of bilirubin, edema, and prothrombin time.

- The baseline survival curve corresponds to a fictional patient with who starts with fairly average covariate values *and then never changes* in those covariates.

- The survival for such a subject — if such a person even exists — would be quite good.

The fundamental issue with time dependent covariates is not a computational one but a conceptual one:

If a time-dependent covariate is in the model, then to produce a baseline survival curve one must specify not just baseline values for the hypothetical subject of the curve, but rather an entire *covariate path* for that subject over time.

Creating such a covariate path is difficult; it is all too easy to create baseline hazards that correspond to a subject who is misleading, uninteresting, or impossible.

SAS printout

```
data temp1; set save.pbcseq;
    lbili = log(bili);
    lpro  = log(protime);
    lalb  = log(albumin);

proc phreg data=temp1;
    model (start, stop)futime* status(0,1) = age lbili lpro lalb edema;
    baseline  out=temp3 survival=surv;
```

```
WARNING: Since the counting process style of response was
specified on the MODEL statement, the SURVIVAL=
statistics in the BASELINE statement should be used with
caution.  These statistics are based on the empirical
cumulative hazards function rather than the product-limit
estimates.  They may not be appropriate as estimates
associated with a survivor function.
```

- First statement is true and important.
- Second statement is true and completely irrelevant. (Substitute ”The sky is blue” if you like.) The Kaplan-Meier like computation is somewhat difficult to

set up in this situation; both SAS and S-Plus use the Nelson-Aalen estimate because it's easier.

- Third statement is false (in my opinion).

S-Plus does allow a true time-dependent curve to be created.

- Create a data frame that contains the entire covariate path for a subject. It must contain all of the variables used in the `coxph` call:
  - `(start, stop]` show the intervals (variable names matching the data set of course)
  - The covariates over that interval.
  - The strata that applies over that interval (for a stratified model).
  - `event` is ignored but must be present
- Use the `individual=T` option of survfit, to declare that one curve is desired.
- Caveat Emptor

Individual S_i from a Cox model

- Nelson-Aalen estimator
  - Variance of the estimate worked out in Tsiatis
  - most common
  - based on hazards
  - easily extended to multiple events
- Kalbfleisch-Prentice estimator
  - Kaplan-Meier like
  - Somewhat harder to compute (iteration for tied deaths)
  - the default in SAS
- Fleming-Harrington estimator
  - Variant on the Nelson for tied deaths (identical logic to the Efron approximation)
  - the default in S-Plus

These are produced with the `baseline` and `survfit` statements.

Which method you choose makes little difference except in the tail of the curve (where the std error is huge anyway).

SAS can do the first two, S-Plus all three.

Population Expected Survival

Create an expected survival curve for a hypothetical *group* of subjects.

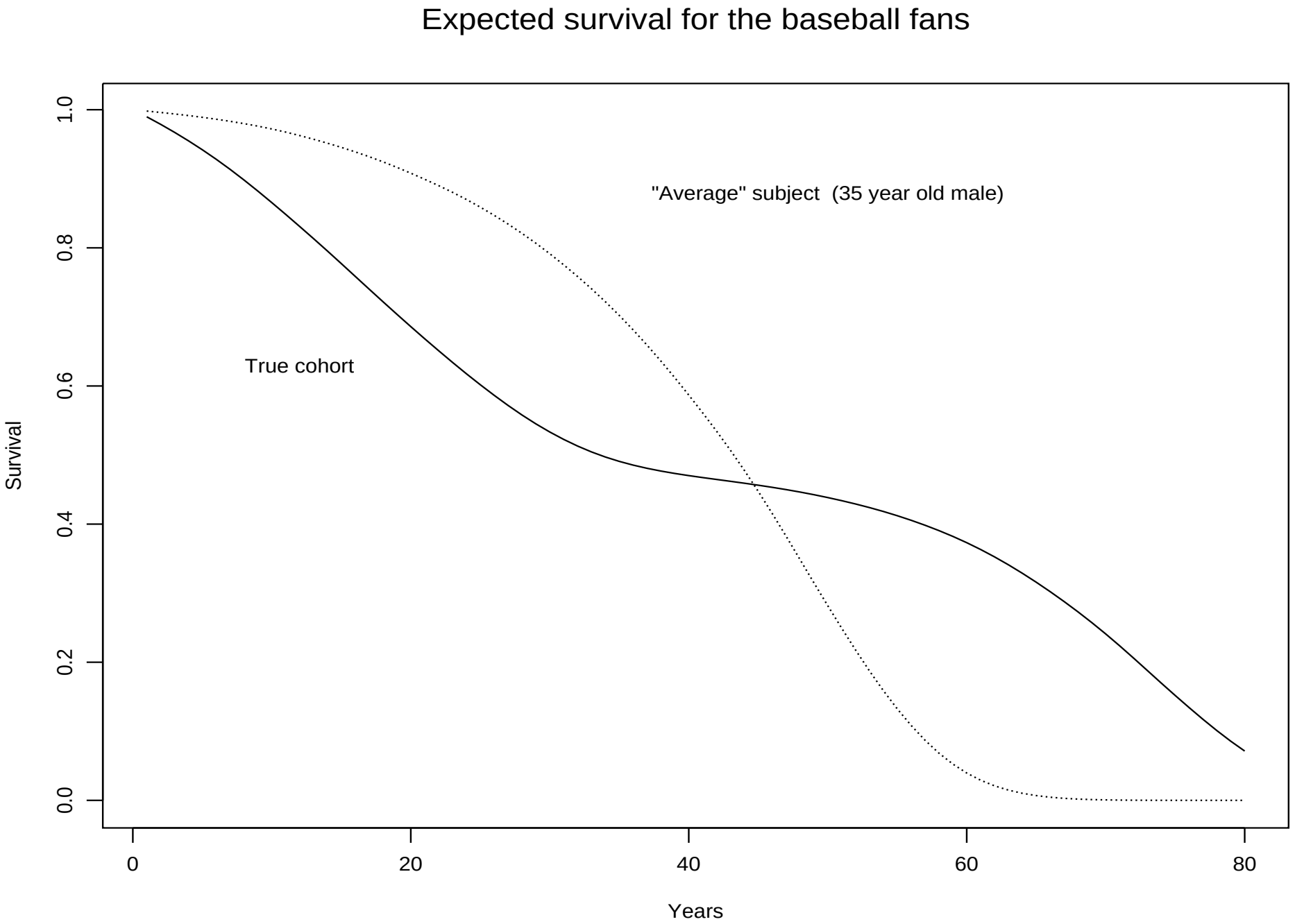
Estimators

- naive
- Ederer (exact)
- Hakulinen (cohort)
- Conditional

Naive estimate: Use the expected survival of the average individual.

What's wrong with the naive estimate?

Imagine a group of 30 grandfathers and 30 grandsons at a baseball game. For some reason we want to know the expected survival of this mixed cohort of 10 and 60 year olds (perhaps for planning reunions). It is *not* the same as that for an average subject, i.e., a 35 year old male.


Expected survival for the baseball fans
"Average" subject (35 year old male)
True cohort
Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
Years

Naive estimate for a Cox model: Compute the mean covariate for the new data set, and use the `baseline` statement to get a single curve.

This estimate was used in Neuberger et al (1986), and justly criticized in Thompsen et al (Stat in Med 1991).

It is no more correct than the baseball example. It may sometimes be close: imagine a cohort all aged between 30 and 35 for instance.

Ederer (or exact) estimate: Compute the expected survival of each individual, separately, and average the curves.

$$S_e(t) = \frac{1}{n} \sum_{i=1}^{n} S_i(t) \, .$$

This is the expected survival curve of a hypothetical matched control group, *assuming complete follow-up*.

F Ederer, LM Axtell and SJ Cutler (1961), The relative survival rate: a statistical methodology, National Cancer Institute Monographs.

In the Cox model case this has been rediscovered and renamed as the “direct adjusted survival curve”.

RW Machuch (1982), Adjusted survival curve estimation using covariates, J Chron Dis.

Liver Transplantation for PBC

Liver transplant is felt to be the only curative procedure available for patients with primary biliary cirrhosis.

Due to a variety of factors, however, including the high cost and risk of the procedure and the limited number of donor organs, this premise has never been subjected to a comparative trial.

When a donor organ becomes available a liver transplant team, either locally to a center or collectively via the procedures of UNOS (United Network for Organ Sharing) must decide which of the multiple needy recipients will receive it. A randomized trial is socially unsaleable in this environment, both now and in the conceivable future.

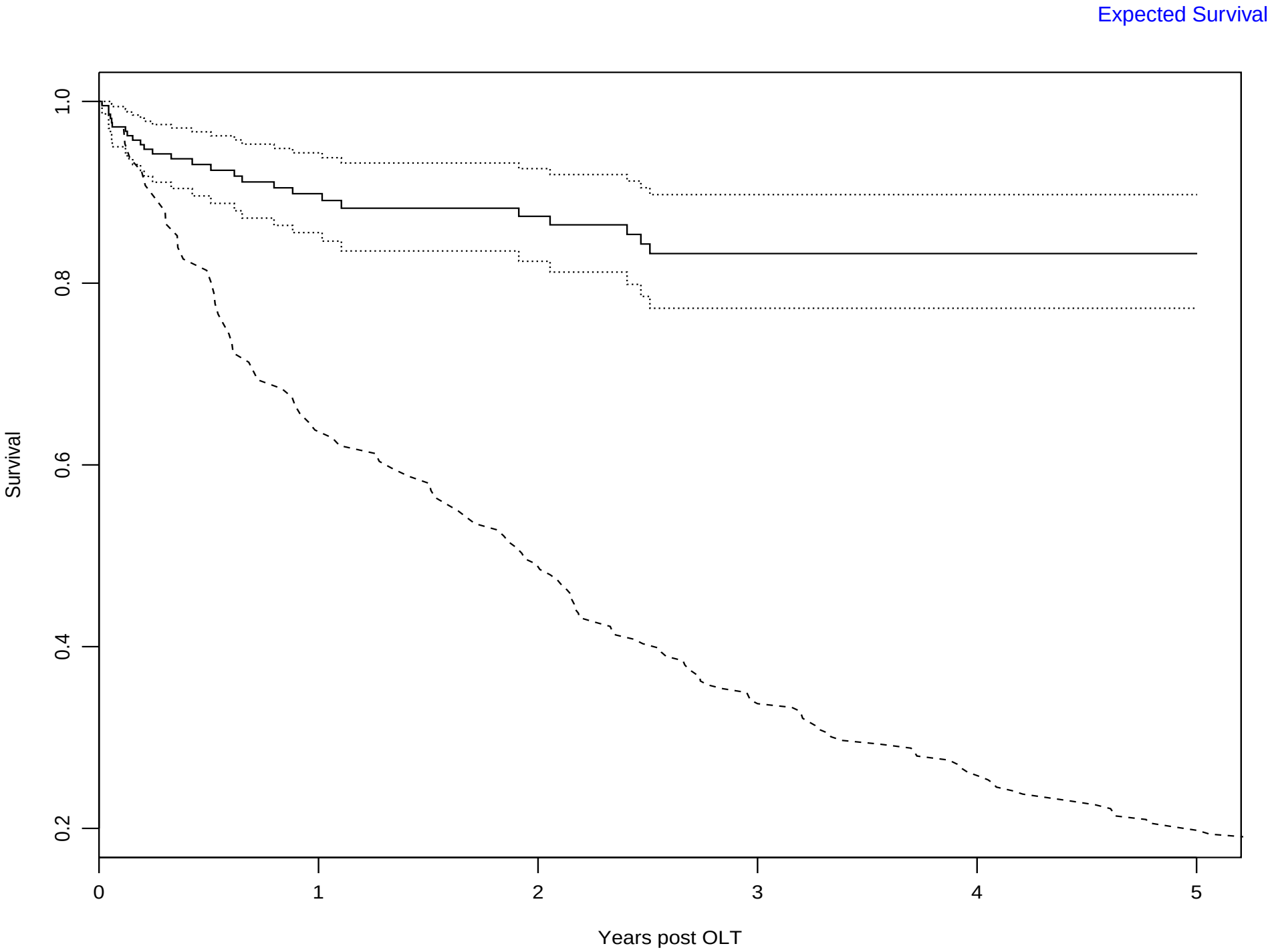
One option is to compare post-transplant survival to what “would have happened” to a matched but untransplanted control.

The data set `olt` contains the survival post orthotopic liver transplant for 214 liver transplants of PBC patients, done at Mayo Clinic, Baylor College of Medicine, and Pittsburgh Medical Center from April of 1985 through September of 1994.

```
> fit1 <- survfit(Surv(futime, death) ~1, olt)
> plot(fit1, mark.time=F, xscale=365)

> pbcfit <- coxph(Surv(futime, status==2) ~ age + edema + log(bili)
        + log(protime) + log(albumin), data=pbc)
> exp1 <- survexp(~ 1, data=olt, ratetable=pbcfit)  #Ederer curve
> lines(exp1, xscale=365, lty=3)
```

Other than status, the variable names in the `olt` data set are identical to those in the `pbc` data set.

The advantage of liver transplant for these subjects is quite impressive. If one looks very closely, there is a small early disadvantage for transplant corresponding to 6 early transplant deaths at 5, 16, 16, 19, 21 and 22 days.


Expected Survival
Survival
1.0
0.8
0.6
0.4
0.2
0
1
2
3
4
5
Years post OLT

SAS computation

```
data temp1; set save.pbc;
    lbili = log(bili);
    lpro  = log(protime);
    lalb  = log(albumin);

data temp2; set save.olt;
    lbili = log(bili);
    lpro  = log(protime);
    lalb  = log(albumin);

proc phreg data=temp1;
    model futime* status(0,1) = age lbili lpro lalb edema;
    baseline covariates=temp2 out=temp3 survival=surv /nomean;

proc sort data=temp3; by futime;
proc means noprint data=temp3;
    by futime;
    var surv;
    output out=ederer mean=surv;

proc gplot data=ederer;
    plot surv * futime;
```

In SAS we first generate all 214 individual curves, then average them using

`proc means`. A plot that overlays the Kaplan-Meier with the expected curve is left to the reader. (In other words, SAS/graph defies my understanding).

One sample test

A formal test of observed=expected for the transplant data can be based on the one-sample logrank.

```
> exp2 <- survexp(futime ~ 1, data=olt,
                      ratetable=pbcfit, cohort=F)
> survdiff(Surv(futime, death) ~ offset(exp2), data=olt)

n=205, 9 observations deleted due to missing.
   Observed Expected    Z   p
       26     88.9    17.1  0
```

The single change in the call to `survexp` is inclusion of the `cohort=F` argument, the function then produces per *individual* results rather than cohort values.

The object `exp2` is simply a vector of 214 values, each of which is the value of the appropriate per-subject survival curve (such as would be computed with `survfit`), at the last follow-up time for that subject. The one sample log-rank test is

$$\frac{(\sum N_i - E_i)^2}{\sum E_i}$$

which is considered to be a chi-square on 1 degree of freedom.

Doing the one-sample test in SAS

Not easy

- The `baseline` statement produces 214 curves, we need 214 values.
  - Value 1 = value of curve 1 at the time subject 1 died or was censored.
  - Value 2 = value of curve 2 at the time subject 2 died or was censored.
  - . . .
- The curves are piecewise horizontal (like a K-M).
- This SAS code is also left to the reader . . .

Other estimates

The “exact” estimate is not always the best, in spite of its name. As with other statistical procedures, it is often the “exact computation of the wrong thing”.

- Ederer or exact: the expected survival curve of a matched control group, assuming complete follow-up of the control.
- Hakulinen: The expected survival curve of a matched control group, assuming censoring in the controls.
- Conditional: The expected survival curve of — not really anything.
  - ⋆ easy to compute
  - ⋆ asymptotically equivalent to Hakulinen if observed = expected.

The Ederer and Hakulinen estimates can differ markedly if patient characteristics change over the course of enrollment.

Verhuel et al (Lancet 1993) examine the survival of 634 consecutive patients over the age of 20 who had an aortic valve replacement at Amsterdam Medical Center, from 1966 to 1986. The proportion of patients over age 70 was

- 1%: first 10 years
- 27%: second 10 years

There are 4 curves

1. The Kaplan-Meier as we see it today, which is systematically too flat in the tail (the last 10 years of the curve contains only young patients).

2. The Kaplan-Meier as it will be.

3. The expected survival of the matched cohort, assuming they have similar censoring to the patients (Hakulinen).

4. The expected survival of the matched cohort, with no censoring in the cohort (Ederer).

Placement of 1 and 4 on the same graph is misleading.

Hakulinen's method (cohort method).

Each of the fictional matched subjects is assumed to have entered the study at the same time as his/her matched case. Each of them thus has a *maximal* follow-up time, and the expected curve is computed with censoring due to follow-up.

For the aortic valve study, the exact estimate of expected was 63% at 15 years, the estimate adjusted for censoring is 75%. The observed survival rate at 15 years was 50% (first year mortality of 11%). Surgery was not as life-saving as it seemed.

This will be an issue with any study that has significant changes in its patient mix during the accrual period.

The conditional method.

Match each case with 10^3 controls. If a control dies, note it, and replace him/her with a new one. If the case dies or is censored, however, stop all follow-up of his/her controls at that point.

The “lab-rat” estimator: if the case dies sacrifice the controls.

Proposed in Ederer and Heise “Instructions to IBM 650 programmers in processing survival computations” (NCI report).

Cohort Survival for the Cox model

- The Hakulinen method has been rediscovered and called the “Bonsel” estimate.

- The conditions for which Hakulinen’s estimate is imperative seem to rarely arise in Cox model situations,

  - Long recruitment time
  - Censoring at an analysis date such that different subjects have very different amounts of contributed time
  - Large changes in enrollment over time, with respect to the important covariates in the Cox model.

- The conditional estimate has been called the “direct survival curve”.

S-Plus can calculate both curves, SAS neither.

Assume that `futime` is the follow-up time for the subjects, and that `ctime` is the constructed censoring time variable

- For censored subjects `ctime = futime`
- For subjects who have died, `ctime` = the time at which they would have been censored, usually analysis date - enrollment date.

```
> exp.hakulinen  <- survexp(ctime ~ 1, data=olt, ratetable=pbcfit)
> exp.conditional<- survexp(futime ~1, data=olt, ratetable=pbcfit,
                        conditional=T)
```

Thomsen et al (1991) point out the difference between the naive and Ederer estimates, and explain the difference in another way. They also discuss the conditional estimator, but say that it is “difficult to interpret”.

The sometimes practice of comparing S_0 or S_e to the Kaplan-Meier as a goodness of fit test is DUMB. In fact: Cox model + K-P estimates of individual curves + conditional = Kaplan-Meier (sharp pencil + grad student + algebra).

The variance of S_0 is found in Tsiatis, and is available from both SAS and S-Plus. The variance of S_e is found in Gail and Byar; it is a mess to compute. The variance of S_h and S_c?

Summary: Cox models

1. The naive estimate is not a good estimate of cohort survival.

2. The Ederer (direct adjusted) and Hakulinen (Bonsel) estimates have not measurably differed in published comparisons. (But I have a recent example where they do.)

3. The method used for the individual curves S_i has little impact. (For the Gronigen curves the last point on the curve is .0931 for PL vs .0946 for CH).

4. Variance estimates are available, but not in software yet.

	SAS	S-Plus
Population		survexp
Individual		$\surd$
Ederer		$\surd$
Hakulinen		$\surd$
Conditional		$\surd$
Cox model	phreg	coxph
individual	baseline	survfit
direct adjusted	*	survexp
Bonsel		survexp
conditional		survexp
Person years		$\surd$

* SAS has no software for cohort estimates, but since each of the S_i can be generated the Ederer estimate is computable with a macro.

Hazard formula, Cox model

Let $\hat{r}_i = \exp(X_i\hat{\beta})$ be the risk score for each subject.

Under a Cox model, the Nelson-Aalen estimate for a hypothetical subject with covariates $X^\dagger$ and risk $r^\dagger$ is

$$\begin{aligned}\hat{\Lambda}(t) &= r^\dagger \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n Y_j(s)\hat{r}_j(s)} \\ &= r^\dagger \int_0^t \frac{d\overline{N}(s)}{\sum_{j=1}^n Y_j(s)\hat{r}_j(s)}\end{aligned}$$

where $Y_i(t)$ is 1 if subject i is at risk at time t.

If $\beta = 0$ this reduces to the usual Nelson estimate.

Most computer programs subtract the means from each column of X before doing this computation, to avoid overflow in the exponential function.

$$e^{x\beta} = e^{(x-60)\beta} e^{60\beta}$$

Variance

The variance of $\hat{\Lambda}$ consists of two terms

- Term 1 = the variance of the N-A estimator, assuming that β was known
- Term 2 = the extra variance due to uncertainty in $\hat{\beta}$

$$T_1(t) = \left(r^\dagger\right)^2 \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\left(\sum_{j=1}^{n} Y_j(s)\hat{r}_j(s)\right)^2}$$

If $\beta = 0$, this is exactly the variance of the Nelson-Aalen estimator as presented earlier.

Again, subtracting the means from each column of X makes the computer program more stabile.

$$
\begin{aligned}
T_2 &= q'(t)\mathcal{I}^{-1}q(t) \\
q(t) &= r^\dagger \int_0^t [X^\dagger - \bar{X}(s)] \frac{d\overline{N}(s)}{\sum_{j=1}^n Y_j(s)\hat{r}_j(s)}
\end{aligned}
$$

- The second term is an average distance from the covariates $X^\dagger$ of the hypothetical subject and the average X of the data set.
- This is similar to confidence limits for a regression line:
  - ⋆ constant term
  - ⋆ + quadratic term that grows as one gets further from the mean
- If fact, if one were to draw the variance at a particular time t as a function of $X^\dagger$, it would look much the same as a regression CI.

Tied death times

When there are tied event times, an important variation of the estimate is the analogue of the Fleming–Harrington estimate. For example, if observations 1 to 5 were at risk at time t with observations 1 to 2 experiencing an event, then the term for that time point

$$\frac{1}{r_1 + r_2 + r_3 + r_4 + r_5} + \frac{1}{r_1 + r_2 + r_3 + r_4 + r_5}$$

would be replaced by

$$\frac{1}{r_1 + r_2 + r_3 + r_4 + r_5} + \frac{1}{.5r_1 + .5r_2 + r_3 + r_4 + r_5}.$$

This is identical to the computations for $\hat{\beta}$ that correspond to the Efron approximation.

Kalbfleisch-Prentice estimate

$$S(t, \hat{\beta}) = \prod_{t_{(k)} \leq t} \alpha_k,$$

where $t_{(k)}$ are the unique event times, and α_k satisfies the equation

$$\sum_{i=1}^{n} dN_i(t_{(k)}) \frac{r_i(t_{(k)})}{1 - \alpha_k^{r_i(t_{(k)})}} = \sum_i r_i(t_{(k)}),$$

where $dN_i(t_{(k)})$ is 1 if subject i has an event at time $t_{(k)}$.

- The left term is thus a sum over the subjects with an event at that time, and
- the right is a sum over all subjects at risk.

If there is only one event at time $t_{(k)}$ then the equation can be solved for α_k,

$$\alpha_k = \left(1 - \frac{r_{(k)}}{\sum_i r_i(t_{(k)})}\right)^{1/r_{(k)}},$$

with $r_{(k)}$ the risk score of the subject who experienced the event.

If there are multiple events, solve by iteration. (It is known that $0 < \alpha < 1$).

Variance: Use the variance of the Nelson-Aalen

A Greenwood style variance can be created, but has not been explored.

Variance for the cohort estimates

- A closed form estimator for the direct adjusted estimate (Cox + Ederer) has been worked out by Gail et al.
  - ⋆ It has not, however, appeared in any packages
  - ⋆ Because the n individual curves are all correlated, due to a common $\hat{\beta}$ and baseline hazard Λ_0, the estimator is complex
- Estimates for the conditional estimate have been worked out by Keiding
  - ⋆ Based on martingale arguments
  - ⋆ However, the conditional estimate is hard to interpret

Bootstrap based estimate

- A large number of times B (100 – 1000)

  1. Draw a sample of size n, from the original data set of size n, with replacement. (time, status, X)
  2. Compute the fitted model for the data
  3. For each subject in the reference data, compute an expected curve
  4. Compute the average, and save the result

- Tabulate the results

  - ⋆ as shown previously for the KM, it is best to summarize on a $\log(S)$ or logit(S) scale
  - ⋆ pointwise standard errors (requires interpolation)
  - ⋆ pointwise bias (best ignored)

Start-stop data

- Many different types of things can be set up by creating a (start, stop] data set.
- For some of them a survival curve makes sense, for some not
- For all, the software will produce a curve

1. Multiple events

- max 1 event per strata (UDCA study)
  - ⋆ Assumption that someone who has had one event type is still informative for other event types
- Multiple events in a strata (Andersen-Gill model)
  - ⋆ last year's lecture: S estimates the time to first bad thing
  - ⋆ this year's lecture: no it doen't
  - ⋆ To estimate time to first event, use a time to first event data set.

2. Alternate time scales
   - Estimates the time to event, for someone who starts at time 0
   - This may or may not be useful

3. Delayed entry (left truncation)
   - For instance, a time scale of “time since diagnosis”, and a delayed entry of “came to Mayo”
   - Curve estimates survival from diagnosis
   - Almost certainly useful
   - Can have a problem with small n at the early time points.

4. Intervals without risk
   - Estimates the survival curve for a subject who was always at risk

5. Your data set
   - Think, think, think

Adjusted survival curves

Suppose that you want to look at survival of males vs females in a study, but

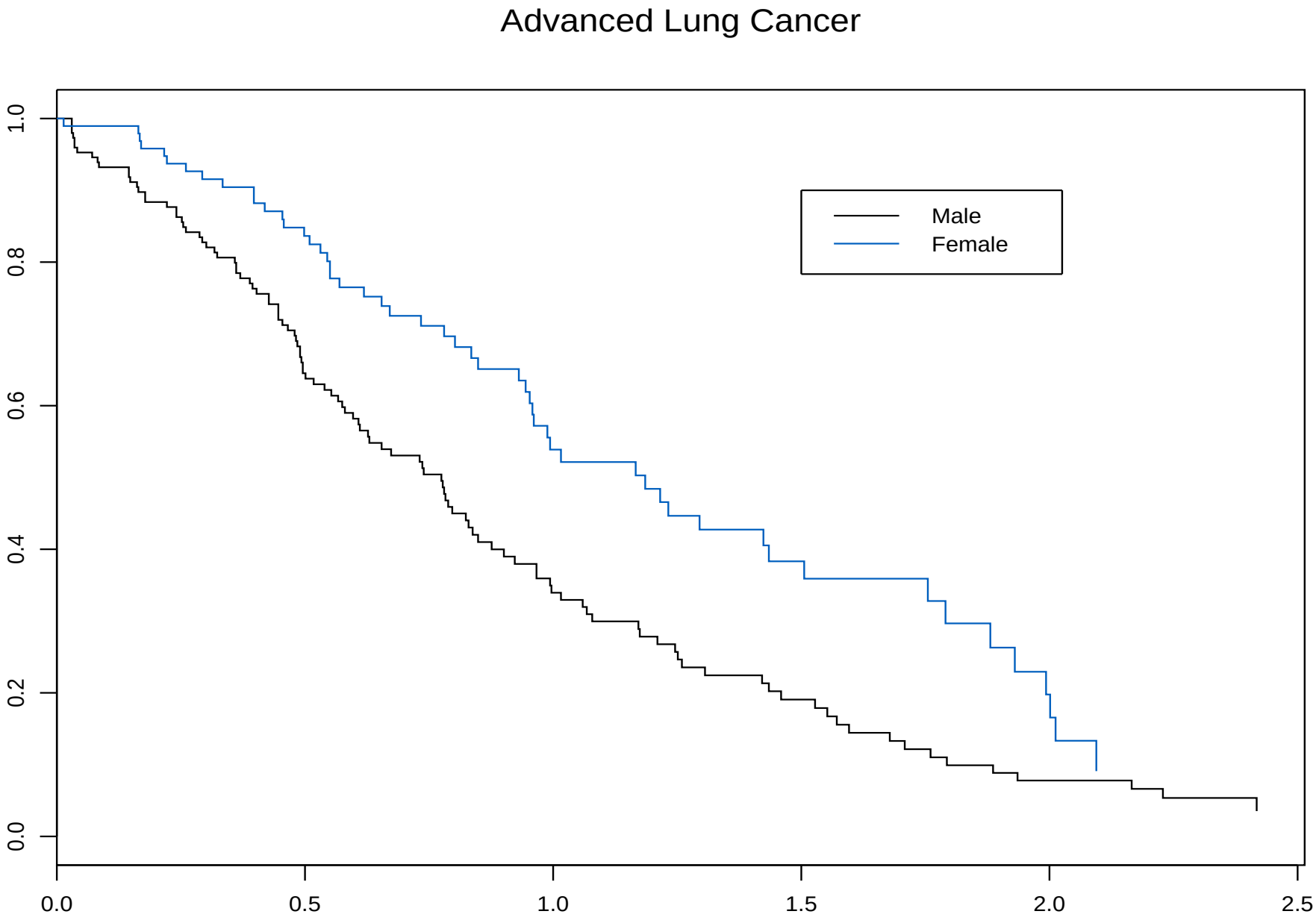
- the two groups differ wrt several covariates $x1, x2, x3, \ldots$
- those covariates have an effect on survival

If

- the covariates can be summarized in a Cox model
- that model will adequately capture the effects
  - ⋆ correct functional form, PH, etc
- the covariate effects can be assumed to be the same for M and F
  - ⋆ (if not, the concept of 'adjusting' has no validity anyway)
- fit a Cox model with sex as the strata, and the important covariates

```
fit <- coxph(Surv(time, status) ~ age + ph.ecog + strata(sex)
                data=lung)
curves <- survfit(fit, newdata=data.frame(age=60, ph.ecog=1))
plot(curves, col=1:2, xscale=365.25)
legend(1.5, .9, c('Male', 'Female'), lty=1, col=1:2)
title('Advanced Lung Cancer')
```

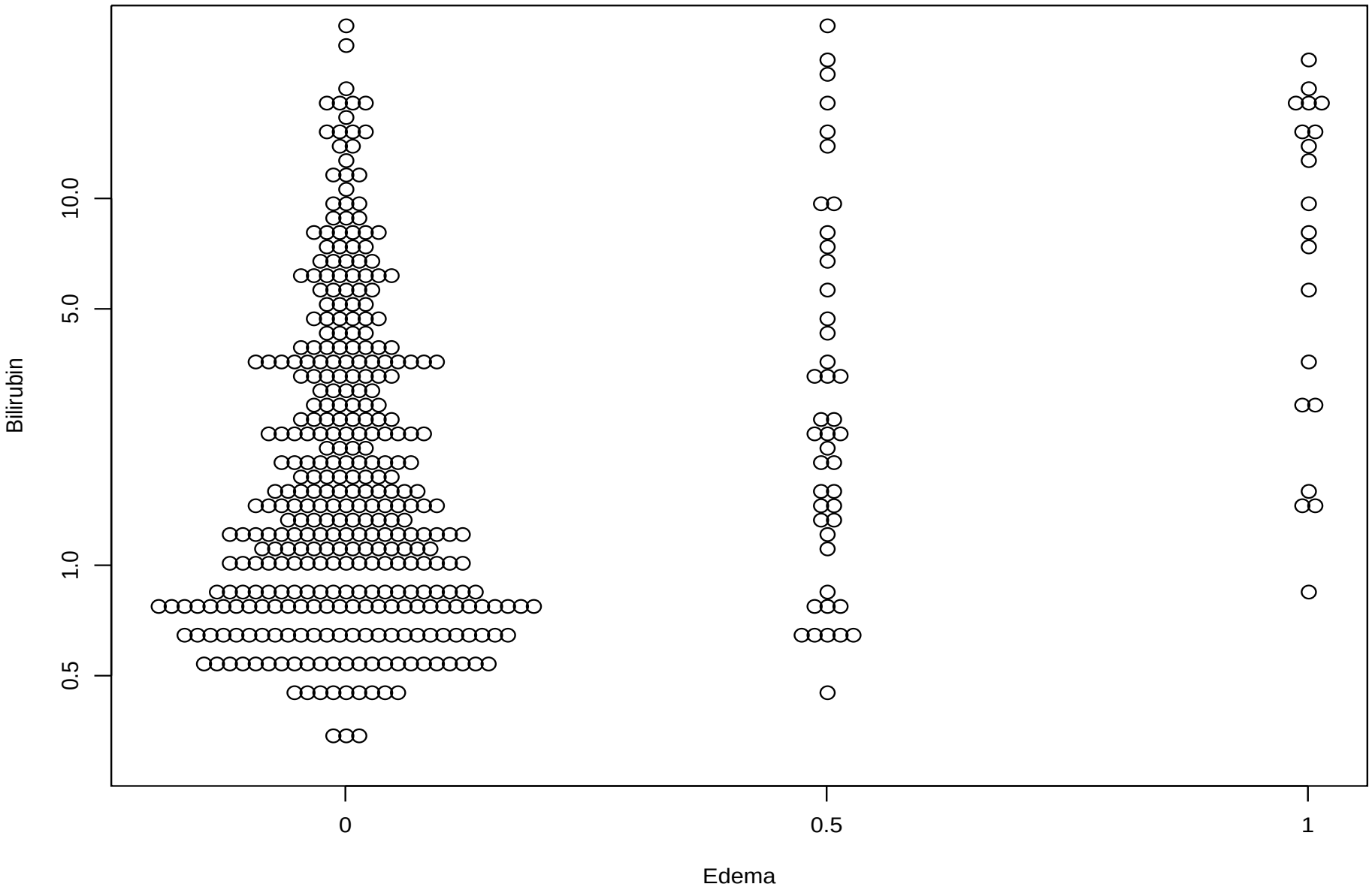
Expected Survival



PBC data

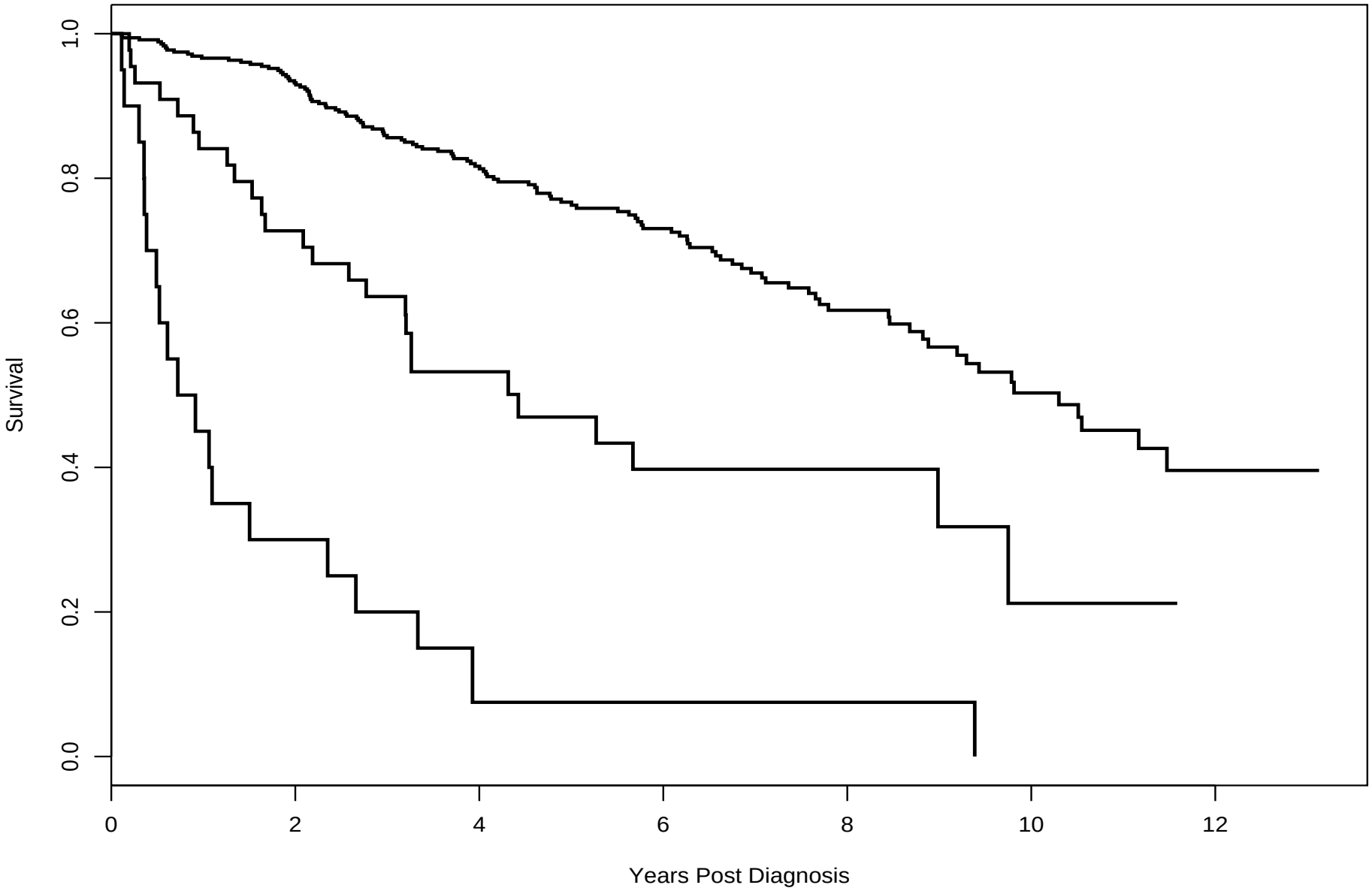
Look at a more realistic example.

- In the primary biliary cirrhosis data (PBC) there is a strong relationship between all of the liver measures
  - bilirubin
  - edema
  - prothrombin time
  - albumin
- Bilrubin is known to be the most important predictor
- Can we get a picture of the effect of edema
  - after adjusting for bilirubin
  - without assuming PH for edema


Bilirubin
10.0
5.0
1.0
0.5
0
0.5
1
Edema

Expected Survival

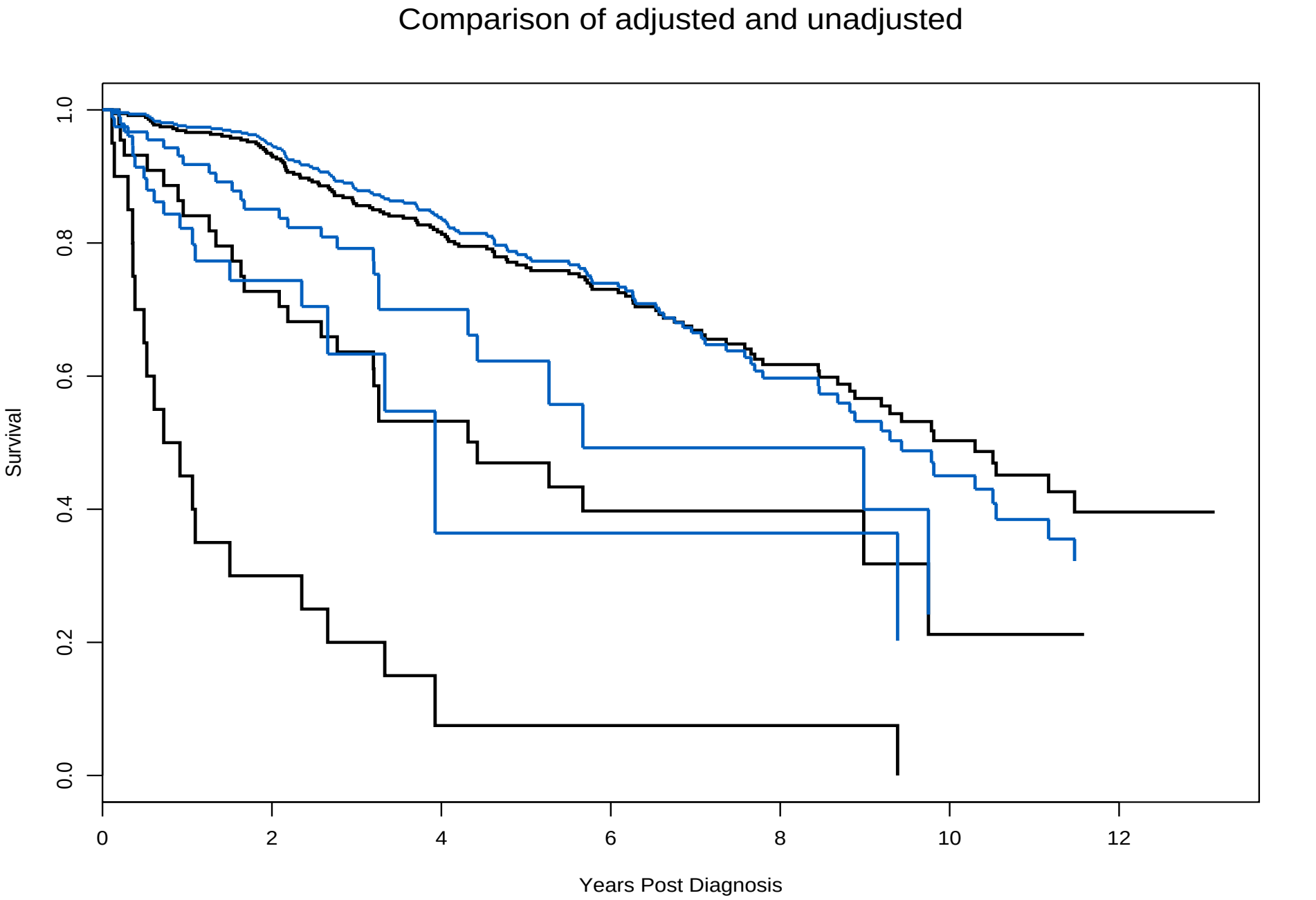
PBC Study, Survival by Edema



Adjusting for the mean bilirubin

```
> sfit1 <- survfit(Surv(futime, status==2) ~ edema, data=pbc)

> fit1 <- coxph(Surv(futime, status==2) ~ log(bili) +
                        strata(edema), data=pbc)

# The geometric mean of bili is 1.77
> sfit2 <- survfit(fit1, newdata= data.frame(bili=1.8))

> plot(sfit1, mark.time=F, xscale=365.25, lwd=2,
      xlab="Years Post Diagnosis", ylab="Survival")
> lines(sfit2, xscale=365.25, col=2, lwd=2)
```

Expected Survival



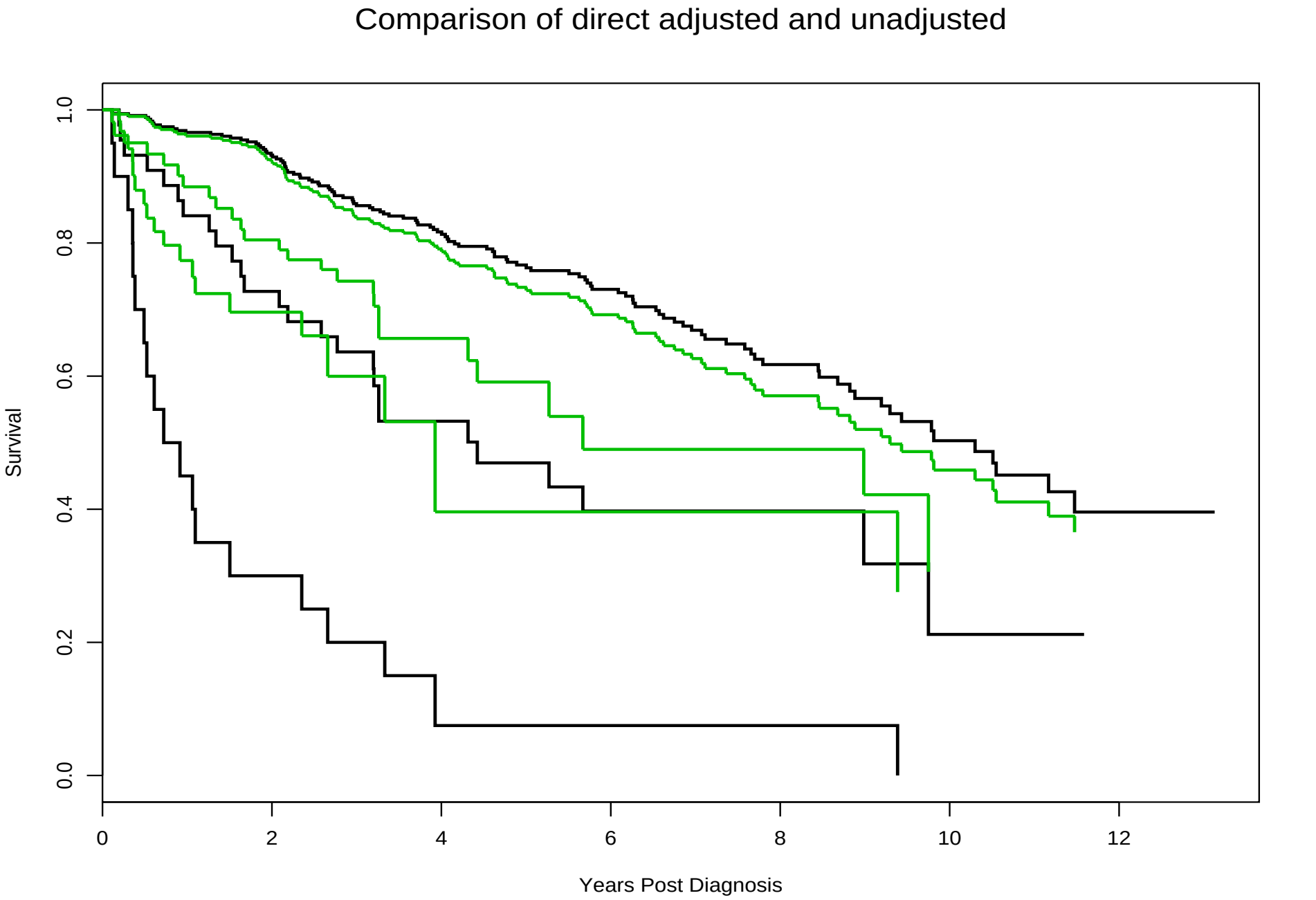
Population adjustment

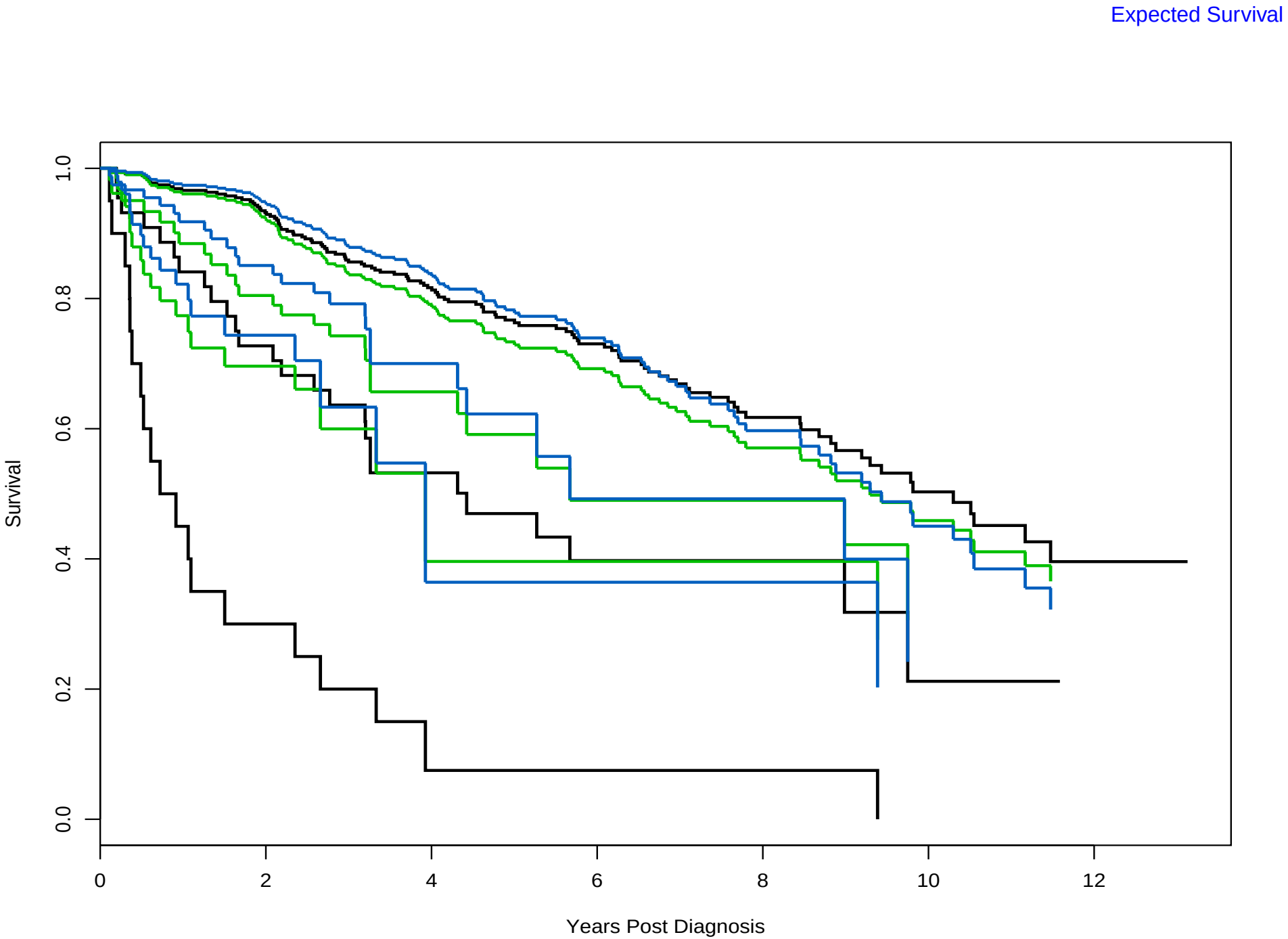
- Because
  - bilirubin is continuous
  - risk is very well modeled by log(bilirubin)
  - the Cox model had the “right” variable in the right form
- Then we can get away with using the mean curve as a display.
  - actually no, not even in this case
  - major mistake in other data sets
- Better is to use a population average

- Population average
  - What the KM “would have been” if all three edema groups had the same distribution of bilirubin values
  - What to use as the ‘same distribution’?
    - the overall distribution of bilirubin
    - the distribution of bilirubin in one particular group

```
> sfit3 <- survexp(~1, data=pbc, ratetable=fit1)
 Problem in survexp: survexp cannot handle stratified Cox moc

> sfit4 <- survfit(fit1, newdata=pbc)
> sfit4$surv <- rowMeans(sfit4$surv)
> plot(sfit1, mark.time=F, xscale=365.25, lwd=2,
      xlab="Years Post Diagnosis", ylab="Survival")
> lines(sfit4, xscale=365.25, col=4, lwd=2)
```

Expected Survival




Expected Survival
Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
12
Years Post Diagnosis

Reprise

- The KM or FH estimator, on raw data, gives the survival for a *group* of people

- When dealing with a rate table, one can compute survival for either a *group* or for a *person*
  - ⋆ survival for a ‘person’ is of course the hypothetical survival for a large number of identical persons
  - ⋆ (the actual survival for one person is a step function that goes from 1 to 0 on the day of their demise)

Standard Populations

- Individual survival
  - What is the survival of a 53 year old male, still alive on Oct 30, 2006?
  - What is the expected number of years remaining?
  - What is the probability of 12 more years?
  - Use the US rate tables (or Minnesota, or . . .)
  - In S/R: survexp routine with only 1 subject in the data set

- Cohort survival
  - Given cohort of subjects with a given age/sex makeup, what is the expected survival curve for the group as a whole?
  - The appropriate thing to compare to a Kaplan-Meier
    - corresponsds to a hypothetical study where one had actually recruited an age/sex matched control for every study subject.
    - In S/R, use survexp with cohort=T and multiple subjects
      - it calculates all n individual survival curves
      - averages them appropriately (Ederer, Hakulinen, conditional)
  - Why are age, sex, and race always the covariates?

- Those are the variables in the standard rate tables.
- One would match on even more, if you could.
- Therneau et al have an example that matches on years of smoking.

Cox models

Issues are exactly the same as before

- Individual survival
  - What is the predicted survival for a 53 year old PBC patient with bilirubin of 2, no edema, prothrombin of .5, and albumin of 3.5?
  - In S/R use the survfit routine, with a Cox model fit as the first argument.
  - In SAS, use the baseline statement
  - Often applied to the immediate subjects of a study
  - Less often, a new patient and an older definitive study.
- Cohort survival
  - Using predictions from a prior Cox model, what is the expected survival of a given cohort of subjects?
  - Sometimes used for ‘what if’ questions.
  - Sometimes used to adjust a survival curve.
  - Sometimes used to display the results from a study.
- What variables?

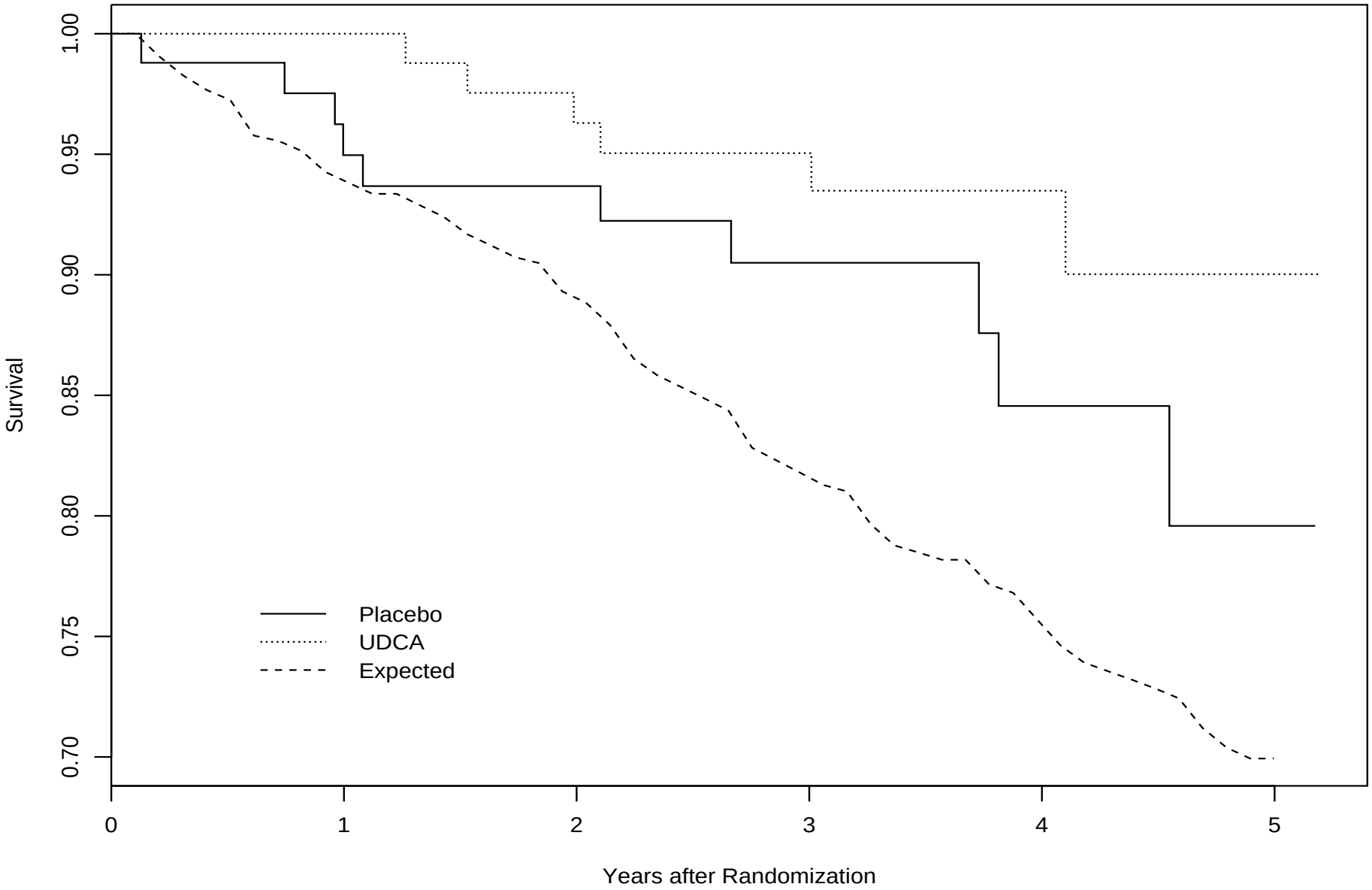
- whatever was important in the Cox model

Expected survival in UDCA treated patients

(Section 10.4.5)

- D-penicillamine, the agent used in the trial for the PBC data set, was shown not to be effective; several other drugs have since been evaluated in this disease.
- A randomized double-blind trial of a new agent, ursodeoxycholic acid (UDCA), was conducted at the Mayo Clinic from 1988 to 1992 and enrolled 180 patients.
- The data are reported in Lindor et al. 1994, UDCA nearly halves the rate of adverse outcomes.

Expected Survival



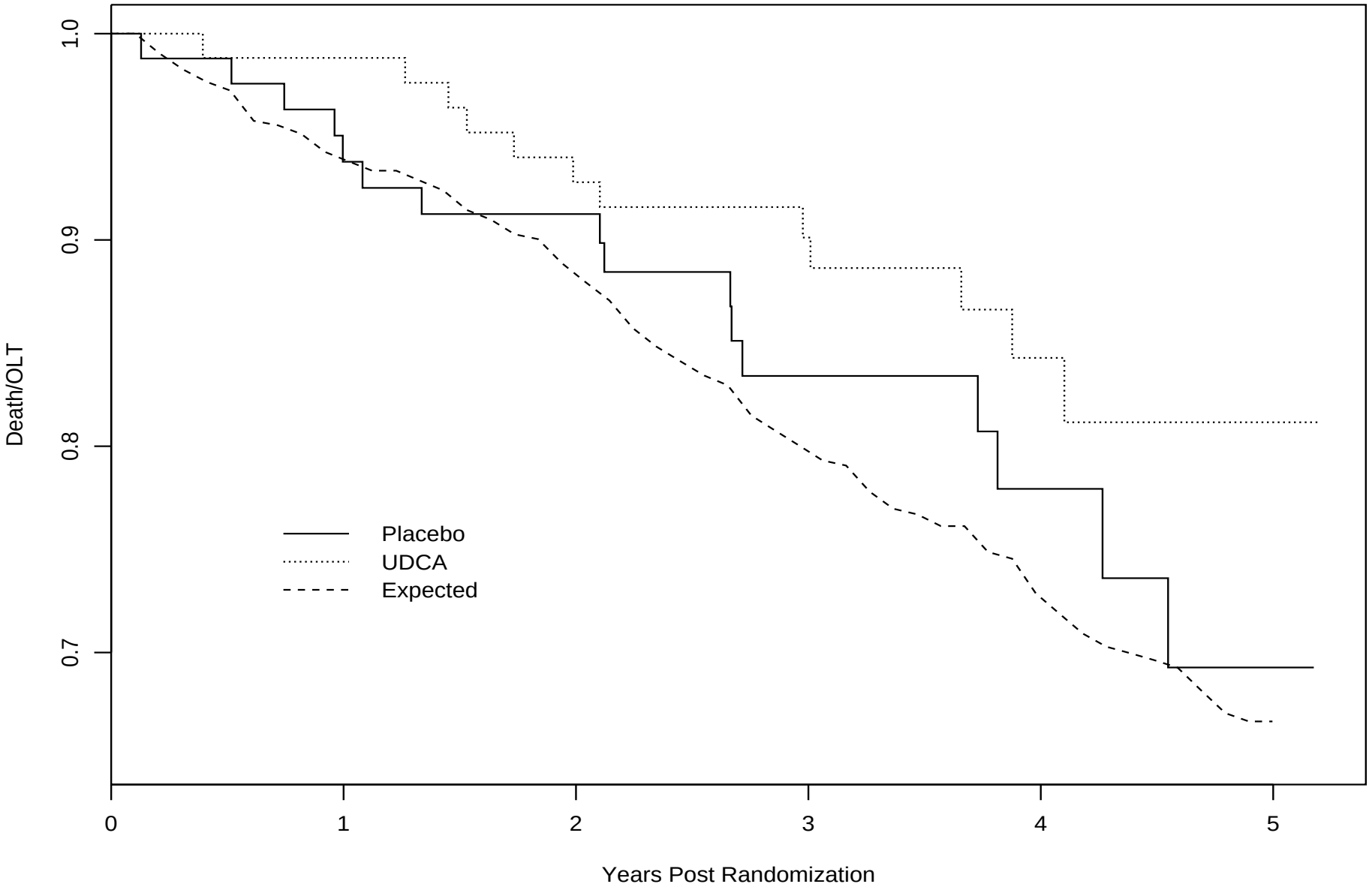
- The plot shows the UCDA and placebo KM curves
  - (Note that in editions 1 and 2 of the book, the UDCA/placebo labels are switched).
- Along with an expected curve based on
  - The PBC natural history model
  - All 180 subjects in the trial
- Why does the placebo group have better survival than expected?
  - Suspect: the rise in orthotopic liver transplant (OLT)
  - At the time of the PBC natural history trial, transplant was new
  - At the time of the UDCA trial, transplant was a recognized therapy
    - the sickest patients are selected for transplant
    - the treatment ‘rescues’ those who are about to die

Combined death/old analysis

- One solution would be to use a death/olt endpoint
- Refit the pbc data, using death/olt
  - `coxph(Surv(futime, status>0)  age + ...`
- However....
  - We would like to retain the coefficients of the PBC natural history model
  - The model has been validated in 3 other populations, with outstanding stability
  - Change only the baseline hazard Λ_0
  - Similar to changing only an intercept
- Key trick: iter=0

```
> pbcfit <- coxph(Surv(futime, status==2) ~ age + log(bili) +
                edema + log(albumin) + log(protime), data=pbc)

> risk <- c(pbcfit$x %*% pbcfit$coef)

> fit2 <- coxph(Surv(futime, status>0) ~ age + log(bili) +
                edema + log(albumin) + log(protime), data=pbc,
      init= pbcfit$coef, iter=0)
> expect2 <- survexp( ~ 1, data=udca3,
                                      ratetable=fit2)
# Book is wrong here too
```

Expected Survival



Alternate code

- Create a fixed pbc risk score
  - rscore = .0396*age + .8635*log(bili) + . . .
- Add `rscore` to the model as an *offset* variable
- Works in either SAS or S/R

Utility of the risk score

Question:

- UDCA improves survival
- UDCA improves liver chemistry values (better bilirubin, etc)
- The PBC natural history risk score is based on liver chemistries
- Is the PBC risk score useful, in a patient on UDCA?
  - ⋆ Does it correctly rank patients
  - ⋆ Does it predict survival accurately?

Approach: Obtain the observed and expected event counts, by subsets of risk score.

- If the risk score is worthwhile
  - ⋆ Observed and expected should rise together
- If the risk score is worthwhile and *calibrated*
  - ⋆ The observed will (approx) equal expected

- Complication
  - Some patients crossed over from placebo to UDCA
  - This occured at the planned end of the study
  - UDCA takes about 6 months to saturate.
  - How to deal with them?
    - Only use time up to crossover
    - Change groups at crossover (risk score at crossover)
    - Use the risk score after 6 months of treatment, along with time after 6 months

```
> tdata <- na.omit(udca3[, c('risk', 'rx', 'status', 'futime')])
> group <- cut(tdata$risk, c(0, 4.4, 5.4, 6.4, 10))
> expect3 <- survexp(futime ~ 1, data=tdata,
                     ratetable=fit2, cohort=F)

> temp1 <- table(group, tdata$rx)
> temp2 <- tapply((tdata$status >0), list(group,tdata$rx), sum)
> temp3 <- tapply(-log(expect3),list(group, tdata$rx), sum)
```

		Placebo			UDCA		After 6 Months UDCA		
Score	n	Obs	Exp	n	Obs	Exp	n	Obs	Exp
≤ 4.4	24	0	1.0	29	2	2.4	52	2	3.0
4.4–5.4	38	5	3.1	26	1	6.4	17	2	3.2
5.4–6.4	20	1	4.3	20	5	10.7	11	6	4.7
> 6.4	9	7	7.1	13	7	21.1	5	3	5.2
Total	91	13	15.4	88	15	40.5	85	13	16

Table 3: Observed and expected numbers of death/OLT events

Competing risks

Monoclonal Gammopathy

The plasma cell lineage comprises only a small portion of human blood cells, $<$3%, but is responsible for the production of immunoglobulins, an important part of the body's immune defense.

Each cell creates a distinctive molecule, allowing the body to respond to a broad spectrum of threats; when the immunoglobulins are assayed using protein electrophoresis one sees a roughly uniform density of molecular weights over the defined range.

In the case of a plasma cell malignancy (multiple myeloma, macroglobulinemia, amyloidosis, and other related disorders) the assay will often reveal a sharp spike — in spite of its inappropriate (unbounded) growth, the malignant clone is still manufacturing its unique product.

The presence of a monoclonal peak in persons without evidence of overt disease is termed “monoclonal gammopathy of undetermined significance” (MGUS); it may often be discovered inadvertently when protein electrophoresis is performed for other diagnostic reasons.

The population prevalence of MGUS increases with age from about 1% at age 50 to 3% for patients over 70.

Given its connection to serious plasma cell diseases, the potential prognostic importance of MGUS is of interest. Is it:

- a precursor state to malignancy
- an incidental finding of no prognostic importance
- something in between?

There are multiple hazards operating:

$$\begin{aligned} \lambda_m(t) &= \text{multiple myeloma} \\ \lambda_o(t) &= \text{other plasma cell malignancies} \\ \lambda_d(t) &= \text{death from other causes} \end{aligned}$$

Estimating the individual hazards, e.g. with a Cox model, is simple.

Estimating the “survival” in the presence of competing risks is harder.

The data set has 1384 subjects. If there were no censoring, one useful thing to examine would be the *crude hazard*:

- $CI_{\text{prog}}(t) =$(# progressions by time t)/1384
- $CI_{\text{death}}(t) =$(# deaths w/o progression by time t)/1384

Notice that if the death rate were to increase, the progression rate would decrease. This estimate, also called the *cumulative incidence* or *marginal incidence* estimator, shows the influence of one cause upon another.

With censoring, it can be computed using one of several algorithms

An alternative estimate is to compute separate Kaplan-Meier estimates

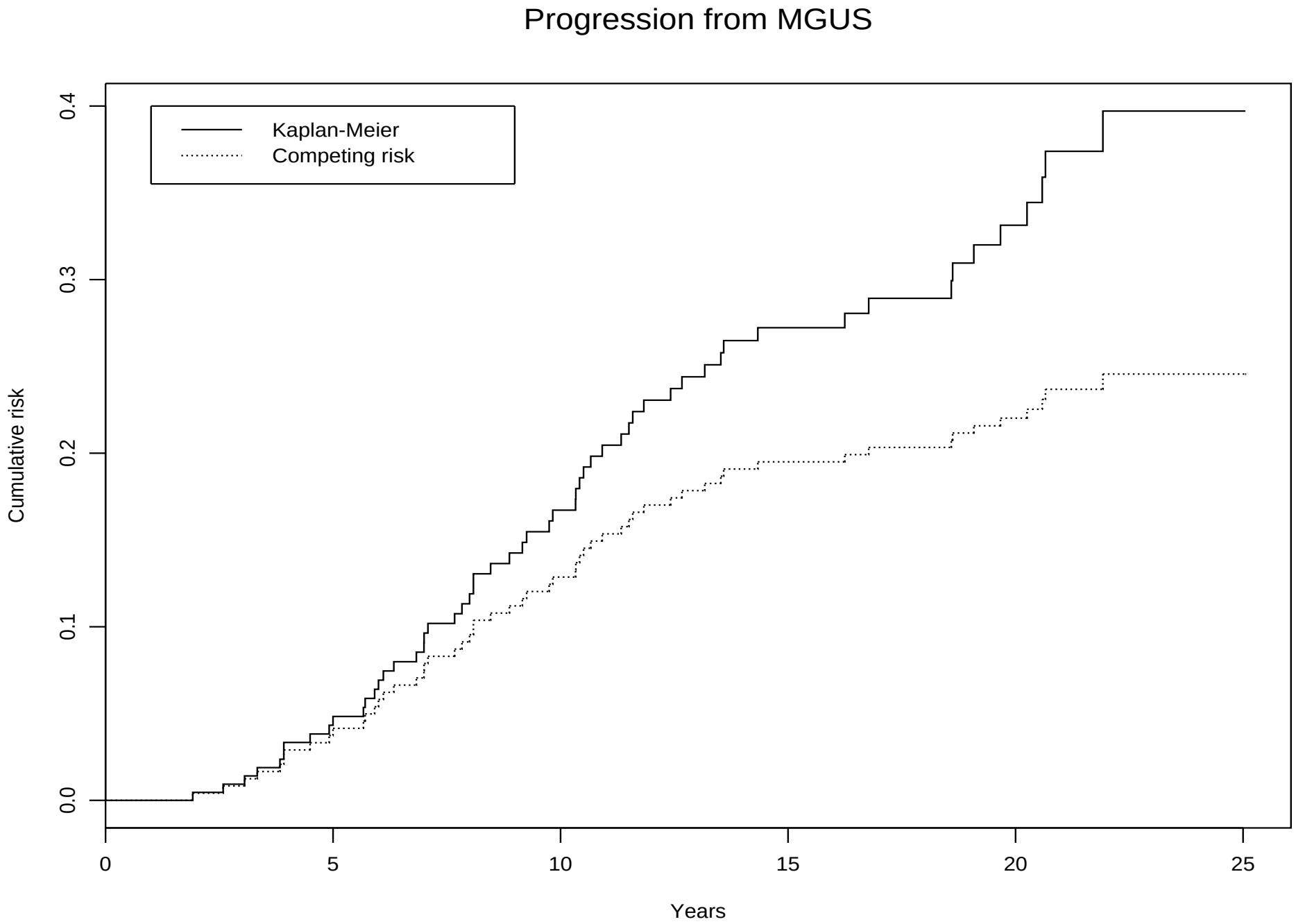
- $K_{\text{prog}}(t) = 1-$ KM(treating death as a censoring event)
- $K_{\text{death}}(t) = 1-$ KM(treating prog as a censoring event)

One is treating each death as “if this subject had not died, their future disease course would be just like those who survived him/her”.

For this to make sense

- The thought experiment must make sense: “progression with all other causes of death eliminated”.
- Death must not be informative for future disease.

Competing risks



Neither estimate is wrong.

The CI curve answers the insurer's question:

- "How many cases of MGUS will actually occur".
- The curve begins to flatten out at older ages, as the other infirmities increase.
- This was the correct curve for sample size estimation.
- The correct way to compute *lifetime risk*.

The KM curve answers the physician/biologist's question:

- "What is a patient's ongoing risk, given that he/she is still coming in to see me?"
- "Does the rate of disease increase with time?"
- But it requires a key assumption about independent censoring.

The KM curve is used much more than it should be, because software is available.

Computing CI estimates

Let $S(t)$ be the overall Kaplan-Meier, "time to first bad thing".

Method 1

1. If S has a jump at time t (some event happened), let $\Delta(t) =$ the jump size at $t = S(t) - S(t-0)$.

2. Partition the jump across the event. If there were 3 events at time t, 2 of type 1, 0 of type 2 and 1 of type 3, then

   - $CI_1(t)$ will increase by $(2/3)\Delta(t)$ at this point
   - $CI_2(t)$ does not increase
   - $CI_3(t)$ will increase by $(1/3)\Delta(t)$ at this point

Method 2: formal equation

$$CI_i(t) = \int_0^t \lambda_i(s) S(s-) ds$$

This is extensible to more complex situations (like the Cox model), but not very intuitive.

Method 3: redistribute-to-the-right

Code (latest R survival)

```
sfit <- survfit(Surv(time, status) ~ sex, data=mgus, etype=event)
```

Averaging Survival

Given a mixed population

- p_1 subjects from population 1
- p_2 subjects from population 2
- ...
- Survival curves S_1t, S_2t ... S_kt

The overall survival curve is

$$S(t) = \sum_{i=1}^{k} p_i S_i(t)$$

Proof

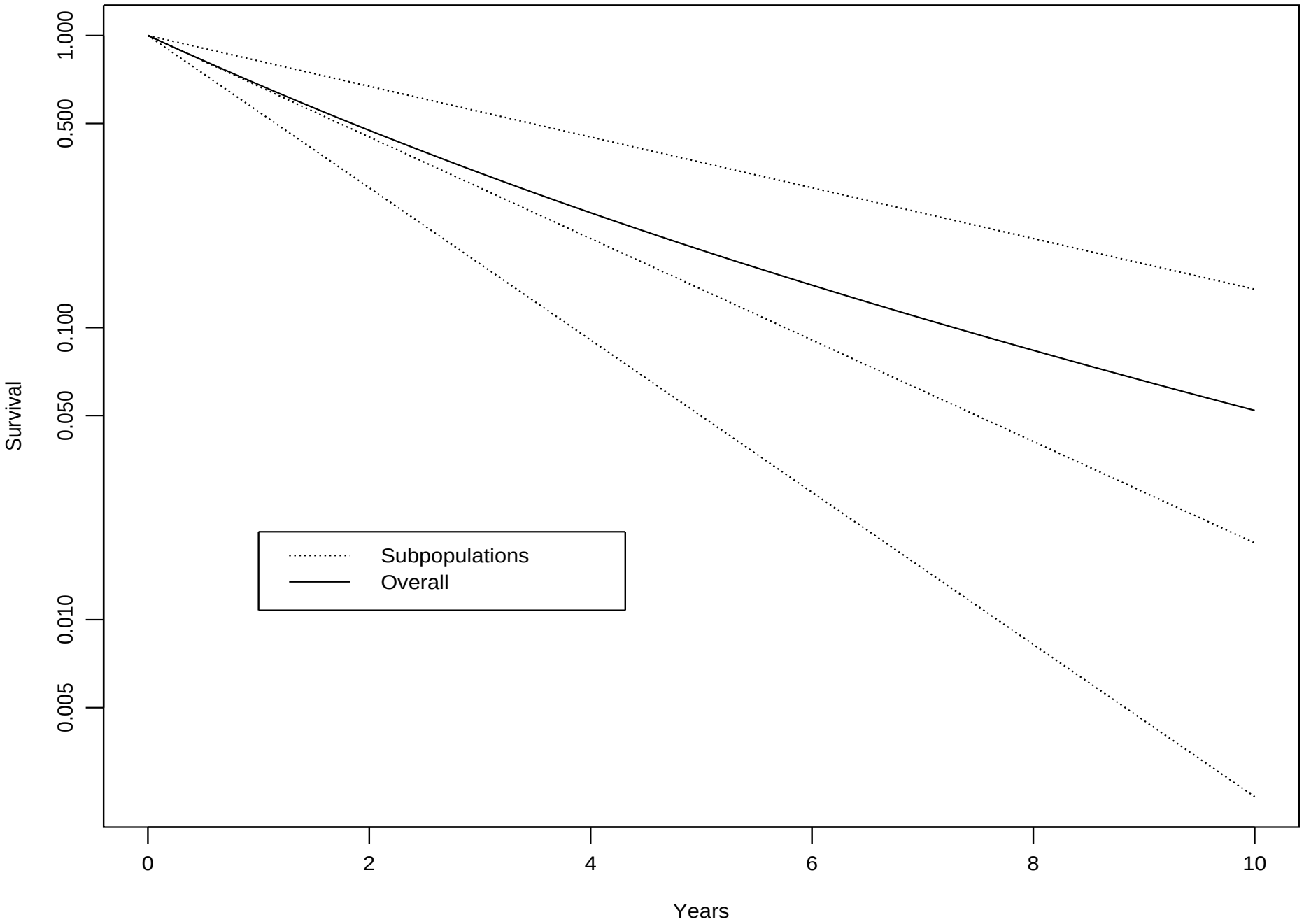
$$
\begin{aligned}
S(t) &= \text{Pr(survival beyond t)} \\
&= Pr(\text{subject from population 1}) \cdot Pr(\text{survival}|\text{pop1}) \\
&\quad + Pr(\text{subject from population 2}) \cdot Pr(\text{survival}|\text{pop2}) \\
&\quad \ldots \\
&= \sum_{i=1}^{k} p_i S_i(t)
\end{aligned}
$$

Average Hazard

$$\begin{aligned}\lambda(t) &= \text{death rate at t, among those still at risk at time t} \\ &= \frac{\sum_{i=1}^{k} \lambda_i S_i(t)}{\sum_{i=1}^{k} S_i(t)} \neq \sum_{i=1}^{k} p_k \lambda_i(t)\end{aligned}$$

- When S is estimated by a Kaplan-Meier
  - ⋆ we want the Pr ("in the drawing")
  - ⋆ the KM drops at a death time
  - ⋆ we need $S(t-\epsilon)$

Competing risks



Cumulative Number of Events

Definition

$$\begin{aligned} \lambda(t) &= \frac{f(t)}{S(t)} \\ CI(t) &= \int_0^t f(s)ds = \int_0^t \lambda(t)S(s)ds \end{aligned}$$

In a competing risks situation

$$\lim_{t \to \infty} CI(t) < 1$$

- Not everyone will experience cause k
- f_k is a ‘defective’ distribution

The redistribute-to-the-right algorithm

One way to approach the competing risks problem is via the representation of the Kaplan-Meier estimate using the redistribute-to-the-right (RTR) algorithm. In this approach, the computation is divided into two distinct s.

- Redistribute the weights

  1. Let each observation have an initial case weight of $w_i = 1$.
  2. Find the smallest censoring time; say it belongs to observation j, and that n_j of the observations have survival or censoring times that are greater than observation j. That is, there are n_j observations at risk at time $t_j + 0$. Redistribute the weight for observation j to these others, incrementing each of them by the amount w_j/n_j, and then set $w_j = 0$. The picture is of someone exiting a partnership, who distributes all of their shares of stock equally to the remaining members.
  3. Proceed forward to the next censoring time, and repeat step 2, until all the censored times have been processed. If the largest observation time in the data is censored, it's weight is given to a fictitious observation placed at $t = \infty$.

- Using these new weights, calculate an ordinary cumulative distribution

function.

Consider the following small data set of AML patients (from Miller), with survival times in months of: 9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, and 161+.

	Step				
Time	0	1	2	3	4
9	1	1	1	1	1
13	1	1	1	1	1
13+	1	0	0	0	0
18	1	1 +1/8	9/8	9/8	9/8
23	1	1 +1/8	9/8	9/8	9/8
28+	1	1 +1/8	0	0	0
31	1	1 +1/8	9/8 + 9/40	27/20	27/20
34	1	1 +1/8	9/8 + 9/40	27/20	27/20
45+	1	1 +1/8	9/8 + 9/40	0	0
48	1	1 +1/8	9/8 + 9/40	27/20 + 27/40	81/40
161+	1	1 +1/8	9/8 + 9/40	27/20 + 27/40	0

The final Kaplan-Meier estimate has a step of size 1/11 at time 9, (1/11)*(27/20)= .1227 at time 34, and so on.

In the competing risks problem:

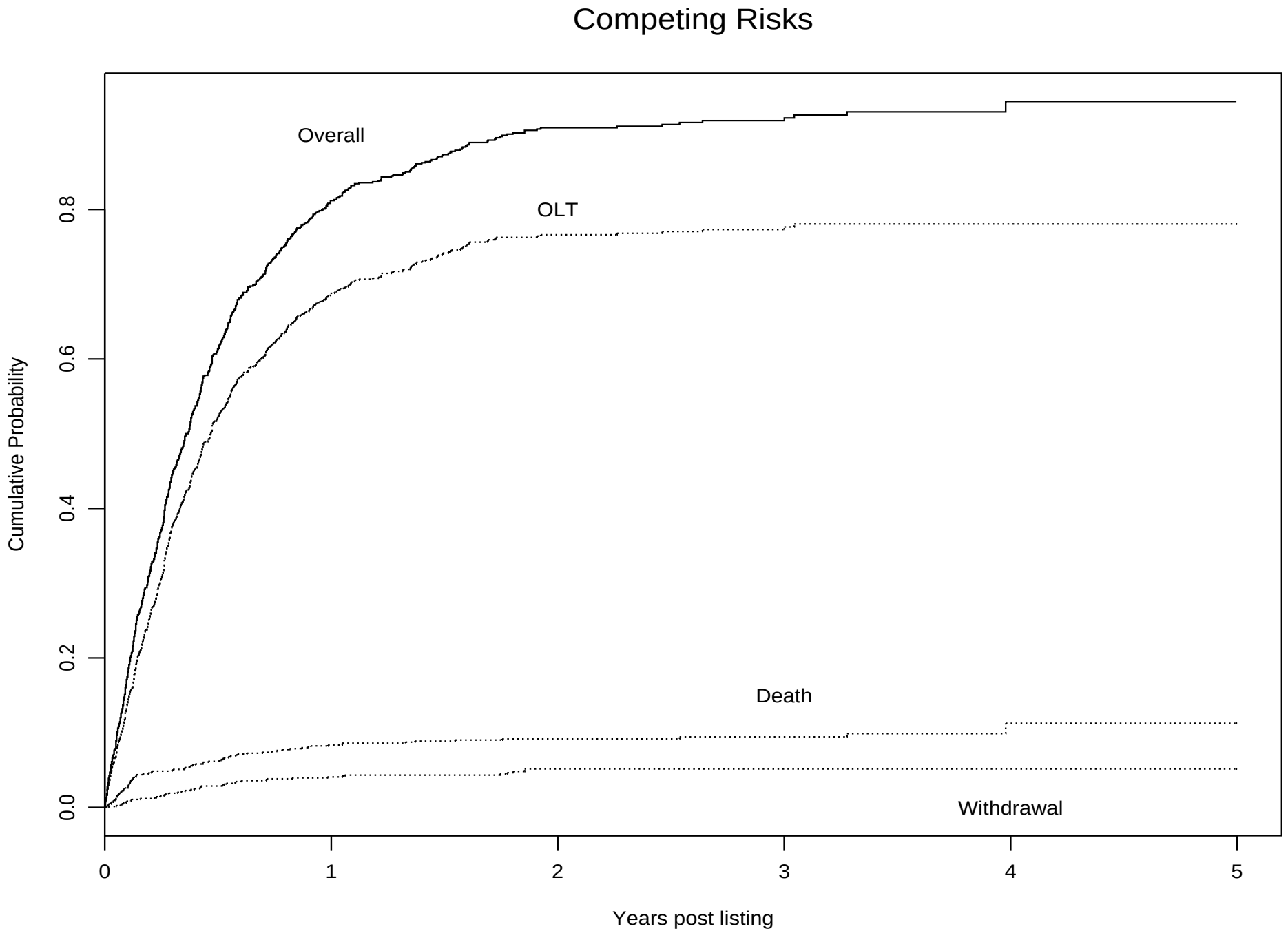
- ordinary KM: redistribute the weights of those who die of other causes. Treat removals due to other causes as though they were censored alive.

- CI estimate: do not redistribute weights of those who fail due to other causes.

Survival on the waiting list

- Feb 1990 through Aug 1999
- 861 patients registered to the waiting list at Mayo Clinic
- At closure
  - 639 OLT
  - 79 died while waiting (12 “withdrawn”)
  - 41 withdrawals from the list
  - 102 remaining on the list

	1990–92	1993–95	1996–97	1998-99
Listed	173	250	215	223
Transplanted	157	204	164	114
Age (mean)	49.3	50.5	49.3	51.5
Male (%)	51	48	60	61
Diagnosis (%)				
Alcoholic	16	12	13	16
Cholestatic	50	38	33	21
Viral hepatitis	14	22	23	34
Other	20	28	31	29
Blood type (%)				
A	46	43	33	36
B	12	13	12	14
AB	5	4	4	8
O	38	40	52	43
Meld (median)	18	15	13	13

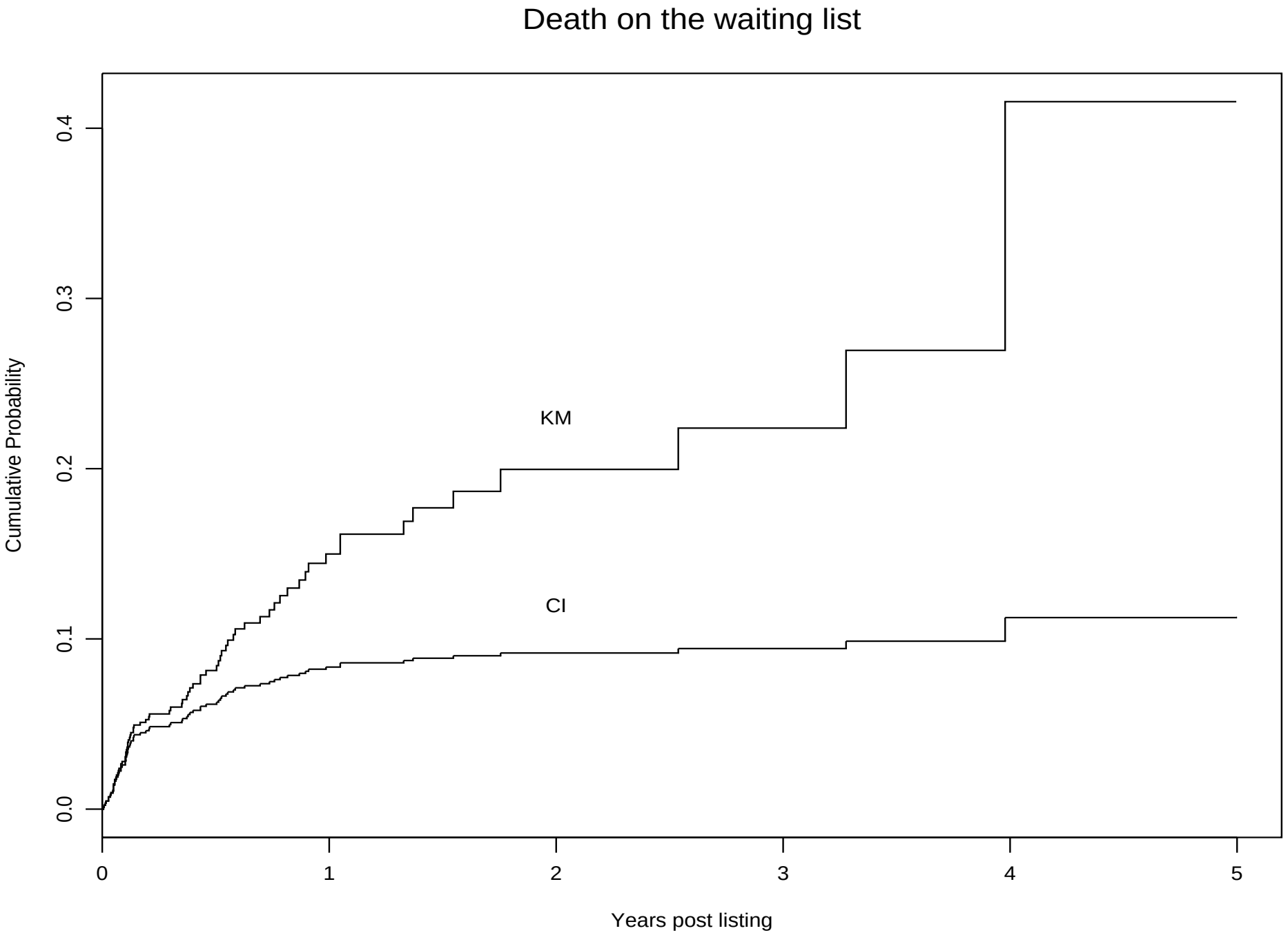

Competing Risks
Overall
OLT
Death
Withdrawal
Cumulative Probability
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
Years post listing

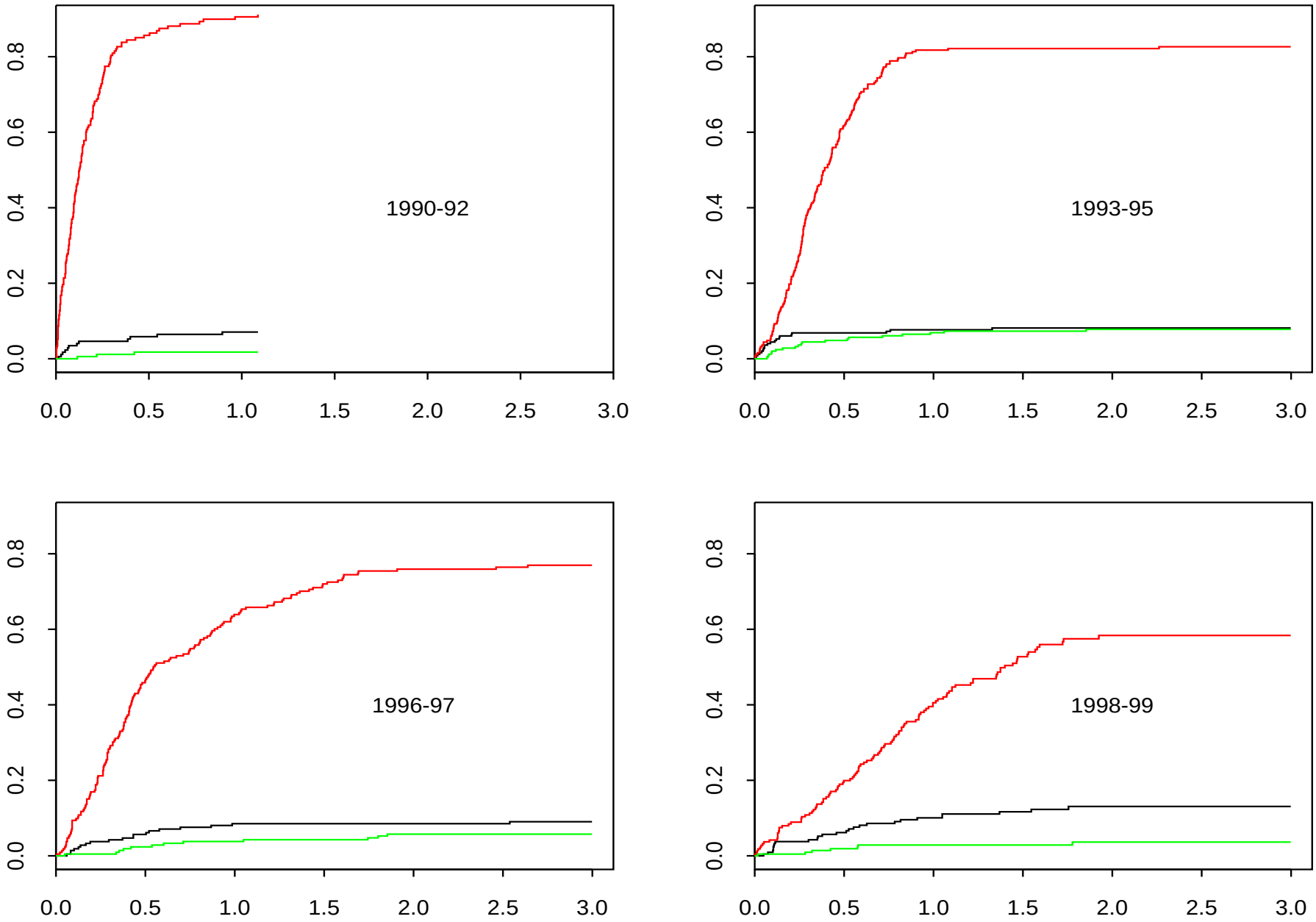
Note that the cumulative incidence curves sum to the total one.

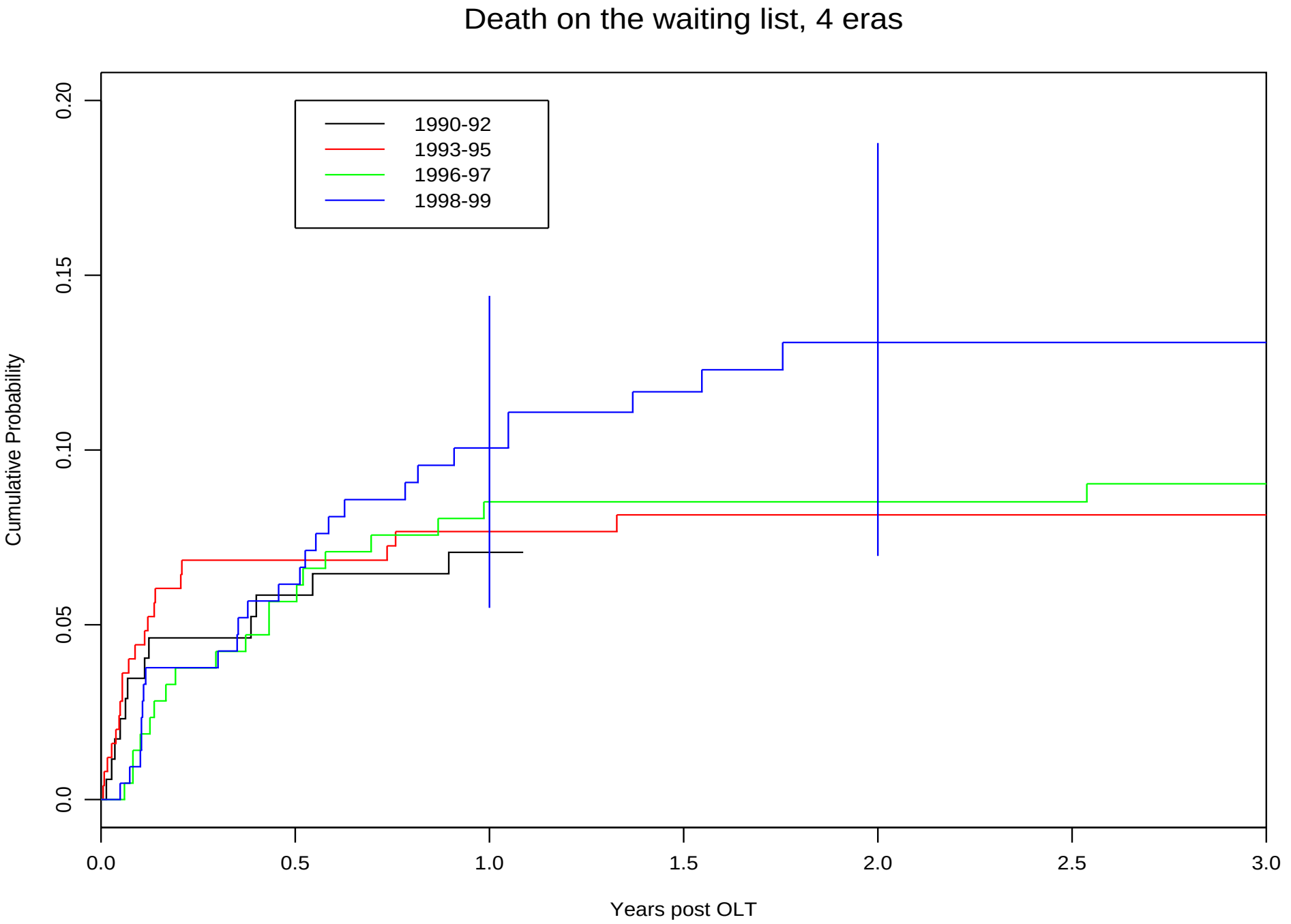
The CI estimate *partitions* the causes of failure.

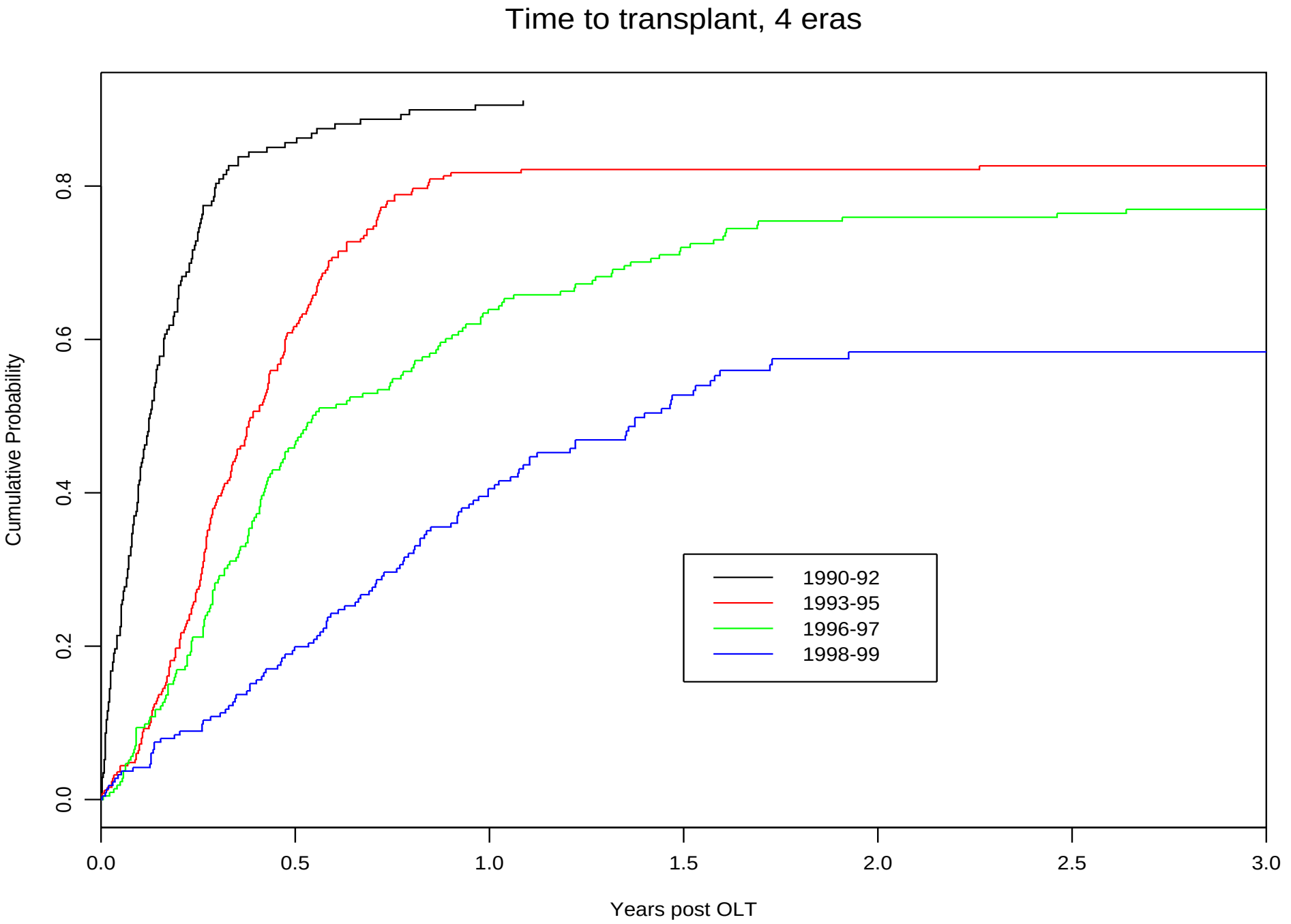
The KM for death and the CI for death in this data set are quite different

- The KM estimates the expected death rate, among waiting list patients, if OLT and withdrawal were to become unavailable.
- It makes a very strong, and in this case completely untenable assumption that those who do get a transplant on a given day are no different than those who did not, with respect to future survival.
- In this data, the *actual* effect of changes over time is the more interesting question. This is what the CI curve estimates.


Death on the waiting list
Cumulative Probability
0.0
0.1
0.2
0.3
0.4
KM
CI
0
1
2
3
4
5
Years post listing

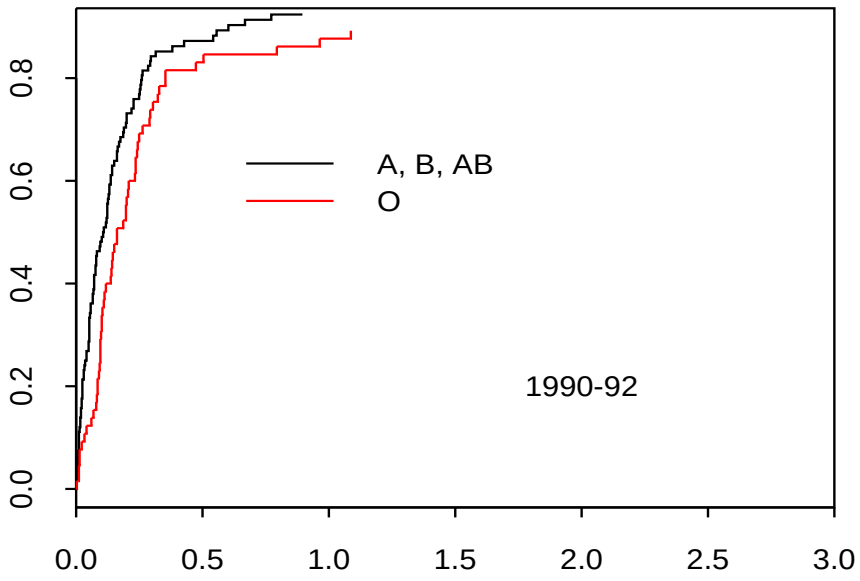

1990-92
1993-95
1996-97
1998-99
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0

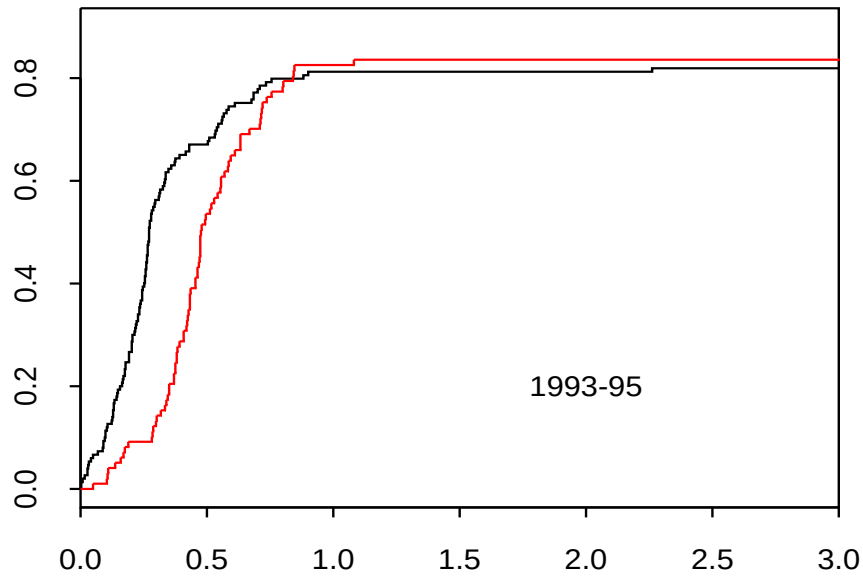

Death on the waiting list, 4 eras
1990-92
1993-95
1996-97
1998-99
Cumulative Probability
0.0
0.05
0.10
0.15
0.20
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Years post OLT

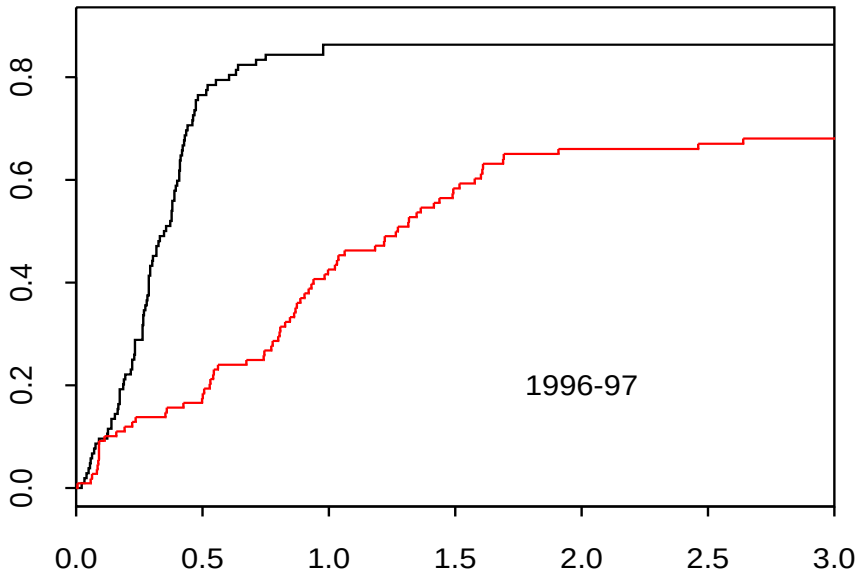

Time to transplant, 4 eras
Cumulative Probability
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Years post OLT
1990-92
1993-95
1996-97
1998-99

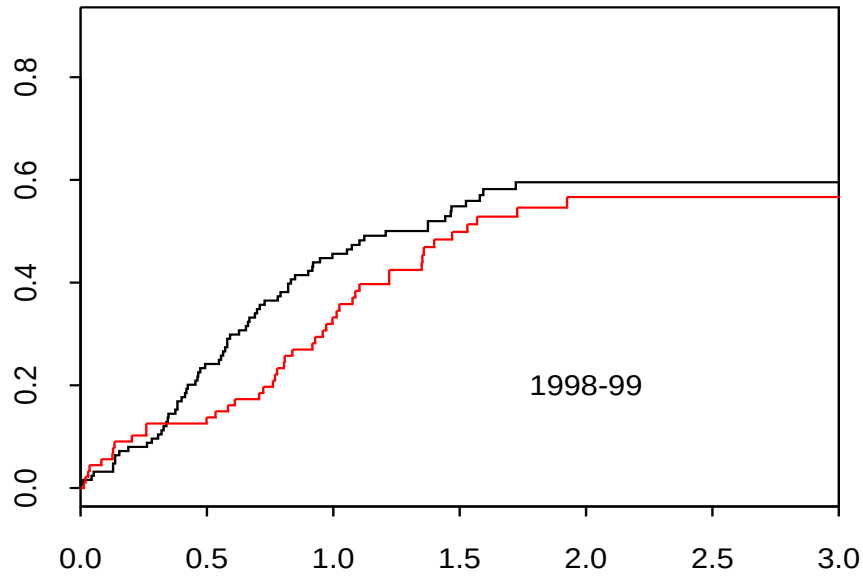
- Transplant has been significantly delayed over the time period.
- Death on the list has not changed substantially.

It is interesting to look at transplant delay, by blood type.


A, B, AB
O
1990-92
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0


1993-95
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0


1996-97
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0


1998-99
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Extension to Cox models

Again, one must choose a particular ”subject” for the curves.

```
fit <- coxph(Surv(time, status) ~ hgb + (age + mspike) * strata(etype),
        data=mgus2)
tdata <- data.frame(hgb=13, age=65, mspike=2)
sfit  <- survfit(fit, newdata=tdata)

# Get the 3 hazard curves
# Total hazard = sum
# Total survival = exp(- total hazard)
# CI = sum(hazard * survival)
```

Key Issue

1. The existence of a competing risk λ_2 does not change a risk λ

- Starting to skydive does not lower the risk of cancer
- But the added accident risk may lower the lifetime risk of cancer occurrence

2. A competing risk does not usually change to Cox model

- A model of instantaneous risk
- Requires weak independence of causes
  - ⋆ That relative hazards are undisturbed

3. A competing risk totally changes the survival curves after a model

- Individual curves become smaller

Example

Imagine 2 causes of death A and B

- Difficult to distinguish
  - ⋆ one-third of the A's are mistakenly called B's
  - ⋆ Essentially random
- Causes are not independent
- Cox analysis for A, treating B as censoring, is not affected
- CI curve is seriously biased

Informative Censoring

When dividing the weight equally among all those still at risk, the redistribute algorithm makes a key assumption:

The censored subject is a random choice from among those at risk.

In the case of liver transplant, this is patently false.

Solutions:

- Give the weight unevenly, to those “most like” the censored subject.
  - ⋆ In a Cox model *that has all the correct variables and coefficients* this is not necessary.
  - ⋆ Leads to time-dependent weights in the Cox model.
- Give the weights to fictitious subjects.

Liver transplant

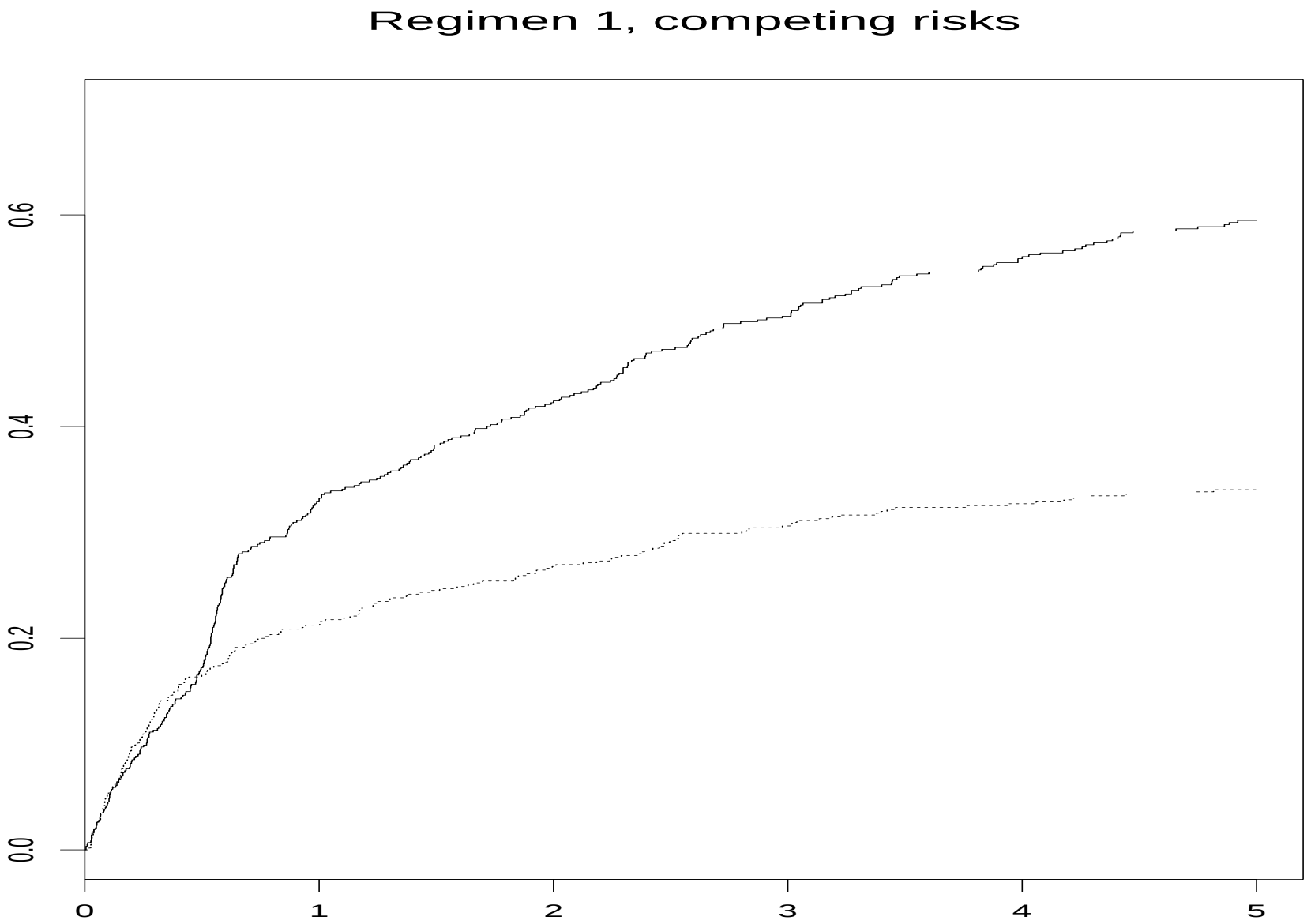
For the waiting time data, competing risks analysis has been very useful.

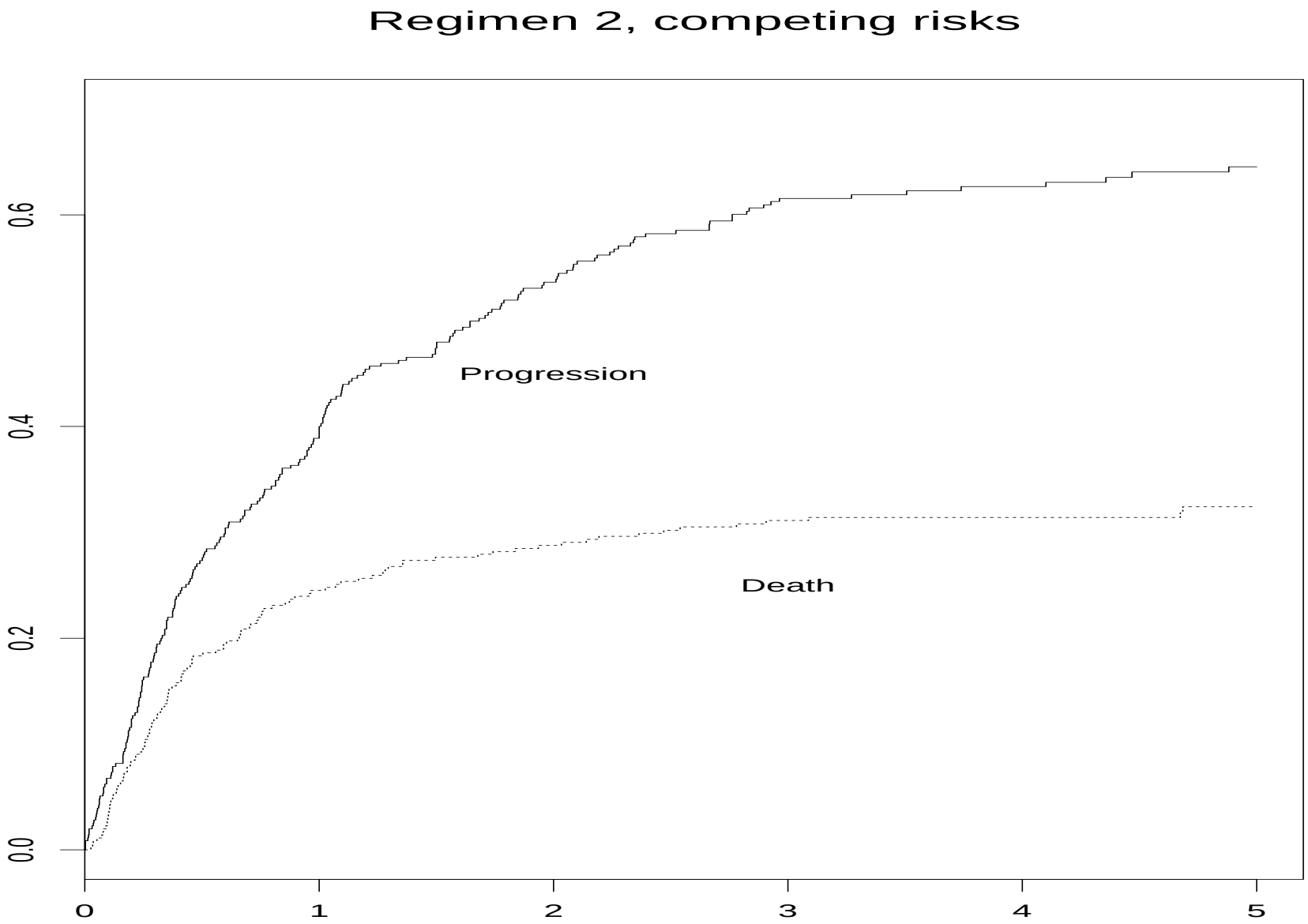
For examining the natural history of disease, for which transplant is a nuisance, alternate distribution of weights has been useful.

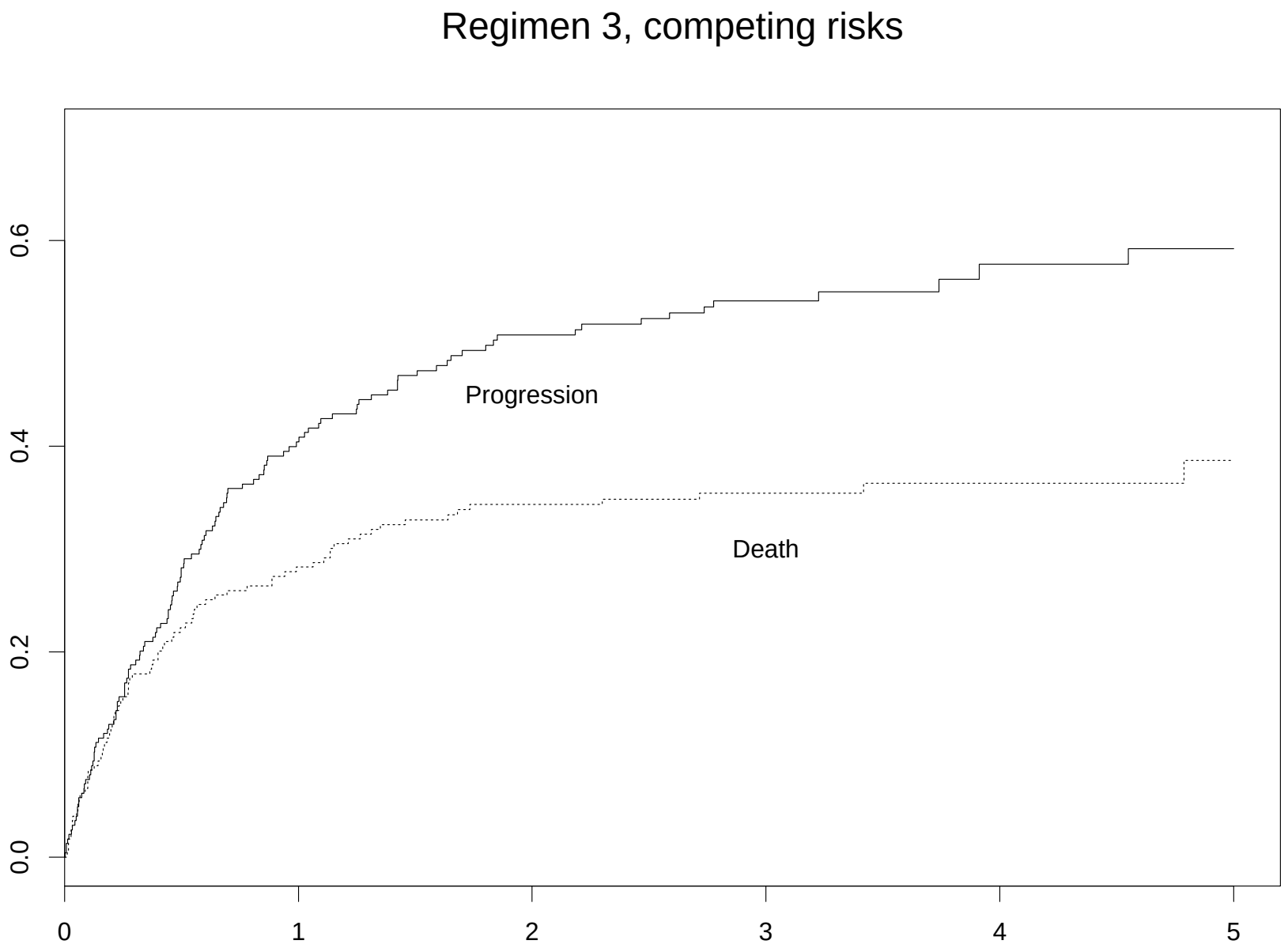
Progression of Multiple Myeloma

- 1027 MM patients from 1/1985 to 12/1998
- Subset to the 578 with “local management” of their treatment.
- What is the overall course of these subjects?

	Current status		
Treatments	alive	dead	
1	12	211	578
2	16	114	355
3	23	80	225
4	12	46	112
5	2	29	64
6+	4	29	33
Total	69	509	


Regimen 1, competing risks
0.0
0.2
0.4
0.6
0
1
2
3
4
5


Regimen 2, competing risks
Progression
Death
0.0
0.2
0.4
0.6
0
1
2
3
4
5

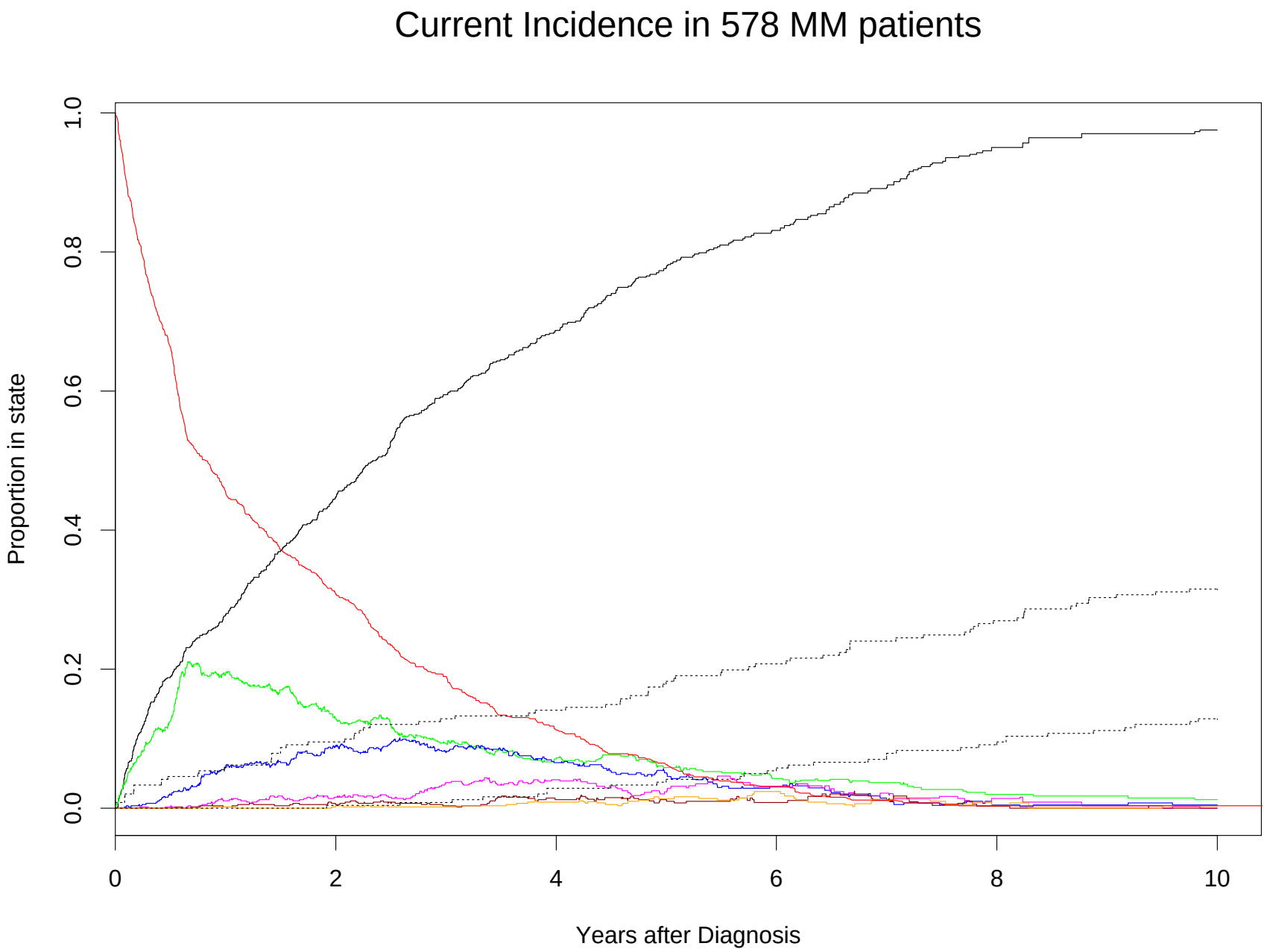

Regimen 3, competing risks
Progression
Death
0.0
0.2
0.4
0.6
0
1
2
3
4
5

Current prevalence curve

An overlook of “current patient status”

- Data set of 1377 observations
- ctime = time from start of current regimen to departure
- status= 0/censored, 1/death or new treatment
- endpoint = went to treatment 1/2/3/4/5/6/death

```
fit1 <- survfit(Surv(ctime, status) ~ 1, event=endpoint, id=clinic,
                                data=mm1)
plot(fit1, col=c(1,3:8), fun='event', mark.time=F, xscale=365.25,
                  xmax=3652)
lines(fit1$time/365.25, 1-apply(1-fit1$surv,1,sum), col=2)
```


Current Incidence in 578 MM patients
Proportion in state
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Years after Diagnosis

Delayed entry

For a large study of 3914 multiple myeloma (MM) patients seen at the Mayo Clinic from 6/47 through 2/96 we had

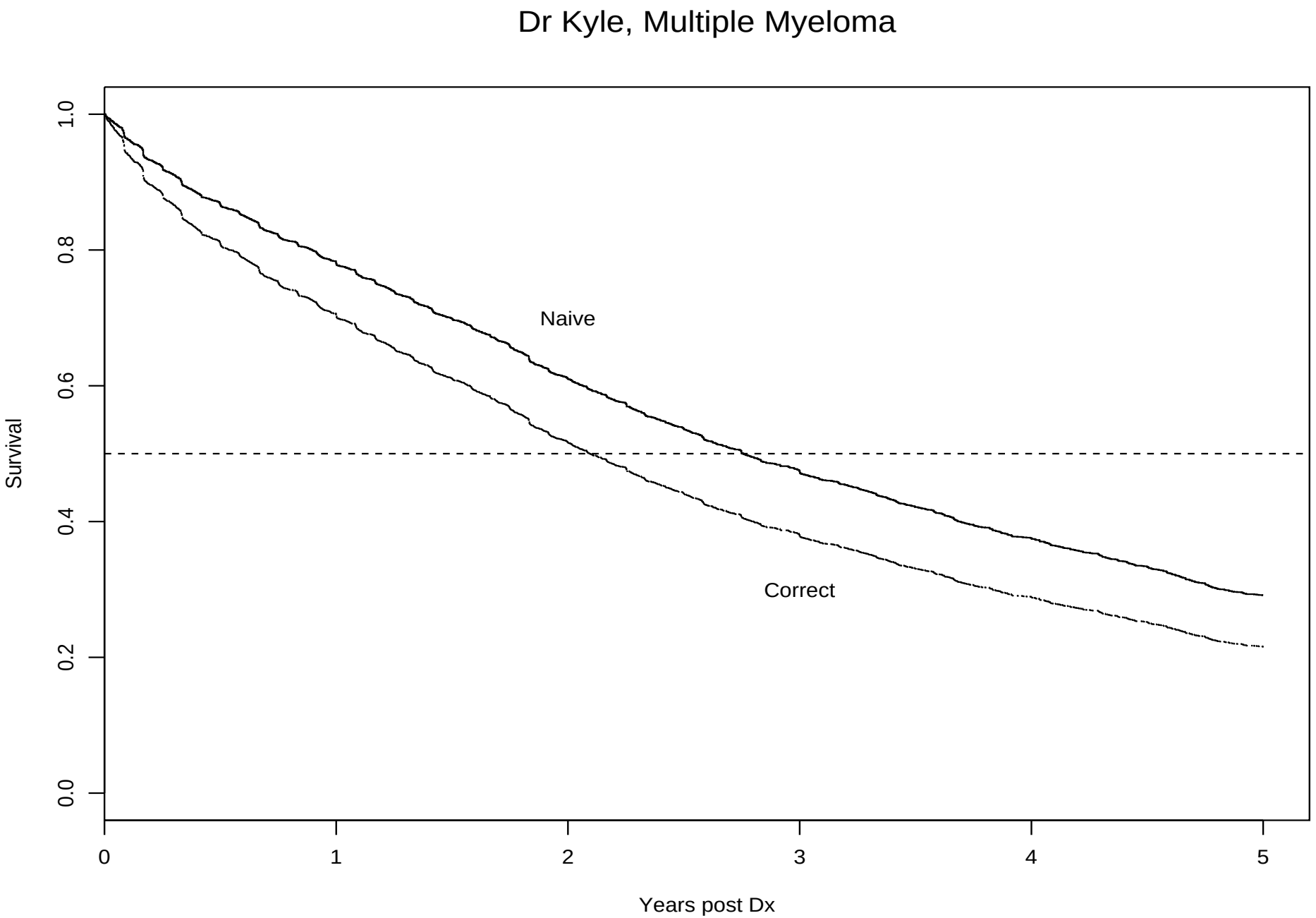
- Date of diagnosis of MM
- Date of examination at Mayo
- Date of death or last follow-up
- Status at last follow-up.

For clinical reasons, the principle investigator (Dr. R. Kyle) wanted to use "survival from diagnosis of MM", rather than "survival from arrival at Mayo".

On this time scale the data is left-truncated: a patient who comes to Mayo 1 year after diagnosis *would not have been seen* had death occurred at < 1 year.

- Long survivors are over-represented.

- The usual survival curve will be biased upward


Dr Kyle, Multiple Myeloma
Survival
1.0
0.8
0.6
0.4
0.2
0.0
Naive
Correct
0
1
2
3
4
5
Years post Dx

Computation

SAS

```
%survtd(strttime=entry, stoptime=futime, event=status,
        data=myeloma, ...)

proc phreg data=myeloma;
   model (entry, futime) * status(0) = ;
   baseline out=newdata surv=surv;
```

S-Plus

```
> fit1 <- survfit(Surv(futime, status) ~1, data=myeloma)
> fit2 <- survfit(Surv(entry, futime, status) ~1, data=myeloma)

> plot(fit1, mark.time=F, conf.int=F, xmax=5,
        xlab="Years Post Dx", ylab="Survival")
> lines(fit2, mark.time=F, lty=2, xmax=5)
> abline(h=.5, lty=3)
> text(locator(2), c("Naive", "Correct"))
```

With delayed entry, e.g., referral patients who enter a group some time after their diagnosis, the Kaplan-Meier curve is

$$S(t) = \prod_{s \leq t} \left(1 - \frac{d(s)}{r(s)}\right)$$

where

$$\begin{aligned} d(t) &= \text{number of deaths at time } t \\ r(t) &= \text{number at risk at time } t \\ &= \sum_{i=1}^{n} Y_i(t) \\ Y_i(t) &= \begin{cases} 1 & \text{if subject } i \text{ is at risk at time } t \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

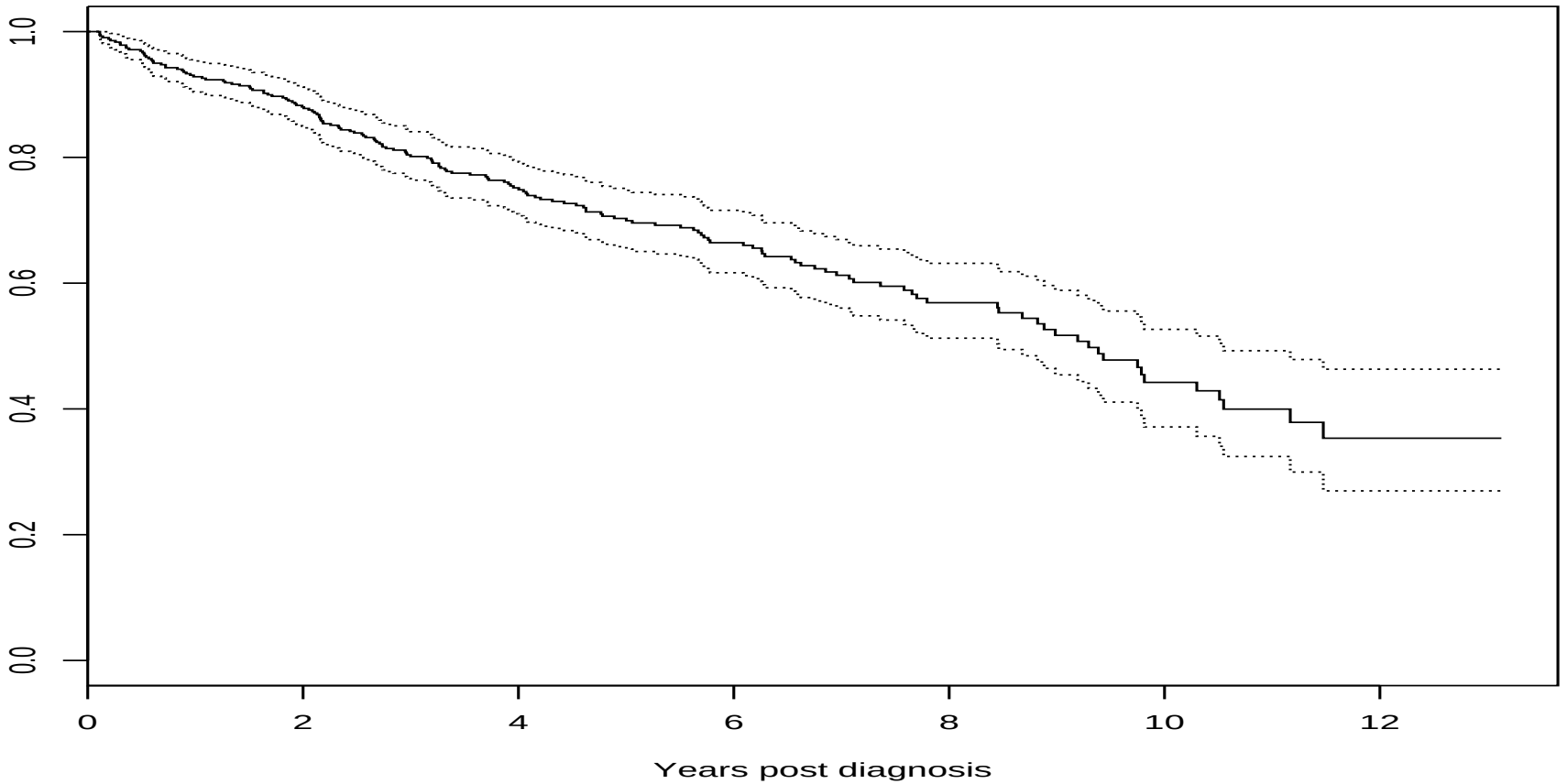
- The usual KM has $r(t)$ = number who have not yet died at time t
- The difference here is that one does not enter the denominator until you have arrived
- You can't be in the denominator unless you can be in the numerator.
- "Not counted as being in the lottery drawing, unless you could have won the drawing."

PBC study

Between January, 1974 and May, 1984 the Mayo Clinic conducted a randomized double-blind study in primary biliary cirrhosis of the liver (PBC), comparing the drug D-penicillamine (DPCA) with a placebo. There were 424 patients who met the eligibility criteria of the trial, of whom 312 agreed to participate.

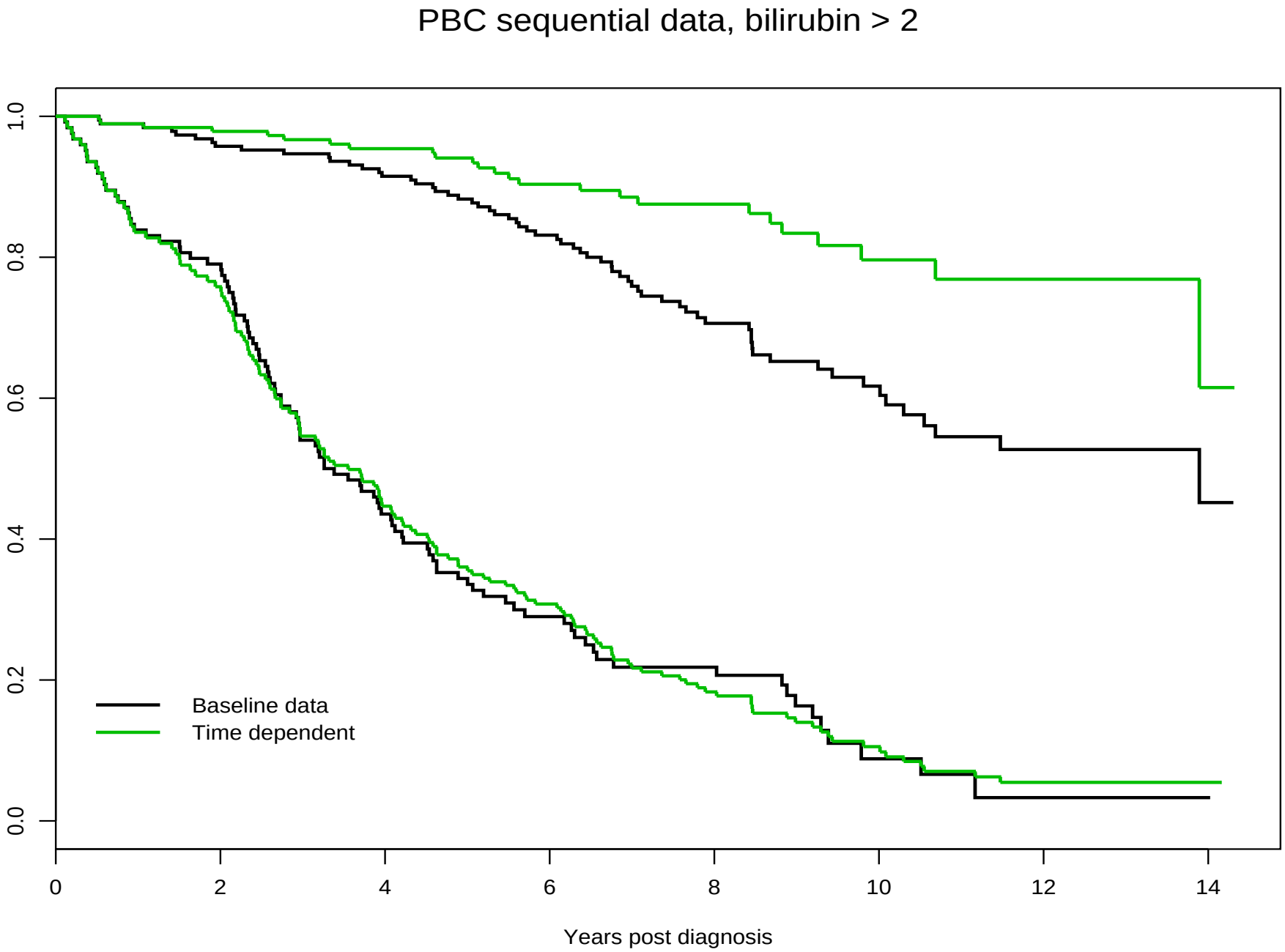
PBC is a rare but fatal chronic liver disease of unknown cause, with a prevalence of about 50 cases per million population. The primary pathologic event is destruction of the interlobular bile ducts, possibly by an immunologic mechanism. Progression is slow but inexorable, with a median survival of about 8 years. Females comprise 90% of the cases.

Patients were scheduled for a repeat visit at 6 and 12 months, and then yearly thereafter. The primary analysis was conducted in July, 1986, two years after the last patient had been entered. At that time 125 of the 312 patients had died, eight patients had been lost to follow-up, and 19 had undergone orthotopic liver transplantation (OLT).


1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
12
Years post diagnosis

What is the impact of a bilirubin greater than 2?

```
fit1 <- survfit(Surv(fu.days, status>0) ~ I(bili >2),
                data=pbcseq, subset= (start==0),
                xlab="Years post diagnosis")

fit2 <- survfit(Surv(start, stop, event>0) ~ I(bili>2),
                data=pbcseq)

plot( fit1, mark.time=F, xscale=365.25, lwd=2)
lines(fit2, mark.time=F, xscale=365.25, col=4, lwd=2)

title("PBC sequential data, bilirubin > 2")
legend(.1, .2, c("Baseline data", "Time dependent"), l
       ty=c(1,1), col=c(1,4),
       lwd=2, bty='n')
```


PBC sequential data, bilirubin > 2
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
12
14
Years post diagnosis
Baseline data
Time dependent

- Why is the second curve different?
  - ⋆ but only for the good subset (bili ≤ 2?
- What do the two curves mean?

- As PBC progresses, bilirubin stays in control for a time, then gradually rises, then more rapidly rises

- For the *bilirubin* $\geq$ *2* curve
  - ⋆ If Pr(failure $\mid$ bili > 2) is independent of how long ago the bilirubin exceeded the threshold
  - ⋆ Then the lower curve is an estimate of failure for those who start with a high bilirubin
    - ∗ more precise due to a larger n
  - ⋆ If not, which is often likely, then the lower curve is an estimate of . . . I don't really know

- The upper curve is survival for someone whose bilirubin never goes over 2.

Summary

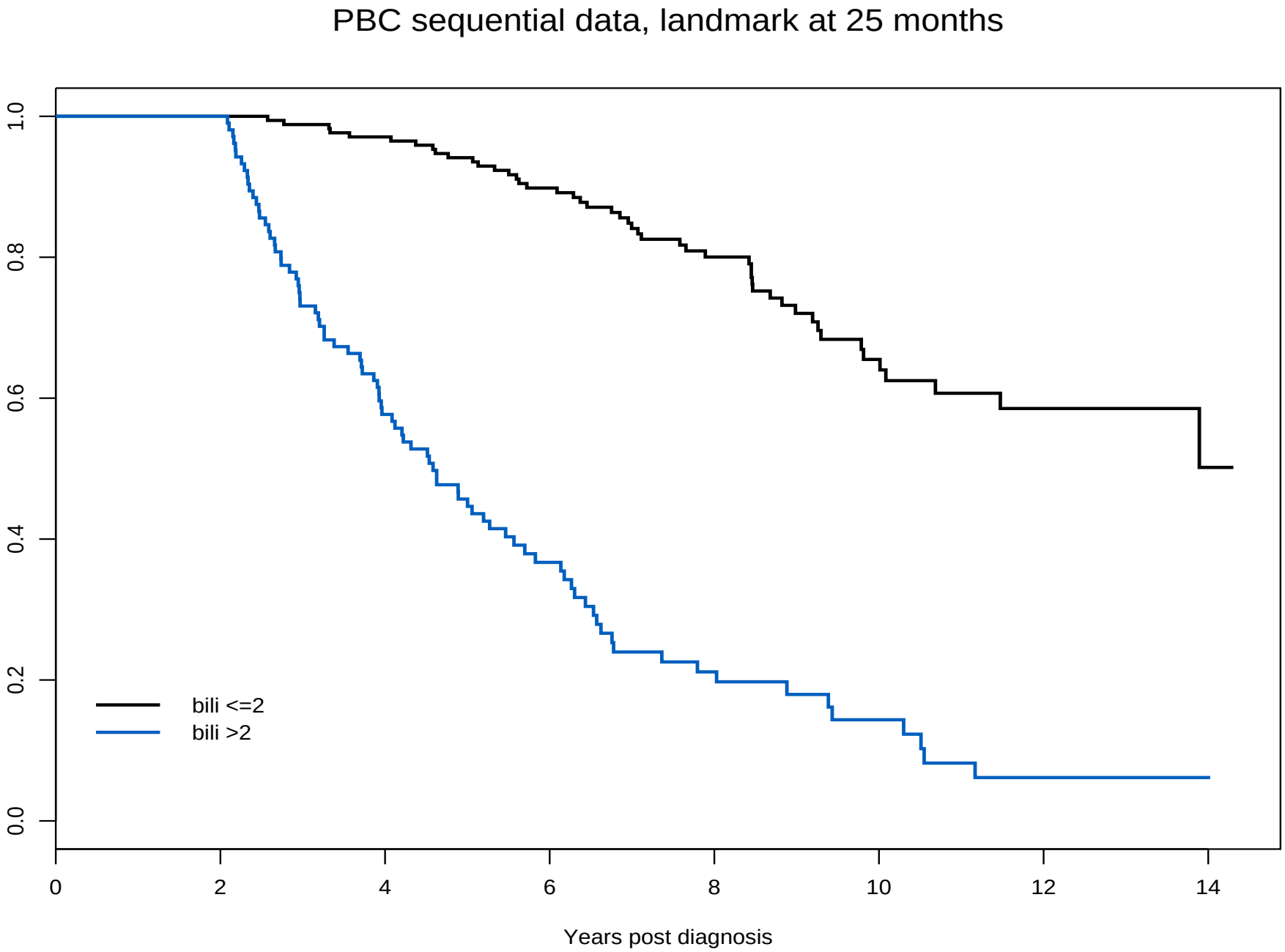
The [start, stop) method produces a nice looking curve.

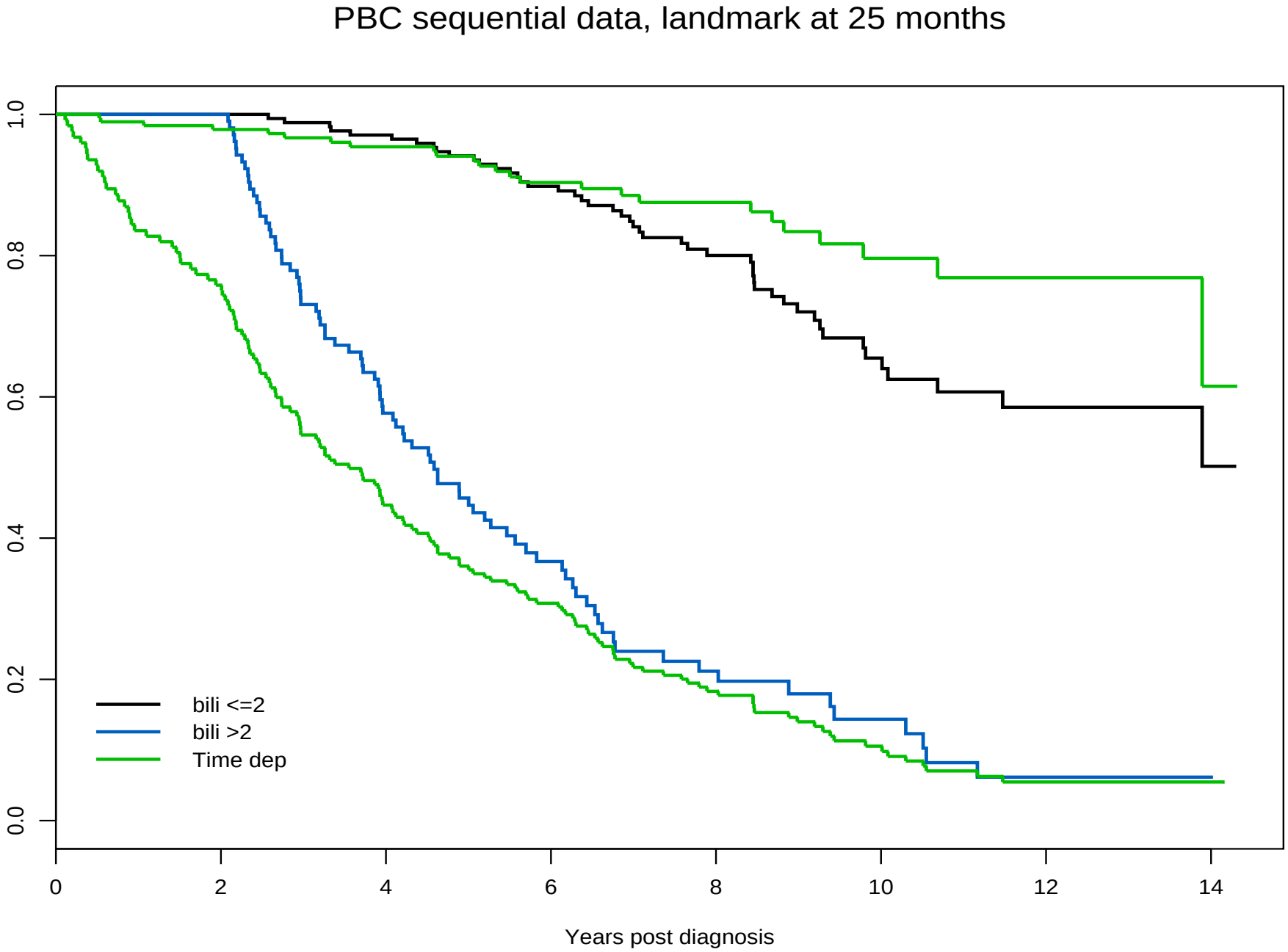
But we have no clear idea of what the curve is.

Landmark curves

The problem with the time dependent covariate curves is that they do not answer a clinical question: “This is the survival of a person with x, y and z”.

- They most certainly are not a description of the *effect of high bilirubin*
- The *landmark method*
  - Given a patient, sitting across the table from the physician
    - 25 months after diagnosis
    - bilirubin of 2.1
  - What is this person’s predicted survival?
  - Method
    - categorize all patients still alive at 25 months into high/low bilirubin
    - using the value of bilirubin *at 25 months*
    - compute and draw a survival curve forward in time, treating this as a *time fixed* covariate


PBC sequential data, landmark at 25 months
1.0
0.8
0.6
0.4
0.2
0.0
bili <=2
bili >2
0
2
4
6
8
10
12
14
Years post diagnosis


PBC sequential data, landmark at 25 months
1.0
0.8
0.6
0.4
0.2
0.0
bili <=2
bili >2
Time dep
0
2
4
6
8
10
12
14
Years post diagnosis

Landmark

Advantages

- Answers an actual question
  - ⋆ What is the survival for a patient who is
    - ∗ alive at 25 months
    - ∗ with a bilirubin ≤ 2
- the question is relevant to someone
- the answer is correct for that someone
- recognizes that some fraction will progress to bilirubin > 2

Disadvantages

- You have to pick a time
  - ⋆ Late enough to have population in both curves

- ⋆ Early enough to be relevant

- Sometimes multiple landmark points must be chosen.

- A Cox model is based on *instantaneous* rates
  - ⋆ Time dependent covariates are not a problem, in theory
  - ⋆ A subject can trade from group to group often
  - ⋆ Only necessary that the *current* covariates decribe the *current* risk.
- A survival curve summarizes *cumulative* experience
  - ⋆ Difficult to describe just exactly what a time dependent grouping means