Short course:

Survival Analysis Models & Statistical Methods

Presenter: **Eric V. Slud**, Statistics Program, Mathematics Dept.,
University of Maryland at College Park, College Park, MD 20742

The objective is to introduce first the main modeling assumptions and data structures associated with right-censored survival data; to describe the most successful methodological ideas for analyzing such data that have arisen over the past three decades in the biostatistical literature; and then to show how the key theoretical developments have resulted in practical computational strategies both for fitting survival models to data and for assessing the fit of those models. Data illustrations will highlight the comparison and modelling of cancer survival data taken from the National Cancer Institute's **SEER 9** database (`www.seer.cancer.gov`), with scripts in **R**.

Outline:

(O) Data-Structure for (Right-) Censored Survival Data

Lexis diagrams; data analysis objectives. SEER data examples.

(I) Parametric Hazard Models

Miscellaneous models illustrating hazard and survival function shapes; Weibull MLE's and confidence intervals displayed for SEER data; hazard-rate estimators for piecewise-constant-hazard models (C.-L. Chiang 1960), 'actuarial' estimates.

(II) Life Tables , Nonparametric Hazard & Survival Curve Estimators

Observed-death and at-risk counting processes; compensators and martingales; ref. 2×2 *table idea of Mantel (1963); Nelson-Aalen & Kaplan-Meier, ref. Aalen (1975), Breslow-Crowley (1974), Gill (1983).*

(III) Comparing populations, cont'd.

Median survival time and Confidence Interval; ref. Slud, Byar and Green (1984), Strawderman and Wells (1997); nonparametric vs. parametric survival curves, e.g. Kaplan-Meier vs. Weibull, ref. Miller (1983); goodness of fit of survival curve estimates, incl. comparison using SEER data.

(IV) Competing Risks.

Competing risks, dependent censoring; ref. Gail (1975), Tsiatis (1975), Peterson (1976), Slud and Rubinstein (1983).

(V) Two-sample test statistics.

(weighted) logrank, power, sample-size, and ARE. (Peto & Peto 1972, Slud 1982, Tarone and Ware 1977) Brief mention of (group-) sequential ideas. (Tsiatis 1982; Slud and Wei 1982; Slud 1984)

(VI) Semiparametric Models for Survival-curve differences.

Cox model and partial likelihood (Cox 1972, 1975, Andersen and Gill 1983, Wong 1986, Slud 1992). Frailty Models, EM vs profile-likelihood fitting method (Nielsen et al 1992, Kosorok et al. 2004, Slud and Vonta 2004). Data analyses and comparisons; time-dependent covariates (Andersen et al. 1993). Goodness of fit, Martingale residuals. (Schoenfeld 1980)

Specific topics to be be covered include: Life Tables, Censoring Mechanisms, Regression Models involving conditional survival functions and hazards, Competing Risks and their Pitfalls, Martingales and Survival Model Residuals, Partial Likelihoods for Survival Data, Model Diagnostics and Time-dependent covariates, introduction to Frailty Models, and Goodness of Fit in Survival Models. Each topic will proceed from new concepts to interpretation of central theoretical results to a data example.

Idea of Life Table

- Define "entry" by one of: *birth, test or diagnosis, surgery, etc.* Keep other important age or cohort variables in re-coded form *as covariates*.
- Record event-time from entry, and whether study endpoint (e.g. failure) or time of loss to followup (censoring/withdrawal).

Key questions: (1) can survival in different cross-classified groups be characterized or compared in terms only of observable data recording numbers of survival events up to t and censoring events up to t ?

(2) If censoring and other conditions differ across groups, can survival functions be estimated ?

(3) Can survival functions be corrected for (some) covariate imbalances and compared across groups ?

Data Example: SEER 9 registry, 1973-2001 mortality of diagnosed Lymphoma patients

COVARIATES: Age at diagnosis, Birth year, stage & grade of tumor, whether single/primary tumor, location of patient, surgery and/or radiation treatment indicator, diagnosis confirmation indicators, Race, Ethnicity, Sex.

DATA FORMAT FOR A SURVIVAL STUDY

Subjects enter at random times E_i , 'followed' until

$E_i + T_i = \min(E_i + X_i,\ E_i + C_i)$ (not both observed)

'death-time' ($X_i =$ *lifetime*), or 'censoring time'
(e.g., $C_i = E_{\max} - E_i + \tau$ *administrative*)

Data: $\{(E_i,\ T_i,\ \Delta_i,\ Z_i)\ ,\ i = 1, \ldots, n\}$ or

$\mathcal{D} = \{(T_i,\ \Delta_i)\ ,\ i = 1, \ldots, n\}$ where

$T_i =$ *time-on-test* or *event time*

$\Delta_i = I_{[X_i \le C_i]}$ *death indicator*

Z_i *auxiliary covariates*, e.g. group indicator ξ_i;
may be time-dependent obs on $[0, T_i)$

Objective: to estimate the marginal survival function $S_X(t) = P(X_1 > t) = 1 - F_X(t)$ consistently from the data $\mathcal{D}$.

Assumptions: random vectors $(E_i,\ X_i,\ C_i,\ Z_i)$ independent & identically distributed (*iid*), $i = 1, \ldots, n$;

also (X_i, C_i) have continuous *joint density*, i.e.

$\lim_{\delta \searrow 0} \frac{1}{\delta^2} P(X_1 \in (x, x+\delta),\ C_1 \in (c, c+\delta)) = f_{X,C}(x, c)$

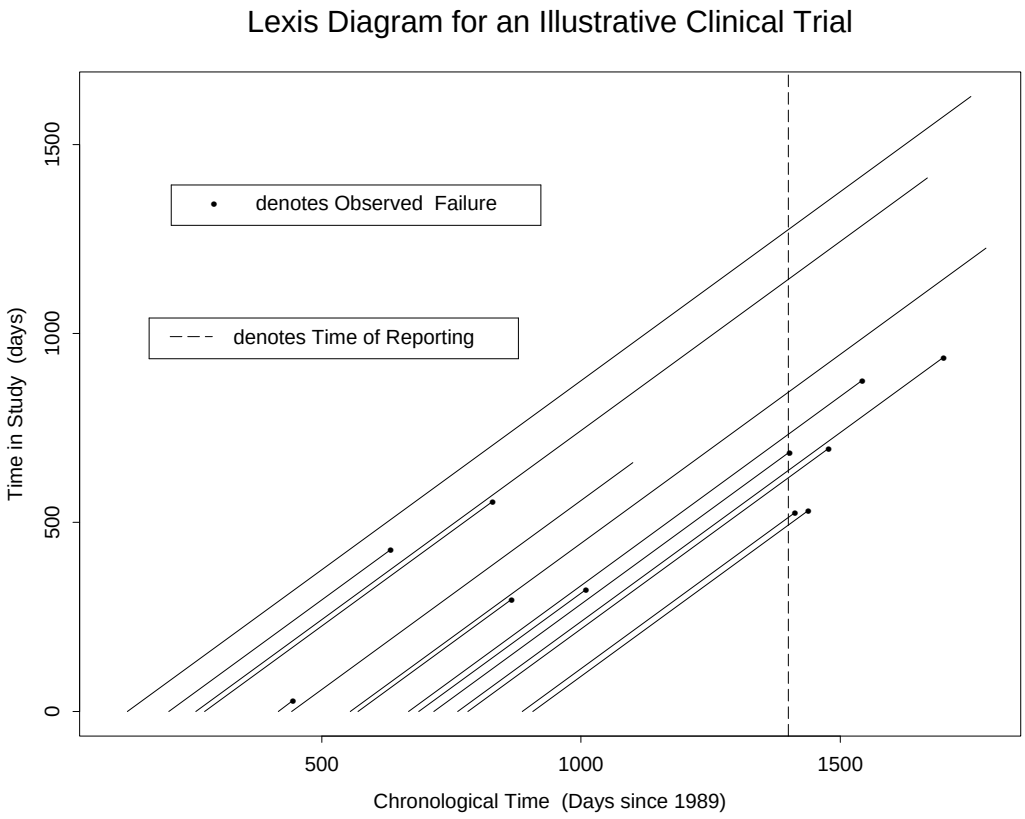

Figure 1: "Lexis Diagram" (from contributed article to Encyclopedia of Biostatistics): from entry, patients' followup is pictured as 45° line: solid dot represents death, line not ending in dot representes censoring.

Death Hazards

In general, define **hazard intensity**

$$h_X(t) \equiv \lim_{\delta \to 0} \frac{1}{\delta} P(X \in (t, t+\delta) \,|\, X > t) = \frac{f_X(t)}{S_X(t)}$$

Then

$$h_X(t) = -\frac{d}{dt} \ln S_X(t) \quad \Rightarrow \quad S_X(t) = \exp\left(-\int_0^t h_X(s)\, ds\right)$$

So **hazard** is instantaneous mortality rate conditional on previous survival, and the integrated form of **cumulative hazard**

$$H_X(t) = \int_0^t h_X(s)\, ds = -\ln S_X(t)$$

is also very useful in specifying survival models.

Major Cases:

(i) *Constant hazard rate*: $h_X(t) \equiv \lambda$

occurs only when $H_X(t) = \lambda t$, $S_X(t) = e^{-\lambda t}$
for Exponential random variable X

(ii) *Increasing hazard rate* = Aging, wearing-out

(iii) *Decreasing hazard rate* = 'Burning-in', mixture of exponential

Examples of Survival Hazards

- 'Multi-hit model' $X = V_1 + V_2 + \cdots + V_r$ with indep. waiting times V_j for 'shocks', mutations, etc.

  If V_j *iid* Expon(λ), then $X \sim$ Gamma(r, λ) increasing-hazard if $r > 1$.

- 'Mixture model' $X \sim \text{Expon}(\tau)$, $\tau \sim G$ r.v.

  Then can prove $h_X(t)$ decreasing : the idea is that individuals (X_i, τ_i) with higher τ_i die early !

- Weibull(λ, γ) power-law hazard $h(t) = \lambda\,\gamma\,t^{\gamma-1}$; scale and power transformation of $V \sim \text{Expon}(1)$:

  $(V/\lambda)^{1/\gamma} \sim \text{Weib}(\lambda, \gamma)$ because:

  $S(t) \;=\; P((V/\lambda)^{1/\gamma} > t) \;=\; P(V > \lambda\,t^{\gamma}) \;=\; e^{-\lambda t^{\gamma}}$

  Hazard $h(t)$ $\nearrow$ for $\gamma > 1$, $\searrow$ for $\gamma < 1$

- *Bathtub-shaped* hazards in *Makeham* model:
  $h(t) \;=\; A + Be^{ct} \qquad (A, B, c > 0)$

  only if we add power-law term $\lambda\,\gamma\,t^{\gamma-1}$, $\gamma < 1$.

Pictures follow:

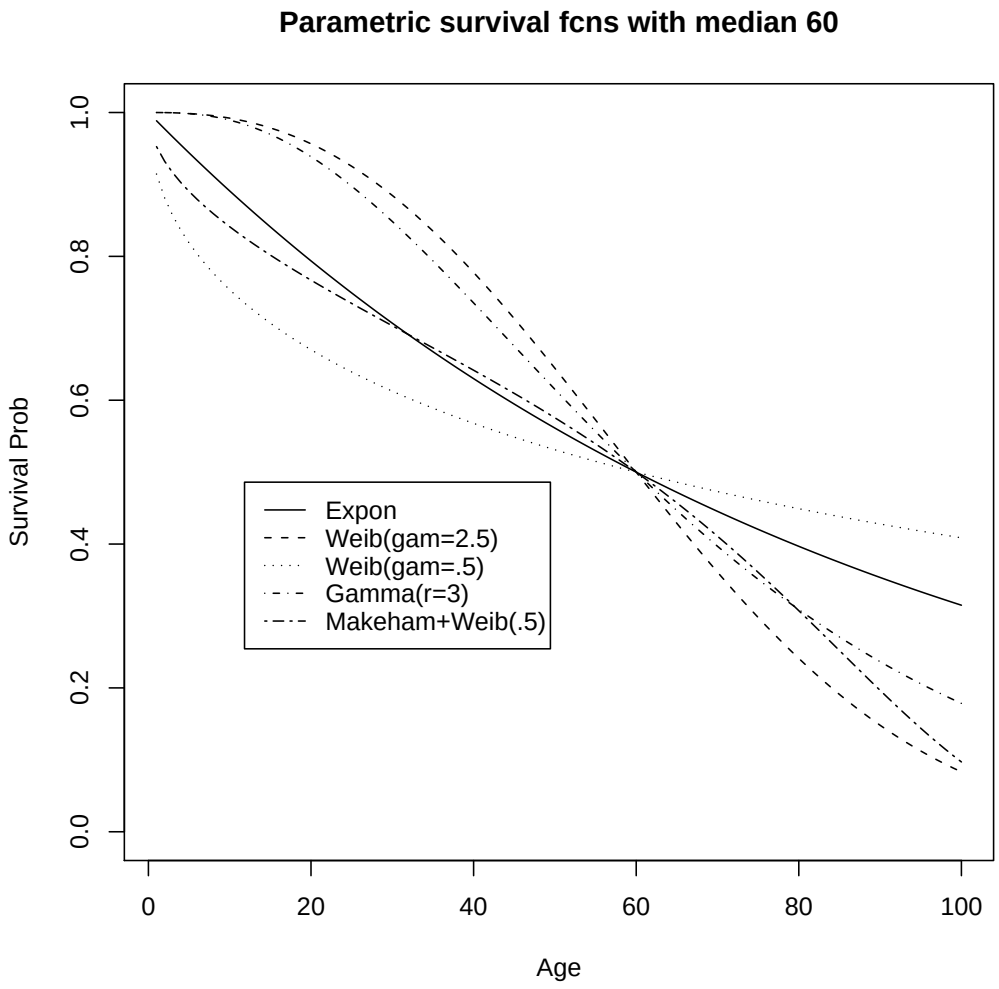

Figure 2: Graphs of survival functions from several parametric models designed to have common median 60.

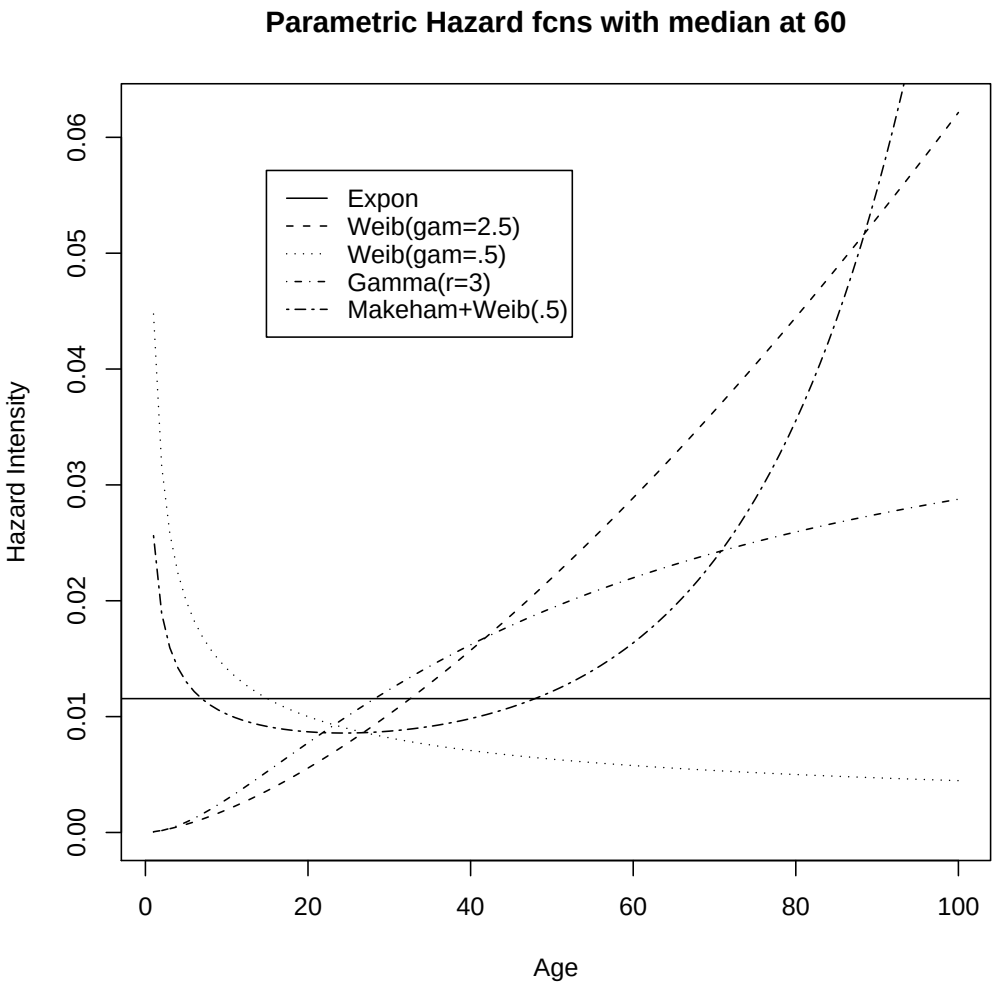

Figure 3: Graphs of cumulative hazard functions of several parametric models designed to have common median 60.

Parametric Survival-Data Likelihood

Suppose that underlying (latent) waiting times X_i, C_i to death and censoring are **independent** for each subject $i = 1, \ldots, n$; densities are $f_X = f_X(\cdot, \vartheta)$, f_C and survival functions are $S_X = S_X(\cdot, \vartheta)$, S_C. For observed data $(T_i, \Delta_i) = (\min(X_i, C_i),\ I_{[X_i \le C_i]}$:

$$Lik(\vartheta) \ = \ \prod_{i=1}^{n} \{(f_X(T_i, \vartheta)\, S_C(T_i))^{\Delta_i}\, (f_C(T_i)\, S_X(T_i, \vartheta))^{1-\Delta_i}\}$$

Censoring density f_C is unknown but does not contain parameters to be estimated, so dropping factors f_C, S_C and writing $f_X = h_X S_X = h_X\, e^{-H_X}$ leaves

$$\text{logLik}(\vartheta) \ = \ \sum_{i=1}^{n} (\Delta_i\ \log h_X(T_i,\ \vartheta) \ - H_X(T_i, \vartheta))$$

$$= \int \{\log h_X(t, \vartheta)\, dN(t) \ - \ \sum_{i=1}^{n} I_{[T_i \ge t]}\, h_X(t, \vartheta)\, dt\}$$

where

$$N(t) \ = \ \sum_{i=1}^{n} \Delta_i\, I_{[T_i \le t]}$$

defines the *observed death counting process*, and the *at-risk process* is

$$Y(t) \ = \ \sum_{i=1}^{n} I_{[T_i \ge t]}$$

MLE's from Parametric Survival Likelihood

$$\text{logLik}(\vartheta) \;=\; \int \left(\log h_X(t,\vartheta)\, dN(t) \;-\; Y(t)\, h_X(t,\vartheta)\, dt \right)$$

leading to likelihood score equation

$$\mathbf{0} \;=\; \int \nabla_\vartheta \, \log h_X(t,\vartheta) \, (dN(t) \;-\; Y(t)\, h_X(t,\vartheta)\, dt)$$

with solution $\hat{\vartheta}$ satisfying in large samples:

$$-\nabla^{\otimes 2} \text{logLik}(\vartheta_0) \, (\hat{\vartheta} - \vartheta_0) \;\approx\; \nabla \text{logLik}(\vartheta_0)$$

(using notation $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}'$), which leads to:

$$\hat{\vartheta} \;-\; \vartheta_0 \;\approx\; \Big(- \int \nabla^{\otimes 2} \, \log h(t,\vartheta_0)\, dM(t) +$$

$$\int (\nabla \log h(t,\vartheta_0))^{\otimes 2} \, Y(t)\, h(t,\vartheta_0)\, dt \Big)^{-1} \int \nabla \log h(t,\vartheta_0)\, dM(t)$$

where $dM(t) \;=\; dN(t) \;-\; Y(t)\, h_X(t,\vartheta_0)\, dt$ is the integrator for martingale *stochastic integrals* and will be seen to have the property that when the model with hazard $h_X(t,\vartheta_0)$ actually governs the data, for each square-integrable function g (wrt $f(t,\vartheta_0)\, dt = h(t,\vartheta_0) S(t,\vartheta_0) dt$) and all large n,

$$E \left(\int g(t)\, dM(t) \right)^2 \;=\; \mathcal{O}(n \int g^2(t)\, f(t,\vartheta_0)\, dt)$$

Under the $f(t,\vartheta_0)$ model, 1st term in $\nabla^{\otimes 2} logLik$ above is $\mathcal{O}(\sqrt{n})$, can be ignored because 2nd is $\mathcal{O}(n)$.

Specialization to Weibull Log-Lik

Likelihood in Weibull(λ, γ) case $[h(t, \vartheta) = \lambda\gamma\, t^{\gamma-1}]$ gives

$$\int \begin{pmatrix} 1/\lambda \\ 1/\gamma + \log(t) \end{pmatrix} (dN(t) - Y(t)\, \lambda\gamma\, t^{\gamma-1}\, dt) = 0$$

First equation uniquely determines $\hat{\lambda}$ in terms of $\hat{\gamma}$ by

$$\sum_{i=1}^{n} \Delta_i = N(\infty) = \int \sum_{i=1}^{n} I[T_i \geq t]\, d(\lambda\, t^{\gamma}) = \lambda \sum_{i=1}^{n} T_i^{\gamma}$$

Second eq'n becomes

$$\int \log(t)\, dN(t)/N(\infty) = -\gamma^{-1} + \sum_{i=1}^{n} T_i^{\gamma} \log(T_i) \,/\, \sum_{i=1}^{n} T_i^{\gamma}$$

and right-hand side is strictly $\nearrow$ in γ.

RESULT. For large n, if Weibull(λ_0, γ_0) model holds, then

$$\sqrt{n}\,(\hat{\vartheta} - \vartheta_0) \approx \mathcal{N}\left(0, \left\{\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1/\lambda_0 \\ 1/\gamma_0 + \log(T_i) \end{pmatrix}^{\otimes 2} \Delta_i\right\}^{-1}\right)$$

If not, large-sample limit ϑ_0 of $\hat{\vartheta}$ still exists and **robust** *misspecified-model variance can also be estimated simply.* (White 1982)

Weibull Models in SEER Data

(16228 White Non-Hisp Hodgkins patients)

(10-yr) Age categories 1-7 (1st 0-20, last $\geq$ 70); Sex; Radiation used; Stage (1, 2, 4 progressively worse; 9 unk)

Summary of estimated Weibull γ parameters over 7 x 2 x 2 x 4 strata.

Min	1st Q	Median	Mean	3rd Q	Max
0.538	0.708	0.8498	0.8820	1.011	1.549

For Male Patients $\geq$ 70:

Stage=		1	2	4	9
Rad=0	λ	.076	.111	.153	.155
	γ	.765	.626	.605	.658
Rad=1	λ	.010	.020	.056	.048
	γ	1.114	.979	.762	.837

(Robust) 95% Conf Int's for λ in the Rad=1 Gp

	Lower	PtEst	Upper	Nsiz
1	.0032	.0103	.0328	65
2	.0067	.0200	.0602	38
4	.0270	.0564	.1178	26
9	.0322	.0485	.0733	108

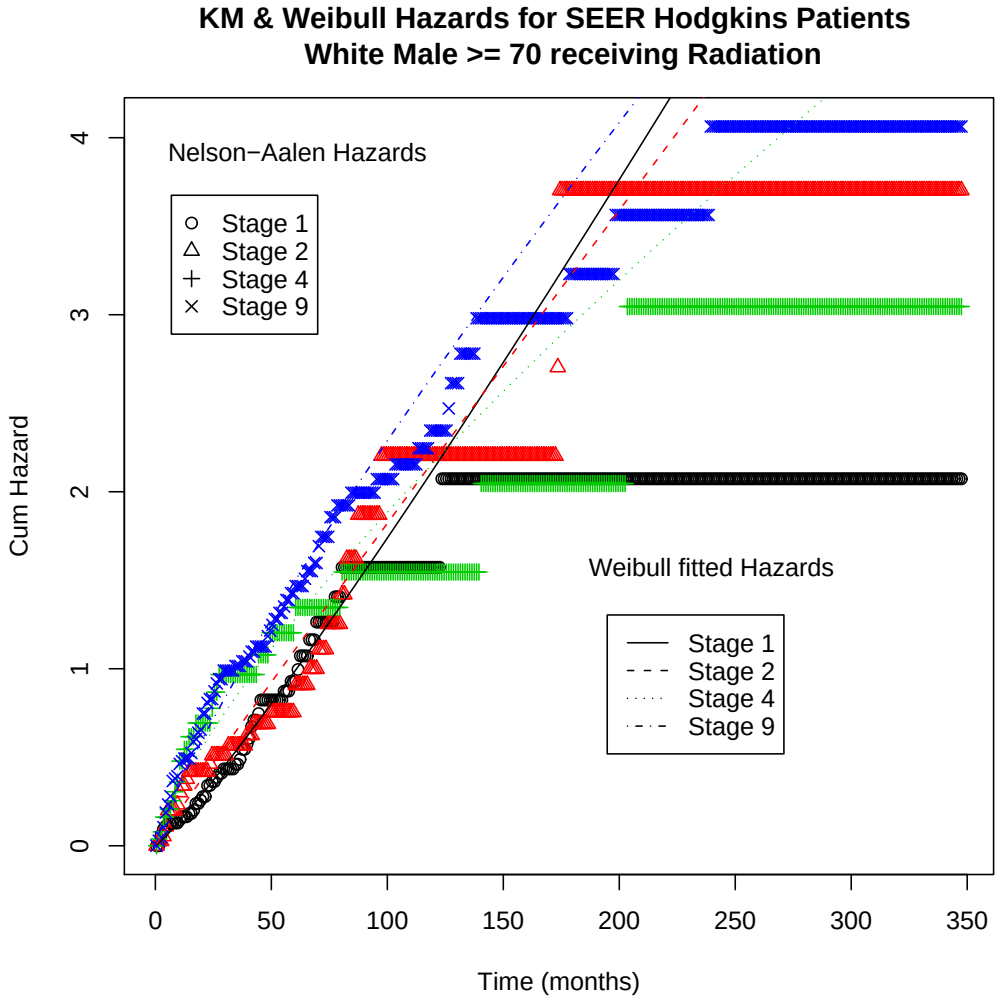

Figure 4: Graphs of estimated Weibull cumulative hazard functions for several SEER groups, Hodgkin White Males ≥ 70 receiving Radiation.

Weibull Fitting, continued

So we can use Weibull parameters **descriptively** with robust variances, which are typically close ($\pm 10\%$) to model- based ones, to describe *and test* population differences.

Pictures (for Stage-wise curves within Male ≥ 70 patients receiving radiation) show some variety, also suggest not-so-good fit in some cases.

Goodness-of-fit of parametric hazard curves still to be considered. Introduce nonparametric curves to test against:

- *Kaplan-Meier* for **Survival**
- *Nelson-Aalen* for **Cum-hazard**

Note that from Weibull, median is estimated $(\ln 2/\lambda)^{1/\gamma}$:

Min	1st Q	Med	Mean	3rd Q	Max
9.75	70.6	262.7	483.8	505.7	7067.

But 348 (months) was the max seen in these data!

Exponential Case

Consider data satisfying Weibull$(\lambda_a, 1) = $ Expon(λ_a) among all surviving uncensored to time a and right-censored at time $b = a + \delta$ $(a \geq 0,\ \delta \leq \infty)$, i.e. :

$$(T_i^*, \Delta_i^*) \equiv (\min(b, T_i), \Delta_i\, I_{[T_i \leq b]}) \ : \quad 1 \leq i \leq n, \quad T_i \geq a$$

This is *left-truncated right-censored* dataset for which maximum likelihood estimator of hazard λ_a specializes (from Weibull formulas with $\gamma = 1$ fixed) to:

$$\sum_{i=1}^{n} I_{[a \leq T_i \leq b]}\, \Delta_i \,/\, \sum_{i=1}^{n} I_{[a \leq T_i]}\, \min(T_i - a, b - a)$$

or number of observed deaths in interval divided by total time on test within exposure-interval $[a, b]$.

RESULT: $\quad \hat{\lambda}_a \ = \ (N(b) - N(a)) \,/\, \int_a^b Y(t)\, dt$

since $\int_a^b Y(t)\, dt \ = \ \Sigma_{i=1}^n \ I_{[T_i \geq a]}\, \min(T_i - a, b - a)$

This says for very small δ that the instantaneous hazard rate $h(a)$ at a is generally estimated by

$$\hat{h}(a) \ = \ (N(a + \delta) - N(a)) \,/\, (\delta\, Y(a))$$

A Binomial Model for Observed Failures

Consider for fixed a, small $\delta > 0$:

$$\underbrace{\{i = 1, \ldots, n : \ \Delta_i = 1,\, T_i \in [a, a+\delta]\}}_{\text{count} \quad N(a+\delta) - N(a)} \subset \underbrace{\{i \leq n : \ \ T_i \geq a\}}_{\text{count} \quad Y(a)}$$

Conditional on 2nd set, each element i falls in the first set by an independent coin-toss with heads-probability

$$\begin{aligned} & P(a \leq X_1 \leq \min(a+\delta, C_1) \,|\, \min(X_1, C_1) \geq a) \\ & \quad \approx \ P(a \leq X_1 \leq a+\delta) \,|\, \min(X_1, C_1) \geq a) \\ & \qquad \text{by smallness of } \ \delta > 0 \\ & = \ h_X(a)\delta \ + \ o(\delta) \qquad\qquad\qquad \text{since } X_1,\, C_1 \text{ indep} \end{aligned}$$

$$\text{Thus} \qquad Y(a) \cdot h_X(a)\, \delta \quad \approx$$

$$E(N(a+\delta) - N(a) \,|\, \{\min(X_i, C_i, a),\ I_{[X_i \leq \min(C_i, a)]}\}_i)$$

or, letting $\mathcal{F}_a$ denote all $\{\ \}$ data observable up to a ,

$$E(N(a+\delta) - N(a) - Y(a)\, h_X(a)\, \delta \,|\, \mathcal{F}_a) \ = \ o(\delta)$$

or for all $t, s > 0$,

$$\begin{aligned} & E(N(t+s) - N(t) - \textstyle\int_t^{t+s} \ Y(a)\, h_X(a)\, da \,|\mathcal{F}_t) \\ & = \ E(\textstyle\int_t^{t+s} (dN(a) - Y(a) h_X(a) da) \,|\, \mathcal{F}_t) \ = \ 0 \end{aligned}$$

Compensated Counting-Process Martingale

We have just seen for the failure-counting and at-risk processes

$$N(t) \;=\; \sum_{i=1}^{n} \Delta_i \, I_{[T_i \le t]} \;, \qquad Y(t) \;=\; \sum_{i=1}^{n} I_{[T_i \ge t]}$$

based on iid survival data with indep. X_i, C_i,

$$M(t) \;=\; N(t) \;-\; \int_0^t Y(a)\, h_X(a)\, da$$

is a **martingale** in the sense that for $s > 0$,

$$E(M(t+s) - M(t) \,|\, N(u), Y(u),\; u \le t) \;=\; 0$$

We will see that many interesting statistics can be written either exactly or approximately (in large samples) as multiples (like $n^{-1/2}$) of *stochastic integrals*

$$T \;=\; \int g\, dM, \qquad g \quad \text{left-continuous}$$

The increments $\int_a^{a+\delta} g dM \approx g(a)(M(a+\delta) - M(a))$ have variance $\approx g^2(a)\, n\, E(Y(a))\, h(a)\, \delta$ and zero covariance when they cover disjoint intervals, and it can be shown *even if* g *is a random left-continuous function*

$$Var(T) \;=\; \int E\left(g^2(t)\, Y(t)\, h_X(t)\, dt\right) \;=\; E\left(\int g^2(t)\, dN(t)\right)$$

Nelson-Aalen Hazard Estimator

We saw that a hazard function $h = h_X$ known to be piecewise constant λ_a on interval $(a, a+\delta]$ has MLE

$$\hat{\lambda}_a \;=\; \hat{h}(a+) \;=\; (N(a+\delta) - N(a)) \,/\, \int_a^{a+\delta} Y(t)dt$$

If the intervals $[a, a+\delta]$ are small but unspecified, this suggests to 'estimate' $\hat{h}(s)ds \;=\; dN(s)/Y(s)$ which does not make sense as a density function but does in the cumulative *Nelson- Aalen estimator* form (Aalen 1975)

$$\hat{H}_X(t) \;=\; \int_0^t \frac{dN(s)}{Y(s)}$$

Recall that the increments of the process $M(t) = N(t) - \int_0^t Y(s)\, dH_X(s)$ have **expectation** 0. Note Nelson-Aalen is the *martingale estimator* which substitutes for the unknown H_X the estimator which makes

$$\hat{M}(t) \equiv N(t) - \int_0^t Y(s)\, d\hat{H}_X(s) \;=\; 0 \quad \text{for all} \quad t$$

Martingale property, hazard estimator, and formula for variance estimator can be understood more clearly in **cohort life-table** *formulation known to actuaries for at least 130 years.*

Discrete-Time Life Table

Consider time grouped into successive intervals of length δ (in SEER, $\delta = 1$ month). For each $0 \le k < t_{\max}/\delta$:

$$\begin{aligned} Y(k\delta) &= \#\text{alive at time } k\delta \\ N((k+1)\delta) - N(k\delta) &= \#\text{obs'd deaths } \in (k\delta, (k+1)\delta] \\ &\approx \text{Binom}(Y(k\delta),\, \delta\, h_X(k\delta)) \end{aligned}$$

$$\hat{H}_X(k\delta) \;=\; \sum_{j=0}^{k-1} \frac{N((j+1)\delta) - N(j\delta)}{Y(j\delta)}$$

(Cond'l) Variance of individual terms in last sum $\approx$

$$\begin{aligned} & Y(j\delta)^{-2}\, Y(j\delta)\, \delta\, h_X(j\delta)\, (1 - \delta h_X(j\delta)) \\ \approx\quad & (N((j+1)\delta) - N(j\delta))/Y^2(j\delta) \end{aligned}$$

So estimate variance of Nelson-Aalen estimator $\hat{H}_{NA}(t)$ by

$$\widehat{\text{Var}}(\hat{H}_{NA}) \;=\; \sum_{s \le t} \Delta N(s)/Y^2(s)$$

These martingale-related estimators are approximately normally distributed via Martingale CLT applied to

$$\hat{H}_{NA}(t) \;-\; H_X(t) \;=\; \int_0^t \frac{1}{Y(s)}\, dM(s)$$

Discrete-time Approximation, Continued

Standard form of variance-estimator used is the cumulative-hazard version of *Greenwood formula* (estimator of Kaplan-Meier Survival-Curve variance, below):

$$\widehat{\mathrm{Var}}_G(\hat{H}) \;=\; \sum_{s \le t} \frac{\Delta N(s)}{Y(s)(Y(s) - \Delta N(s))}$$

Both formulas correct in large samples and available in **Splus**, **R**.

These variance estimators for $\hat{H}$ used only conditional-variance terms, justified by general formula

$$\mathrm{Var}\left(\frac{N((j+1)\delta) - N(j\delta)}{Y(j\delta)}\right) =$$

$$E\left(\mathrm{Var}\left(\frac{N((j+1)\delta) - N(j\delta)}{Y(j\delta)}\right) \mid Y(j\delta)\right)$$

$$+\ \mathrm{Var}\left(E\left(\frac{N((j+1)\delta) - N(j\delta)}{Y(j\delta)}\right) \mid Y(j\delta)\right)$$

since conditional expectation inside last term is
$\approx\ \delta \cdot h(j\delta)$ which is not random and has variance 0.

Comparison of Nelson-Aalen vs Weibull Hazards

We now have methods of estimating hazard functions from parametric model (especially Weibull) and *non-parametrically* – without distributional assumption – by Nelson-Aalen. We can check how close they are using estimated variance formulas, on SEER data.

In 796-subject group of Male, Hodgkins, No-Radiation, First-Cancer patients Aged 24–29, plotted (on next page)

$$\hat{H}_{Weib}(t) \;=\; \hat{\lambda}\, t^{\hat{\gamma}} \quad , \qquad \hat{H}_{NA}(t) \;=\; \sum_{s \le t} \frac{\Delta N(s)}{Y(s)}$$

and also the pointwise 95% Nelson-Aalen confidence intervals

$$\sum_{s \le t} \frac{\Delta N(s)}{Y(s)} \;\pm\; 1.96 \left(\sum_{s \le t} \frac{\Delta N(s)}{Y(s)(Y(s) - \Delta N(s))} \right)^{1/2}$$

Despite large sample (243 observed deaths), only at around 250 months do Nelson-Aalen pointwise confidence intervals fail to contain the Weibull curve ! Formal testing would involve confidence **band** taking account of multiple-comparison aspect of looking at many different points, and would therefore not reject the Weibull model. **But** we should take into account that the Weibull hazards were fitted from the same data !

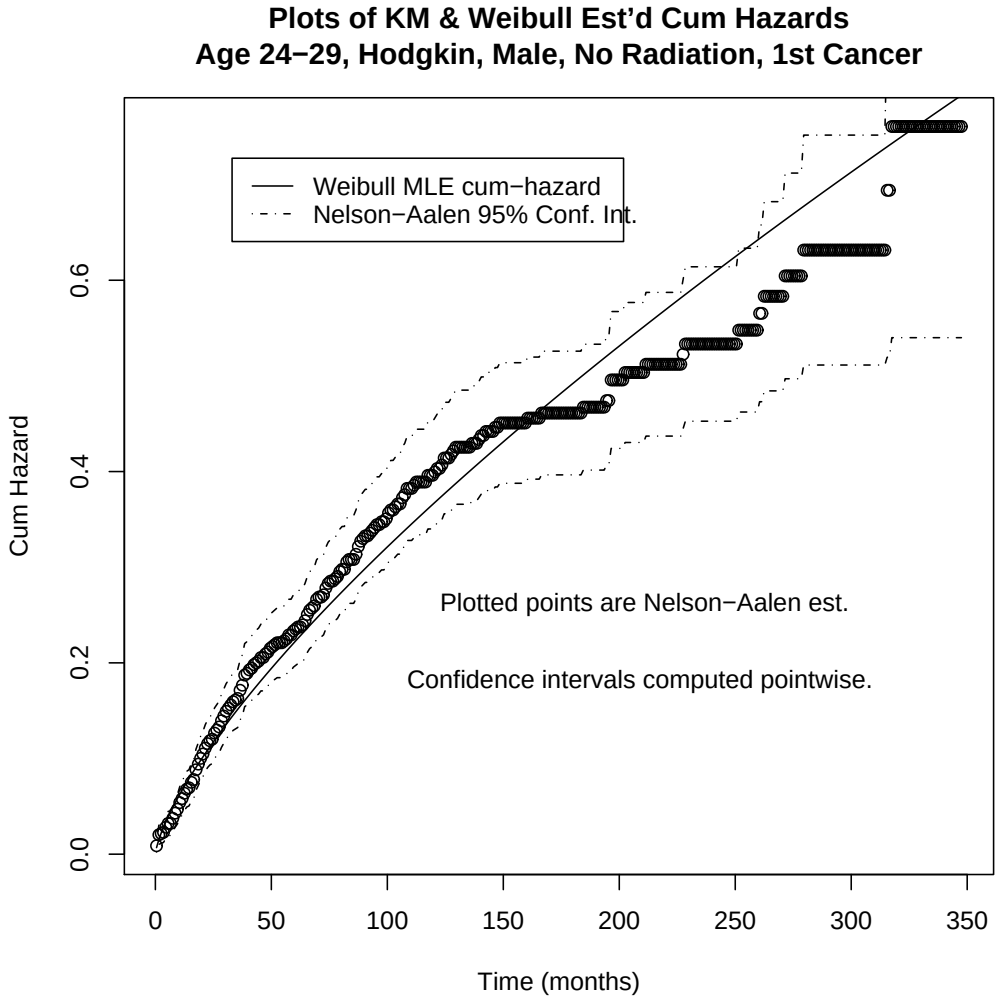

Figure 5: Comparisons of Weibull versus Nelson-Aalen estimated cumulative hazard functions for SEER data, Hodgkin Males.

Joint Sampling Distribution of $\hat{H}_{NA}(t)$, $\hat{H}_{\mathbf{Weib}}(t)$

We illustrate the use of *linearized estimators* and *martingale rep'n* to obtain correct variance for $\hat{H}_{NA}(t) - \hat{H}_{\mathrm{Weib}}(t)$, and to compare variances of the two estimators.

For Weibull estimator, we have from before, with $\vartheta_0 = (\lambda_0, \gamma_0)'$ the assumed true Weibull parameters,

$$\hat{\vartheta} - \vartheta_0 \;\approx\; V^{-1} \int \begin{pmatrix} 1/\lambda_0 \\ 1/\gamma_0 + \log(s) \end{pmatrix} dM(s)$$

where V^{-1} is the inverse of the Weibull information matrix and

$$M(t) = N(t) - \int^t Y(s)\, \lambda\gamma\, s^{\gamma-1}\, ds$$

Linearization refers to Taylor-expansion for fixed t, with respect to λ, γ variables:

$$\hat{\lambda}\, t^{\hat{\gamma}} \;-\; \lambda_0\, t^{\gamma_0} \;\approx t^{\gamma_0} \begin{pmatrix} 1 \\ \lambda_0 \log(t) \end{pmatrix}' (\hat{\vartheta} - \vartheta_0)$$

Thus

$$\hat{H}_{NA}(t) - H_X(t) \;=\; \int \frac{I_{[s\le t]}}{Y(s)}\, dM(s) \;\approx\; \int \frac{I_{[s\le t]}}{n\, S_X(s) S_C(s)}\, dM(s)$$

and $\hat{H}_{Weib}(t) - H_X(t) \;=$

$$t^{\gamma_0} \begin{pmatrix} 1 \\ \lambda_0 \log(t) \end{pmatrix}' V^{-1} \int \begin{pmatrix} 1/\lambda_0 \\ 1/\gamma_0 + \log(s) \end{pmatrix} dM(s)$$

Joint Sampling Distribution, cont'd

Since both centered estimators have martingale representation in the form $\int g_1\, dM$, $\int g_2\, dM$, Martingale CLT yields joint normal large-sample dist'n with means 0 and covariance matrix (and estimator)

$$E\left(\int \begin{pmatrix} g_1^2 & g_1\, g_2 \\ g_1\, g_2 & g_2^2 \end{pmatrix} dN\right) \approx \int \begin{pmatrix} g_1^2 & g_1\, g_2 \\ g_1\, g_2 & g_2^2 \end{pmatrix} dN$$

In our hazard-estimation example, the functions g_1, g_2 for $\hat{H}_{NA}(t) - H_X(t)$, $\hat{H}_{Weib}(t) - H_X(t)$ are respectively (for fixed t):

$$g_1(s) = I_{[s \le t]} \,/\, (n\, S_X(s) S_C(s)) \approx \frac{I_{[s \le t]}}{Y(s)}$$

$$g_2(s) = t^{\gamma_0} \begin{pmatrix} 1 \\ \lambda_0 \log(t) \end{pmatrix}' V^{-1} \begin{pmatrix} 1/\lambda_0 \\ 1/\gamma_0 + \log(s) \end{pmatrix}$$

We had estimators for the variances already, but the asymptotic (i.e., large-sample) covariance is estimated by

$$t^{\hat{\gamma}} \begin{pmatrix} 1 \\ \hat{\lambda} \log(t) \end{pmatrix}' \hat{V}^{-1} \int_0^t \frac{1}{Y(s)} \begin{pmatrix} 1/\lambda_0 \\ 1/\gamma_0 + \log(s) \end{pmatrix} dN(s)$$

SEER Example – Weibull vs Nonparam. Variances

In the picture (two slides earlier) from SEER Hodgkins patients (n=796), two questions remain:

- *Are the Weibull and Nelson-Aalen cumulative-hazard estimators still within range when we use the proper variance for their difference ?*
- *How much precision in CI's is lost by using generally valid nonparametric estimator in place of Weibull ?*

For first question: picture (not included) like previous picture with confidence intervals shows that when variability of estimation of Weibull is taken into account, Weibull always fall within 95% CI of Nelson-Aalen !

For both questions: plotted picture shows standard errors (square roots of variances) cumulative hazard curves from Nelson-Aalen and Weibull and from their difference. In this example, the standard errors of Nelson-Aalen estimator actually look smaller than the parametric Weibull estimators, but this is due in part to lack of fit !

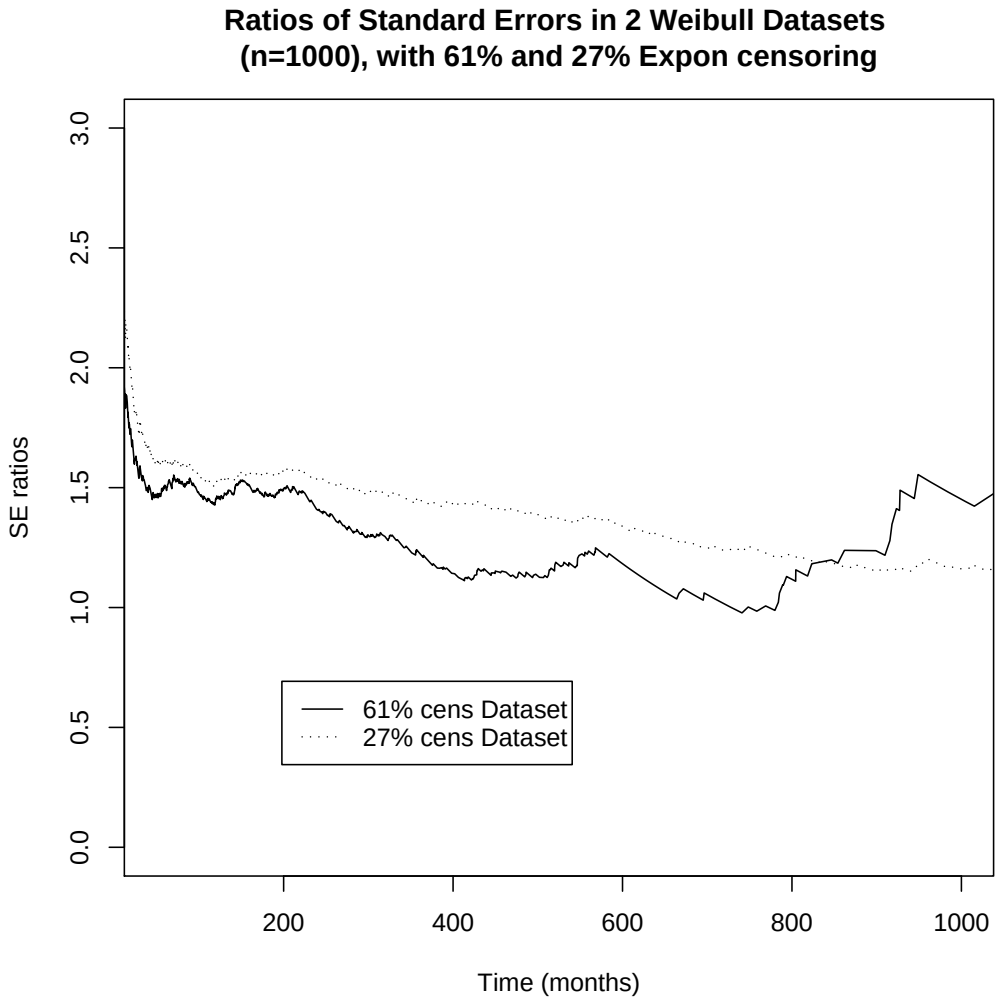

Figure 6: Ratio of standard errors, Nelson-Aalen to Weibull, in two artificial simulated datasets (Weibull survival, Expon censoring). Displayed percent censoring is the theoretical quantity $P(C_1 < X_1) = 1 - E(\Delta_1)$.

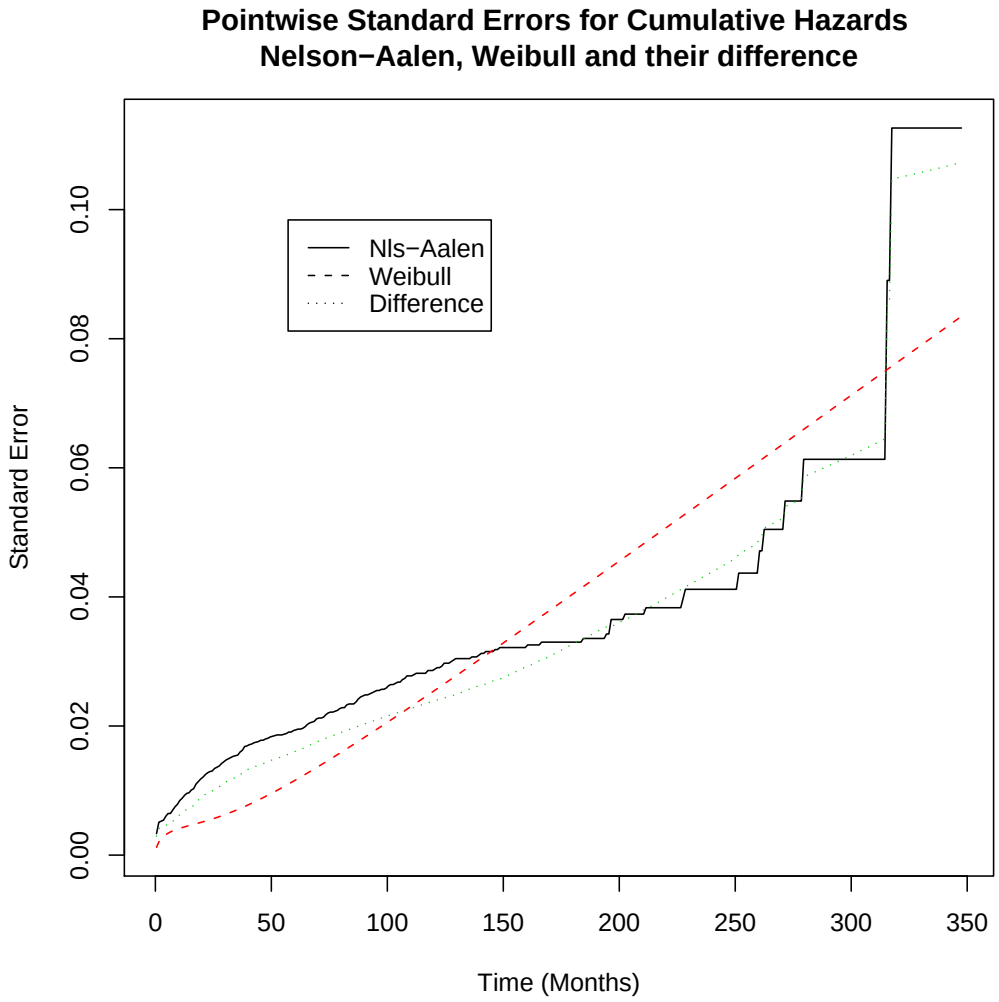

Figure 7: Poitnwise estimated standard errors from parametric Weibull, non-parametric Nelson-Aalen, and from the difference of the two. Same SEER data (n=796), Hodgkin Males aged 24-29.

Weibull vs Nelson-Aalen SE's

More generally, Nelson-Aalen variances may be quite a bit larger than parametric hazard-estimator variances ! (cf. Miller 1983) Some different examples :

In SEER data for Male Hodgkins First-cancer patients aged 30-35 receiving Radiation, 717 subjects with 166 obs'd deaths, Weibull fit is extremely good across the time-line (nearly linear, $\hat{\gamma} = .95$), but

$Q1 = 1.10, \quad \text{Med} = 1.28, \quad \text{Mean} = 1.30, \quad Q3 = 1.37$

Simulated two datasets of size 1000 from the same Weibull dist'n fitted in previous example, with independent exponential censoring times. In first dataset, 610 censored; in second, 266 censored. For both cases, picture shows ratio of Nelson-Aalen to Weibull cum-hazard estimator: on time-interval [50, 1000], the ratio is in range 1.3–1.5 for both datasets.

Such SE comparisons can also be made by theoretical asymptotic formulas, which is what Miller (1983) did.

Conclusion: precision can become much worse in nonparametric setting.

Kaplan-Meier Survival Estimator

$$\hat{S}_X(t) \equiv S^{KM}(t) = \prod_{0 \le s \le t} \left(1 - \frac{\Delta N(s)}{Y(s)}\right) \approx e^{-\hat{H}_{NA}(t)}$$

Relation between S^{KM}, $\hat{H}_{NA}$ follows because $Y(s)$ is of order n for most of the survival curve while $\Delta N(s)$ is much smaller, 0 or 1 in case of continuous distribution, so that

$$-\log(1 - \Delta N(s)/Y(s)) \approx \Delta N(s)/Y(s)$$

Note that all large-sample theory so far has assumed continuous survival distribution.

Recalling $S_X(t) = \exp(-H_X(t))$ and Taylor-expanding the exponential gives

$$S^{KM}(t) - S_X(t) \approx e^{-\hat{H}_{NA}(t)} - e^{-H_X(t)} \approx$$
$$-e^{-H_X(t)} \left(\hat{H}_{NA}(t) - H_X(t)\right) = -S_X(t)(\hat{H}_{NA}(t) - H_X(t))$$

So for $S_X(t)S_C(t) > 0$, $S^{KM}(t)$ is approx. normally distributed with mean 0 and variance estimated by

$$(S^{KM}(t))^2 \, \widehat{\mathrm{Var}}_G(\hat{H}_{NA}(t)) = (S^{KM}(t))^2 \sum_{s \le t} \frac{\Delta N(s)}{Y(s)(Y(s) - \Delta N(s))}$$

(*Greenwood formula* $\hat{\phi}_G(t)$)

References

The material just covered can be found in slightly different form in

C.-L. Chiang (1960) biostatistics book emphasizing the piecewise- constant hazards model, leading to so-called 'actuarial' hazard estimates.

Aalen (1975, *Ann. Statist.*) martingale approach to survival data, includes definition of 'Nelson-Aalen' estimator

Kaplan and Meier (1958, *JASA*) defines the 'product-limit' estimator of the discrete-time actuarial estimators known 80 years earlier, as the interval of discreteness shrinks to 0.

Breslow and Crowley (1974, *Ann. Statist.*) justify large-sample behavior of Nelson-Aalen and Kaplan-Meier estimators; re-done better, with martingale approach, by Gill (1983, *Ann. Statist.*)

Comparison of parametric vs nonparametric variances of survival curve estimators given by R. Miller (1983, "What price Kaplan-Meier?", *Biometrics*)

More on Kaplan-Meier Estimator

$$\hat{S}_X(t) \;\equiv\; S^{KM}(t) \;=\; \prod_{0\le s\le t}\left(1 \;-\; \frac{\Delta N(s)}{Y(s)}\right) \;\approx\; e^{-\hat{H}_{NA}(t)}$$

Kaplan-Meier is generalization to survival data of the (complementary) *empirical distribution function*, and has a martingale representation: if

$$\tau \equiv \max_i T_i \quad , \qquad M(t) = N(t) - \int_0^t Y(s) h_X(s)\, ds$$

then the exact integral formula

$$\frac{S^{KM}}{S}\left(\min(t,\tau)\right) - 1 \;=\; -\int_0^t \left(\frac{S^{KM}(s-)\, I_{[Y(s)>0]}}{S(s-)\, Y(s)}\right) dM(s)$$

shows that $S^{KM}(t)\, I_{[t\le\tau]}$ is unbiased for $S_X(t)\, I_{[t\le\tau]}$. The bias without the indicators is bounded by

$$P(\tau < t) \;=\; \left(P(T_1 < t)\right)^n \;=\; \left(\left(1 - S_X(t)\right)\left(1 - S_C(t)\right)\right)^n$$

Since $S^{KM}(s-)/S(s-)$ in integrand is close to 1 as long as $P(T_1 < t) > 0$, martingale formula shows

$$\mathrm{aVar}\left(\frac{S^{KM}(t)}{S(t)}\right) \;=\; E\int_0^t \frac{dN(s)}{Y^2(s)}$$

another way to get the Greenwood variance formula.

Interlude about R Software

Consider R functions to do the analyses described here:

`survfit` is a standard R function, and `WeibMLE` one I coded (available under RListings.txt file) to calculate MLE's for Weibull parameters. Another, more elaborate customized function to do the other analyses: `SurvEst0`.

```
> SrF1B <- survfit(Surv(Tim,Dth), data=Dtmp1)
## Choose SE option  error="greenwood" or "tsiatis"
## & ConfInt option  conf.type="plain" or "log"
> names(SrF1B)
[1] "n"         "time"     "n.risk" "n.event" "surv"
[6] "type"      "std.err" "upper"  "lower"   "conf.type"
[13]"conf.int" "call"

## For Nelson-Aalen, either: -log(SrF1$surv) , or
> NlsA <- cumsum(SrF1$n.event/SrF1$n.risk)

> WeibMLE   ## Weibull: no 0 death times allowed
function(tim, dth, rsk=NULL, lower=.2, upper=4)

> attach(Dtmp1)
  WeibMLE(Tim + .5, Dth, rsk=c(outer(Tim, Tim,
   function(x,y) x <= y) %*% rep(1,length(Tim))))
[1] 0.01823889 0.67431450
```

Estimating Median Survival Time

Since survival distributions are often very skewed, and data on extremely long survival times generally unavailable, *median* rather than *mean* is the scalar descriptive statistic of choice.

$$\begin{aligned} \text{med}(X) &= S_X^{-1}(.5) = \inf\{t : S_X(t) \le .5\} \\ &= H_X^{-1}(\ln 2) = \inf\{t : H_X(t) \ge \ln 2\} \end{aligned}$$

is estimated (essentially, equivalently) by

$$S^{KM^{-1}}(.5) \qquad \text{or} \qquad \hat{H}_{NA}^{-1}(\ln 2)$$

Here we choose the latter and call it $\hat{m}$.

For sampling behavior and CI, consider

$$\hat{m} > t \iff \hat{H}_{NA}(t) < \ln 2$$

from which it follows $\hat{m}$ is consistent for $m \equiv \text{med}(X)$. Also can show: $H_X(m) = \ln 2 = \hat{H}_{NA}(\hat{m}) + \mathcal{O}(1/n)$ and

$$\sqrt{n}\left\{(\hat{H}_{NA}(\hat{m}) - H_X(\hat{m})) - (\hat{H}_{NA}(m) - H_X(m))\right\} \xrightarrow{P} 0$$

which implies

$$\left(\hat{H}_{NA}(\hat{m}) - \hat{H}_{NA}(m)\right) / \widehat{\text{Var}}_G^{1/2}(\hat{H}_{NA}(t))\,|_{t=\hat{m}} \approx \mathcal{N}(0,1)$$

Alternative Median CI's

The final equation on the previous slide implies, using notation $s_G(\hat{m})$ for the standard error in the denominator, with prob. $1-\alpha$,

$$\hat{H}_{NA}(\hat{m}) - z_{\alpha/2}\, s_G(\hat{m}) \;\le\; \hat{H}_{NA}(m) \;\le\; \hat{H}_{NA}(\hat{m}) + z_{\alpha/2}\, s_G(\hat{m})$$

Applying $\hat{H}_{NA}^{-1}$ throughout the last inequalities turns out to leave the asymptotic probabilities unchanged, yielding the CI:

$$m \;\in\; \hat{H}_{NA}^{-1}\left(\ln 2 \;\pm\; s_G(\hat{m})\right)$$

This is one of a set of competing confidence intervals compared in Slud, Byar, & Green (Biometrics 1984): another is the *test-based* interval of Brookmeyer & Crowley (1982):

$$\left\{t \;:\quad |S^{KM}(t) \;-\; .5| \;\le\; z_{\alpha/2}\,(\hat{\phi}_G(t))^{1/2}\right\}$$

More accurate coverage is given by corrected CI's based on Edgeworth expansions, see Strawderman and Wells (1997).

SEER Examples of Medians, CI's

SEER Hodgkins Lymphoma patients, ≥ 60, Unknown Stage, receiving Radiation. Numbers of patients:

```
              Sex=  F   M
AgeCat= 60-69   81 115
        >=70   119 108
```

These categories chosen for larger hazard rates: unlike other categories, among these older patients almost none survived, ie almost none censored.
Median and Confidence Intervals (LCL,UCL) given by:

AgeCat	Sex	Median	LCL	UCL
60-69	F	68.5	46.5	84.5
	M	49.5	25.5	61.5
≥ 70	F	22.5	14.5	40.5
	M	20.5	12.5	26.5

These intervals are somewhat wide, because moderate sample size goes with respective standard-errors .112, .096, .096, .104 for cumulative hazard at medians.

Competing Risks

In the LATENT FAILURE Competing Risks Model, $S_X(t)$ is interpreted as survival probability if removals due to C_i were suppressed.

This makes clearer sense if C_i is administrative censoring and (entry-time) E_i is unrelated to health than if C_i is due to death from another cause. In latter case, called *Competing Risks*, death-variable X_i following C_i is **counterfactual**.)

No data on deaths following removals !

In case X_i, C_i independent, we saw

$$\frac{N(t+\delta) - N(t)}{Y(t)} \underset{\approx}{P} \frac{n\, P(t < X_1 \leq t+\delta)\, S_C(t)}{n\, S_X(t)\, S_C(t)}$$

which for small δ, $\approx \delta\, h_X(t) \approx H_X(t+\delta) - H_X(t)$.

Note: dependence could arise: (i) because of common dependence of X_i, C_i on underlying medical covariates, or (ii) in an administrative-censoring setting because patients entering at different times E_i (implying censoring times $A - E_i$) have different prognosis (eg 'sicker patients enter later').

Tsiatis (1975) showed $S_X(t)$ not determined by right-censored survival data under dependent censoring !

Dependent Censoring, cont'd

WHAT IF X_i, C_i ARE DEPENDENT ?

Depends on the unknowable counterfactual hazards:

$$\lim_{\delta \to 0} \frac{1}{\delta} P(X \in (t, t+\delta) \,|\, C = s) \qquad \text{for} \quad s < t$$

ANYWAY:

$$S^{KM}(t) \longrightarrow \exp\left(\int_0^t \frac{1}{S_T(s)} \frac{d}{ds} P(T_1 > s,\ \Delta_1 = 1)\, ds \right)$$

Peterson (1976) bounds $S_X(t)$ above and below in terms of identifiable 'subdistribution' functions
$P(T_1 \leq t,\ \Delta = j),\ j = 0, 1.$

Note: could try to proceed with parametric joint- distribution assumption under which S_X could be estimated. BUT assumptions about $f_{T,\Delta}$ can be tested from large datasets using $n^{-1} \sum_{i=1}^{n} I_{[T_i > t,\, \Delta = j]}$, but assumptions about $f_{X,C}$ cannot !

Approach of Slud and Rubinstein (1983) was to find qualitative assumption just enough to render S_X identifiable.

Slud and Rubinstein 1983 bounds on S_X

With dependent-censored data, $S_X(t)$ involves identifiable subsurvival functions plus probabilities like $P(t < X < t+\delta \,|C = s)$, $s < t$. Define

$$\rho(t) = \lim_{\delta \to 0} \frac{P(X < t+\delta \,|\, C < t,\, X > t)}{P(X < t+\delta \,|\, T > t)}$$

Cases.

$\rho \equiv 1$: includes independence of X, C
Kaplan-Meier estimator consistent.

$\rho \approx 0$: minimal, $S_X(t) \approx P(\Delta = 0 \cup T > t)$
censored never die .

$\rho \approx \infty$: maximal, $S_X(t) \approx S_T(t)$
death just after censoring.

Outcome: Each ρ leads to well-defined estimator $\hat{S}_\rho$ generalizing KM; bounds $r_1 < \rho(\cdot) < r_2$ give (consistently estimated) bounds on S_X .

SEER example (Hodgkins, M 24-29, No Rad, 1st Cancer) plotted KM survival, confidence bands, along with series of $\hat{S}_\rho$ curve estimators. **NB:** this is all-cause mortality: cause-of-death given in SEER but not always conceptually clear.

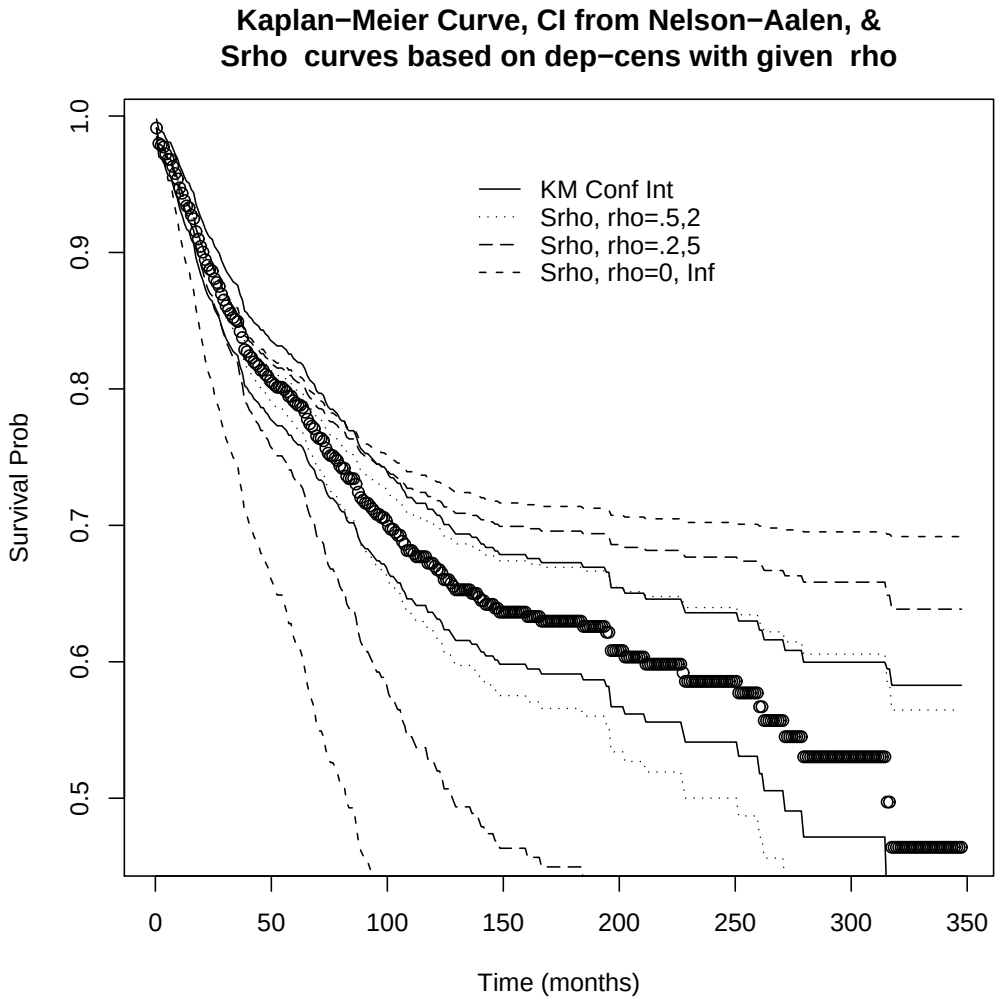

Figure 8: Kaplan-Meier curve, confidence interval, and several $\hat{S}_\rho$ curves in SEER (Hodgkins, M, n=796, 243 observed deaths) data-illustration. Extreme curves (short-dashed, $\rho = 0, \infty$) are the Peterson (1976) bounds.

Two-group Survival Testing – Motivation

Consider problem of testing whether two groups have **same** *(marginal) survival distribution. Particularly important because this is a regulatory question for formal hypothesis testing in randomized clinical trials !*

We already have Nelson-Aalen & Kaplan-Meier **curve** estimators as tools, but want an overall test of equality, with power against one-sided alternatives like

$$H_A: \quad h(t|z=1) \;<\; h(t|z=0), \quad \text{all } \; t$$

or

$$H_B: \quad H(t|z=1) \;< H(t|z=0), \quad \text{all } \; t>0$$

Consider the situation just before & after a death:

Deaths	Others	Totals
$\Delta N^{(1)}(t)$		$Y^{(1)}(t)$ $Gp.\ 1$ at-risk
		$Y^{(0)}(t)$ $Gp.\ 0$ at-risk
$\Delta N(t)$	$Y(t+)$	$Y(t)$ at-risk

$$E(\Delta N^{(1)}(t)\,|\,\text{totals},\ \Delta N(t)>0) \;=\; \Delta N(t)\,\frac{Y^{(1)}(t)}{Y(t)}$$

Hypergeometric under same Survival in 2 Gps

Two-sample Rank Test Statistics

Idea is to compare increments of observed deaths $N^{(1)}(t+s) - N^{(1)}(t)$ in treatment Gp. 1 with numbers based on a pooled at-risk group $Y(t) = Y^{(1)}(t) + Y^{(0)}(t)$ and pooled hazard-estimate $dN(t)/Y(t)$. This is like *Observed minus Expected*.

Think of parameterizing Group differences between survival, leaving nuisance hazard function:

$$S_{X|Z}(t|z) \;=\; Q(S_0(t), z)$$

in many possible ways, e.g.:

$$(S_0(t))^{\exp(\vartheta z)} \qquad \textsc{Lehmann}$$

$$\frac{e^{\vartheta z}\, S_0(t)}{1 - S_0(t) + e^{\vartheta z}\, S_0(t)} \qquad \textsc{Logistic}$$

'Remove' nuisance survival from picture by finding (Peto & Peto 1972) LMP rank statistics

$$E\left(\frac{d}{d\vartheta} \,\text{logLik}\Big|_{\vartheta=0} \;\Big|\; \{\Delta_i,\, Y^{(k)}(T_i),\ \text{all } i,\, k\} \right)$$

Get expressions like

Weighted-Logrank Statistics

$$\int w(t) \left(dN^{(1)}(t) - \frac{Y^{(1)}(t)}{Y(t)} dN(t) \right)$$

Main Example: Logrank Statistic ($w(t) = 1$, Mantel 1963, Peto 1972)

Important Second Example: Gehan Modified-Wilcoxon ($w(t) = Y(t)$, Gehan 1965), sum of scores over pairs of obs $(T_i, \Delta_i, T_j, \Delta_j)$ from Groups 1,0:

$+1$ if $T_i < T_j,\ \Delta_i = 1, \qquad -1$ if $T_j < T_i,\ \Delta_j = 1$

Recall *martingale centering*,

$$\begin{aligned} = & \int w(t) \left(\frac{Y^{(0)}(t)}{Y(t)} (dN^{(1)}(t) - Y^{(1)}(t) h_0(t) dt) \right. \\ & - \left. \frac{Y^{(1)}(t)}{Y(t)} (dN^{(0)}(t) - Y^{(0)}(t) h_0(t) dt) \right) \end{aligned}$$

Variance based either on hypergeometric or on martingale related formula

$$\int w^2(t)\, \gamma(t)\, (1 - \gamma(t))\, dN(t) \quad \overset{\text{logrank}}{\approx} \quad \pi(1-\pi) N(\infty)$$

$\pi =$ is Gp 1 random-allocation fraction; $\gamma(t) \equiv \frac{Y^{(1)}(t)}{Y(t)}$.

Relative-Efficiency and Sample Size Formulas

Normalized Weighted logrank test statistic:

$$S_w \;=\; \frac{\Sigma_{i=1}^n\; w(T_i)\,\Delta_i\;\left(\Delta N^{(1)}(T_i) - \gamma(T_i)\right)}{\left[\Sigma_{i=1}^n\; w^2(T_i)\,\Delta_i\,\gamma(T_i)(1-\gamma(T_i))\right]^{1/2}}$$

Again using *martingale centering* for numerator,

$$\int\; w(t)\;\Big((1-\gamma(t))\,(dN^{(1)}(t) - Y^{(1)}(t)h_0(t)dt)$$
$$-\quad \gamma(t)\,(dN^{(0)}(t) - Y^{(0)}(t)h_0(t)dt)\Big)$$

find mean against local alternative $h_1(t) = h_0(t)\,e^{b(t)/\sqrt{n}}$

$$\approx\; \int\; w(t)\,\gamma(t)(1-\gamma(t))\,Y(t)\,h_0(t)\,\frac{b(t)}{\sqrt{n}}\,dt$$

So under local alternative, statistic is normal with var. 1, mean

$$\approx\; \frac{1}{\sqrt{n}}\;\int b\,w\,\gamma(1-\gamma)dN \Big/ \left[\int w^2\,\gamma(1-\gamma)dN\right]^{1/2}$$

Note: b and w may have different shapes!

w sometimes estimated eg as $(S^{KM})^\rho$ in `survdiff` function in **R**.

Power & Sample-Size, cont'd

Power for one-sided size $\alpha/2$ test using $S_w > z_{\alpha/2}$ is

$$\approx\ 1\ -\ \Phi\left(z_{\alpha/2}\ -\ \frac{\int b\, w\gamma(1-\gamma)dN}{[n\ \int w^2\,\gamma(1-\gamma)dN]^{1/2}}\right)$$

Concrete Logrank example: $w = 1,\ \gamma(\cdot) \approx \pi$. Then for one-sided size $\alpha/2$, and power of $1-\beta$ for this alternative $b/\sqrt{n} = \ln c$, yields:

$$-z_\beta = z_{\alpha/2}\ -\ \frac{(\ln c)\, 0.25\, N(\infty)}{\sqrt{0.25\, N(\infty)}}$$

or

$$N(\infty)\ =\ \frac{(z_{\alpha/2} + z_\beta)^2}{(\ln c)^2\,\pi(1-\pi)}$$

Special case: $c = 2,\ \alpha = .05,\ \beta = .9,\ \pi = .5\ \Rightarrow$ $N(\infty) = 88$.

But $c = 1.5$ changes this to $N(\infty) = 256$.

Choose statistic by shape of imagined log-hazard ratio for efficiency; logrank is by far the most frequent choice !

Combining groups for efficiency is limited by validity since censoring may be different in subgroups !

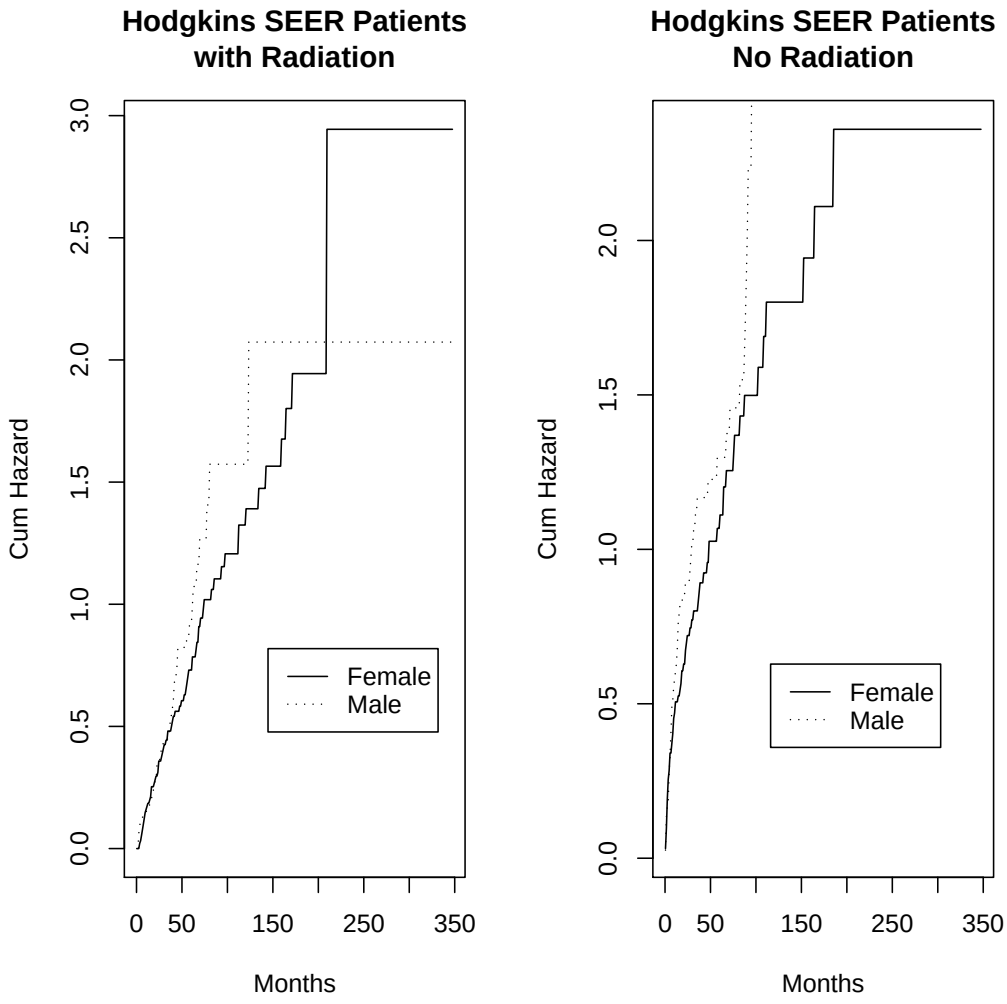

Figure 9: Male vs Female Cum-Haz Curves for SEER Hodgkins data, Rad=0 and 1 groups, illustrating group differences too small to test significant within each Radiation group.

Two-group testing examples:

- Groups of SEER White Stage 1 Hodgkins ≥ 70, F vs. M, with/without Radiation (group sizes 95 F 65 M with Rad, 96 F 76 M without). Pictures show survival differences by Sex.
- Logrank^2 are 1.665, 3.394 separately, 5.915 if groups combined. Purely quantitative question of reaching significance by adequate sample-size.
- With weight-function ("rho=1") $w(t) = S^{KM}(t)$, test statistic values (squared) become: .954, 1.394, 3.382: less power because down-weighting of later deaths! not a good idea according to pictures.
- Stratified statistics: 5.001; expect less power than simple lumping of groups because correcting for nuisance functions in both strata !

```
     ##  R statements
> survdiff(Surv(Tim,Dth) ~ Sex, data=Dtmp3,
     WBO==1 & Stag==1 & Rad==0 & AgeCat==7)$chisq
[1] 3.393823
> survdiff(Surv(Tim,Dth) ~ Sex + strata(Rad),
   data=Dtmp3, WBO==1 & Stag==1 & AgeCat==7)$chisq
[1] 5.001405
```

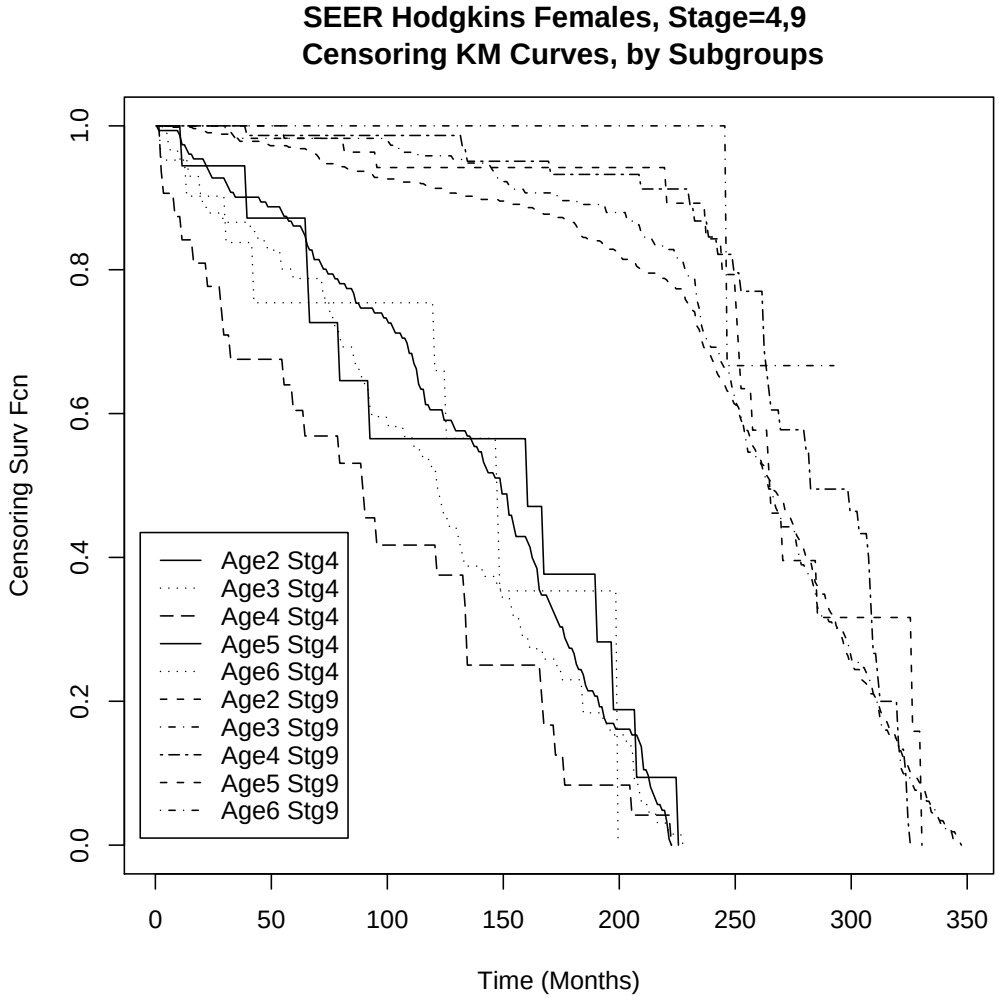

Figure 10: Variety of Kaplan-Meier curves for censoring in SEER White Hodgkins patients, across 5 adult Age-categories and 2 Stages (4 and 9).

Stratified Population Estimates

Imagine 2-gp clinical trials done at a number of distinct hospital centers (different experimental settings). Can assume treatment difference parameter ϑ goes always in same direction, but pop'ns including nuisance survival and censoring might be different !

Picture on previous slide illustrates different censoring distributions across different Stages (within each of Stages 4="Distant", 9="Unknown") censoring seems much the same across age category and sex.

In picture of population-wide KM curves on next slide: "Combination" means that group distinctions are ignored in estimating KM; "Stratification" means that a weighted combination of within-group KM's is taken.

Key issues in stratification are validity of combined model and/or sharing of parameters:

- Independence of death and censoring within group, needed for KM and other analyses, may not persist across groups.
- Nuisance parameters such as baseline hazards are not automatically shared across groups.
- Groups formed, by randomization and common protocol, to achieve homogeneity.

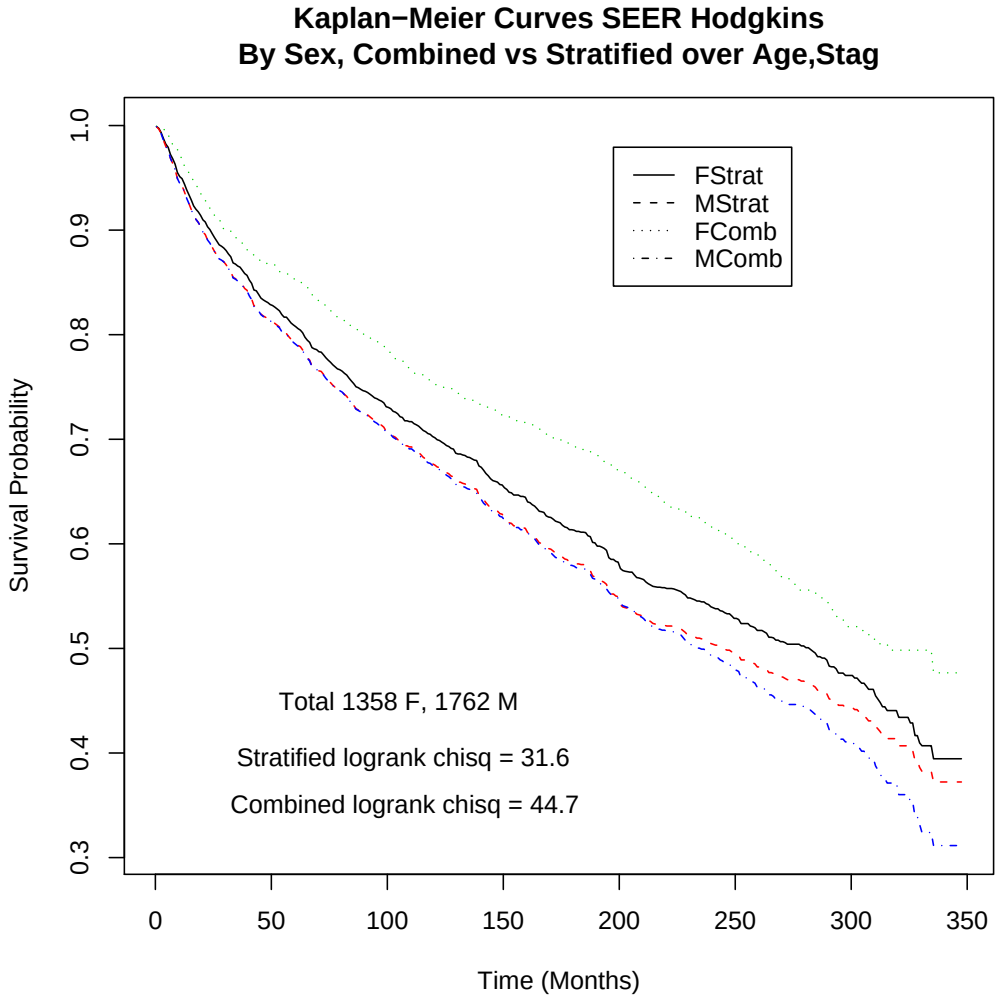

Figure 11: Kaplan-Meier curves for survival by Sex, for 3120 patients across Age and Stage groups, calculated by Combining groups (not legitimate!) versus Stratifying over them.

References

Additional material on Kaplan-Meier from Gill (1983).

Median estimators and confidence intervals based on right-censored survival data, as discussed in Slud, Byar and Green (1984), and Strawderman and Wells (1997).

Competing risks problem reviewed by Gail (1975). Actuaries call this topic "multiple decrement tables" to find e.g. probabilities of wife's survival probabilities following husband's death, for joint-insurance premiums.with nonidentifiability for dependent risks clarified by Tsiatis (1975) and givien in the form of (sharp) inequalities for permissible survival curves by A. Peterson (1976), with gradation between by completely unidentifiable function ρ described by Slud and Rubinstein (1983, *Biometrika*).

Two-sample test statistics, Peto & Peto (1972), Gehan (1965); in stochastic-integral terms, Aalen (1975); general weighted-logrank class, Tarone and Ware (1977).

Repeated significance tests based on two-sample rank statistics: Tsiatis 1982; Slud and Wei 1982; Slud 1984. Later developments in so-called *group-sequential* tests (many authors, beginning with Pocock 1977 and O'Brien and Fleming 1979 now being applied in randomized clinical trials.

Statistical Models for Causative Factors

Imagine a population cohort of individuals observed through a common window of time until either a discrete event ('endpoint') of interest occurs **or** the study ends.

Window may be defined by:

- chronological time-origin,
- individual time-origin e.g. entry into study, surgery,
- another time-scale, such as 'operational' time (reliability) or 'exposure' (epidemiology).

DATA:

(1) **explanatory** part: initial or *baseline* variables Z_i, including group-membership labels, together with time-varying measurements $V_i(t)$ (e.g., cumulative indicators of EKG anomalies or family disease history or blood pressure etc., and maybe age), plus

(2) **at-risk process** $Y_i(t)$ indicator $\{$ alive and under observation at time $t\}$; and

(3) **response** $N_i(t)$ cumulative count of observed events, such as 'death', 'recurrence of tumor', cumulative count of polyps, etc.

Response process initially 0, jumps only at times when $Y_i(t) = 1$.

Conditional Death Hazards

In general, recall **hazard intensity**

$$h_X(t) \equiv \lim_{\delta \to 0} \frac{1}{\delta} P(X \in (t, t+\delta) \,|\, X > t) = \frac{f_X(t)}{S_X(t)}$$

Then

$$h_X(t) = -\frac{d}{dt} \ln S_X(t) \implies S_X(t) = \exp\left(-\int_0^t h_X(s)\, ds\right)$$

In presence of time-varying information:

$$h_{X|W}(t) = \lim_{\delta \to 0} \frac{1}{\delta} P(X \in (t, t+\delta) \,|\, X > t,\, (W(s),\ s \le t))$$

If $W(\cdot)$ process is *not influenced* by X-occurrence, and the influence of W on X is only prospective, can extend the formula:

$$P(X > t \,|\, (W(s), s \ge 0)) = \exp\left(-\int_0^t h_{X|W}(s)\, ds\right)$$

Counting Processes & Martingales

Formalization of Intensity Model :

Def'n: a **counting process** $(N(t),\, t \geq 0)$ is a nondecreasing right-continuous integer-valued process s.t.

$$a.s. \quad \{\Delta N(t)\}_{t\geq 0} \;=\; \{0,\, 1\}$$

Say it is **compensated** by left-continuous increasing process $A(t)$ which is a function of (measurable wrt the *filtration of* σ-algebras generated up to times t by) baseline variables Z and left-continuous processes $(Y(s),\, V(s),\, s \leq t)$ if $N - A$ is a **martingale**, i.e., for all $\delta > 0,\, t \geq 0$

$$E(N(t+\delta) - N(t) - A(t+\delta) + A(t)\,|\,Z,\, (Y(s), V(s))_{s\leq t}) \;=\; 0$$

Interpretation: for small δ

$N(t+\delta) - N(t)$ = indicator of event in $(t, t+\delta)$

$A(t+\delta) - A(t) \approx$ probability known before t

So $N(t+\delta) - N(t) - (A(t+\delta) - A(t))$ is *Observed* minus (conditional) *Expected Count* on $(t, t+\delta)$ from vantage point of just before t.

Intensity Model: $A(t) \;=\; \int_0^t Y(s)\, g(Z,\, V(s))\, ds$

Innovations & Statistics

Innovation means *new independent piece of information* $N(t+\delta) - N(t) - (A(t+\delta) - A(t))$.

In real-data situations, we obtain innovations from each of a large set $N_i(t)$ of counting processes. In some applications, multiple events (e.g. multiple recurrences of nonlethal tumors, polyps, etc.) We search for non-chance *Observed – Expected* patterns among subsets of innovations (times, values i) defined through covariates Z_i, $V_i(t)$.

Limit Theorem: suppose that counting processes and associated predictors $(N_i(\cdot), Y_i(\cdot), V_i(\cdot), Z_i)$ are independent identically distributed across $i = 1, \ldots, n$ with intensities

$$A_i(t) \;=\; \int_0^t Y_i(s)\, g(Z_i,\, V_i(s))\, ds$$

Then for arbitrary sets B, C,

$$\int_0^t \sum_{i=1}^n I_{[Z_i \in B,\, V_i(s) \in C]}\, d(N_i - A_i)(s)$$

are asymptotically independent (for disjoint $B \times C$) normally-distributed variables with mean 0 and variance

$$\int_0^t \sum_{i=1}^n I_{[Z_i \in B,\, V_i(s) \in C]}\, dA_i(s)$$

PROBLEM: formulate, fit, test with data, a model for *prognosis* (probability or rate of event occurrence) as function of explanatory covariates.

Notations from before:

$$N_i(t) = \Delta_i \, I_{[T_i \le t]} \;\; , Y_i(t) \;=\; I_{[T_i \ge t]}$$

Now explicitly consider covariates in conditional prob's

$$P(T < t + \delta \,,\; \Delta = 1 \,|\, T > t,\, C > t,\, Z,\, V(s) :\; s \le t)$$

of essentially immediate death, modelling

$$\lim_{\delta \to 0+} \; \frac{1}{\delta} \, E(\, N(t+\delta) - N(t) \mid Z, (Y(s),\, V(s), s \le t)\,)$$

Multiplicative Intensity Model

Cox (1972), Aalen (1978) introduced the class of models showing effect of current detailed state

$$\begin{aligned} E(N(t+dt) - N(t) \,|\, Z,\, (Y(s), V(s) :\; s < t)) \\ = \; Y(t-) \, e^{\beta' Z + \gamma' V(t-)} \, \lambda(t) \, dt \end{aligned}$$

Parameters (β, γ) describe effect on prognosis of individual subjects, while **nuisance hazard function** $\lambda(t)$ describes the general background population. Exponent usable as *prognostic index*.

Cox Model Considerations

Multiplicative intensity models have the special feature: can analyze them by cancelling out effect of nuisance hazard. Combine notation $(Z_i, V_i(t))$ into $Z_i(t)$ (left-continuous). Note that conditionally given all data $\mathcal{F}_{t-}$ up to just before t, given also $\Delta N(t) = 1$:

$$P(N_i(t+dt) - N_i(t) = 1 \,|\, \mathcal{F}_{t-},\ \Delta N(t) = 1)$$

$$= \frac{e^{\beta' Z_i(t)}\, Y_i(t)\, \lambda(t) dt}{\sum_j\ Y_j(t)\, e^{\beta' Z_i(t)}\, \lambda(t) dt} \ = \ \frac{e^{\beta' Z_i(t)}\, Y_i(t)}{\sum_{j:\, Y_j(t)=1}\ e^{\beta' Z_i(t)}}$$

Fit β by maximizing **Cox Partial Likelihood**:

$$\log PL(\beta) \ \equiv \ \sum_{i=1}^{n}\ \Delta_i\ \log\left(\frac{e^{\beta' Z_i(T_i)}}{\sum_{j:\, Y_j(T_i)=1}\ e^{\beta' Z_i(T_i)}}\right)$$

Can also derive this by likelihood involving (β, Λ):

$$\log L(\beta, \Lambda) \ = \ \sum_i \int \left\{\log(e^{\beta' Z_i(t)} \Lambda'(t))\, dN_i(t) \ - \ Y_i(t) e^{\beta' Z_i(t)}\, d\Lambda(t)\right\}$$

after substituting Λ which is maximized for fixed β at:

$$\hat{\Lambda}(t) \ = \ \hat{\Lambda}_\beta(t) \ \equiv \ \int_0^t \left(\sum_{j:\, Y_i(t)=1} e^{\beta' Z_j(s)}\right)^{-1} dN(s)$$

Model-building: LR tests

In Cox regression as in other parametric modelling, comparisons of models in terms of significance of coefficients and log (Partial) likelihood ratio tests suggest when further model-building steps are needed. Large-sample theory of Partial Likelihood shows:

> If models 1 and 2 are *nested*, with regression parameter space of model 2 larger by d dimensions, then *under model 1*,
>
> $$2\left(\log \text{Lik}_2(\hat{\beta}) - \log \text{Lik}_1(\hat{\beta})\right) \sim \chi^2_d$$

SEER Hodgkins example – 16210 subjects

`Variables: Age Sex Rad Dth Stag Tim`

(Age continuous, Sex & Rad binary, Stag = 1, 2, 4 or 9.)

Model	Variables	logPL	Dim
modfit1	Age, Sex, Rad	-53452.2	3
modfit2	+ Stag	-53240.3	6
modfit3	+ Age:Stag	-53229.8	9
modfit4	modfit2 + all interactions by Sex	-53236.5	11
modfit5	modfit2+sqrt(Age)	-53203.7	7

Cox Models & Stratification

Consider the issue of F vs M comparisons in presence of other covariates (Age, Stag, Rad, + interactions). Could **either** fit single model, with one nuisance hazard function and a Sex coefficient (and possibly Sex-interaction terms with other covariates **or** fit combined *stratified* model (enforce same β coefficients but separate nuisance hazards and logPL contributions) **or** fit models to each Sex (separate coefficients and nuisance hazard).

(Step 1) `modfit4` above (with sex-interactions) hardly better than `modfit2` without, suggesting that when nuisance hazard is the same, the effect of Sex is adequately covered by single coeff.

(Step 2) Stratified model

`> modfit2S <- coxph(Surv(Tim, Dth) ~ Age + Rad + Stag + strata(Sex), data=Dcox4)` has coeff's essentially the same as unstratified model !

(Step3) Two separate models have two nuisance hazard functions, just like the stratified model, but now an extra set of β coefficients.

$$2\left(\log \mathrm{Lik}_{Two} - \log \mathrm{Lik}_{Strat}\right) = 2(-49101.5+49104.4)$$

Value = 5.8, not large for 6 df !

Summary Survival Curves

If Cox model has been fitted, with MPLE parameter estimates $\hat{\beta}$ and estimated **baseline hazard**

$$\hat{\Lambda}(t) = \hat{\Lambda}_\beta(t) = \int_0^t \left(\sum_{i=1}^n Y_i(s) e^{\beta' Z_i(s)} \right)^{-1} dN(s)$$

then predicted/estimated survival function for members of a baseline-covariate-defined group G is

$$\frac{1}{\#(G)} \sum_{i \in G} exp\Big(- \int_0^t e^{\beta' Z_i(s)} \, d\hat{\Lambda}(s) \Big)$$

and if all covariates are non-time-dependent

$$\hat{S}_G(t) \; = \; \frac{1}{\#(G)} \sum_{i \in G} \left(e^{-\hat{\Lambda}(t)} \right)^{\exp(\beta' Z_i)}$$

Now in the two following slides, we look at summary survival curves from comprehensive fitted Cox model `modfit2` across several covariate-defined groups G and compare them with the non-model-dependent groupwise KM curves.

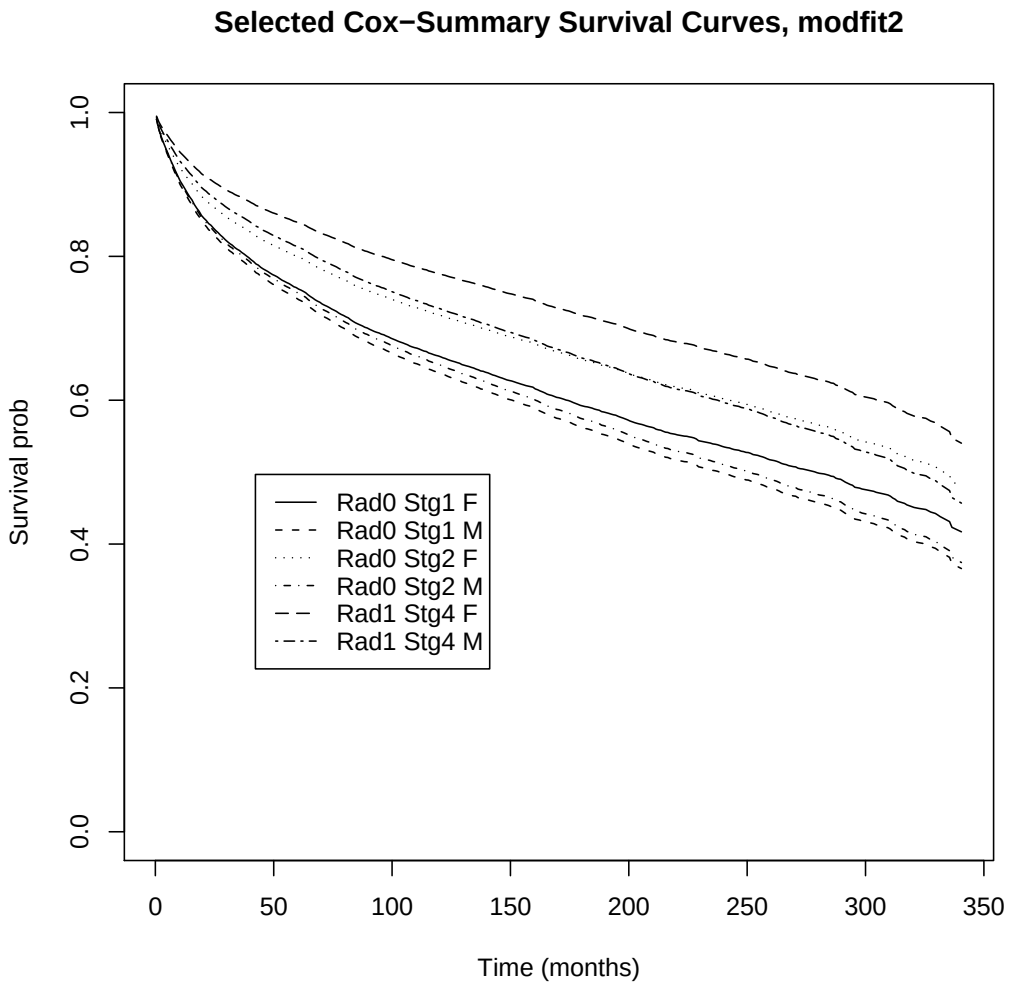

Figure 12: Cox-model summary survival curves fitted from `modfit2` described on slide, showing relation between curves from separate groups.

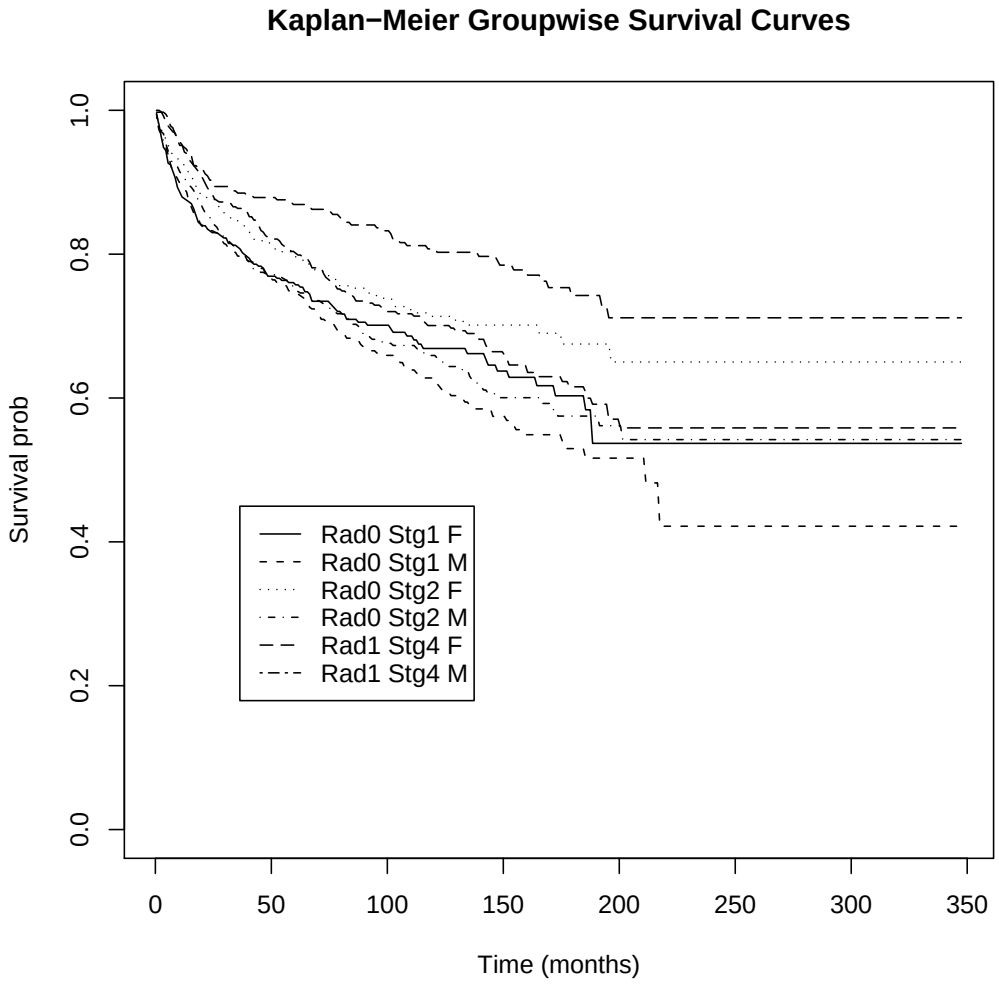

Figure 13: Kaplan-Meier curves fitted without common model on groups of respective sizes 477, 686, 682, 758, 379, and 512, to be compared with Cox-model summary survival curves in previous picture.

Model-Building with 'Residual' Plots

If a specified intensity model is *correct*, then individual terms $N_i(t+\delta) - N_i(t)$ are coin-toss variables (from vantage point of $t-$), with heads-probability $A_i(t+\delta) - A_i(t)$ and variance

$$(A_i(t+\delta)-A_i(t))\,(1-A_i(t+\delta)+A_i(t)) \;\approx\; A_i(t+\delta)-A_i(t)$$

uncorrelated across different $i,\ t$. Plotting cum-sums of

Martingale Residuals $\qquad N_i(\infty) - A_i(\infty)$

and i ordered with respect to some external variable can indicate whether that variable belongs in the Intensity !

Formally define residual over time interval $(a, b]$ and covariate cell C by

$$\sum_{i=1}^{n} \int_a^b I_{[Z_i(t)\in C]} \left(dN_i(t) \;-\; Y_i(t)\, \frac{e^{\hat\beta' Z_i(t)}}{\Sigma_{j:Y_j(t)=1}\; e^{\hat\beta' Z_j(t)}}\, dN(t) \right)$$

and refer them to pointwise variances

$$\mathrm{Var}(t) \;\approx\; \sum_{i=1}^{n} \int_a^b I_{[Z_i(t)\in C]}\, dN_i(t)$$

These residuals formed into time-sequence (or covariate-sequence) pictures give a very powerful tool for detecting lack of fit.

Goodness of Fit — Survival Regression

Note that we cannot expect comprehensive Cox-model to fit on such a large cross-classified dataset (n=16210) as the **SEER** Hodgkins data example above ! So we try some subsets ...

Consider first the full-dataset model `modfit2` with variables `Age, Sex, Stag, Rad`. Residuals in `modfit2$residual` are (in non-time-dependent model)

$$R_i \;=\; \Delta_i - \int \frac{Y_i(s)\, e^{\hat{\beta}' Z_i}}{\sum_{j:Y_j(s)=1} e^{\hat{\beta}' Z_j}}\, dN(s) \;=\; \Delta_i - \hat{\Lambda}(T_i)\, e^{\hat{\beta}' Z_i}$$

We already saw that adding a variable `sqrt(Age)` gave signficant improvement. To see that `Age` is a misspecified variable in `modfit2`, look at plot of residuals

$$a \quad \text{vs.} \quad \sum_{i:\texttt{Age}_i \le a} R_i \Big/ \left(\sum_{i:\texttt{Age}_i \le a} \Delta_i \right)^{1/2}$$

We do this in the next slide successively for the original model (n=16210), the original model with additional `sqrt(Age)` variable, and a restricted dataset (augmented model re-fitted only on cases in SEER registry 27, in Atlanta, with n=1092):

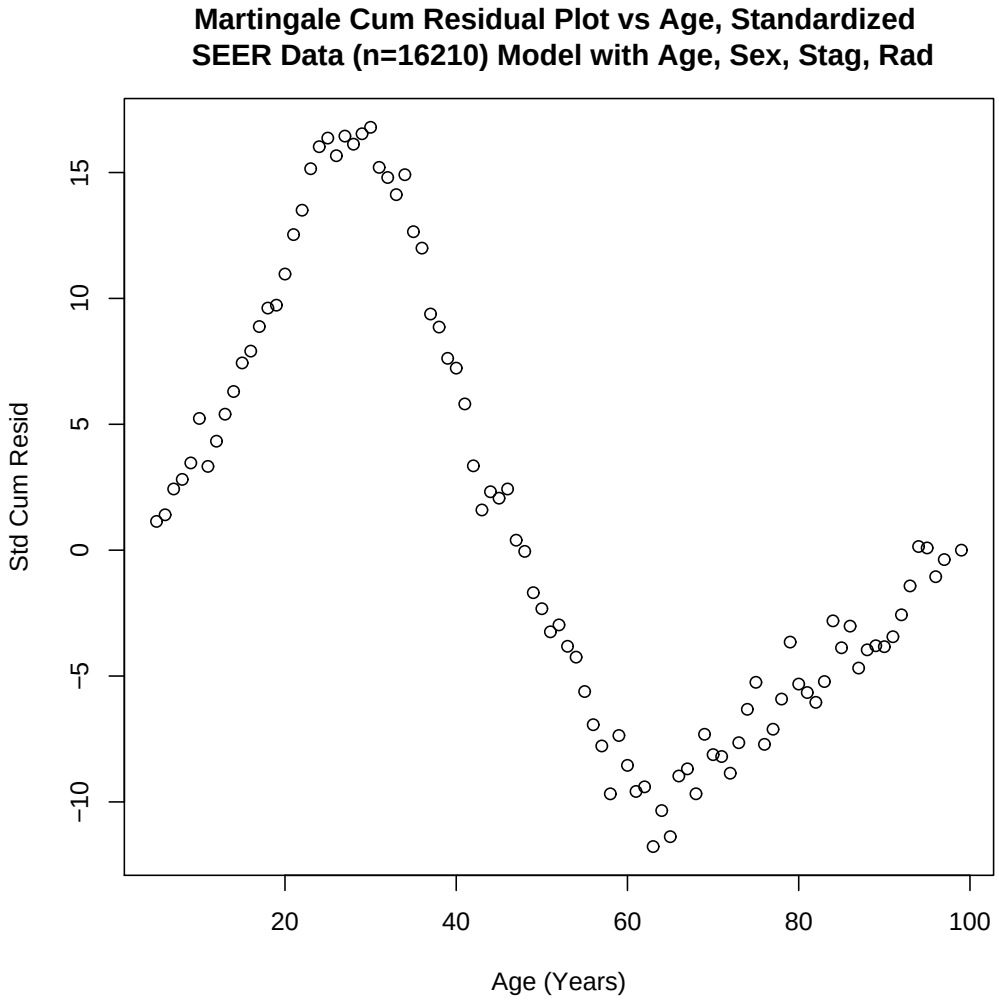

Figure 14: Martingale residual plot for `modfit2` Cox-model fit, with only `Age, Sex, Stag, Rad` variables, fitted to full (n=16210) Hodgkins SEER White dataset. Plot should be compared with standard normal pointwise if the model fits.

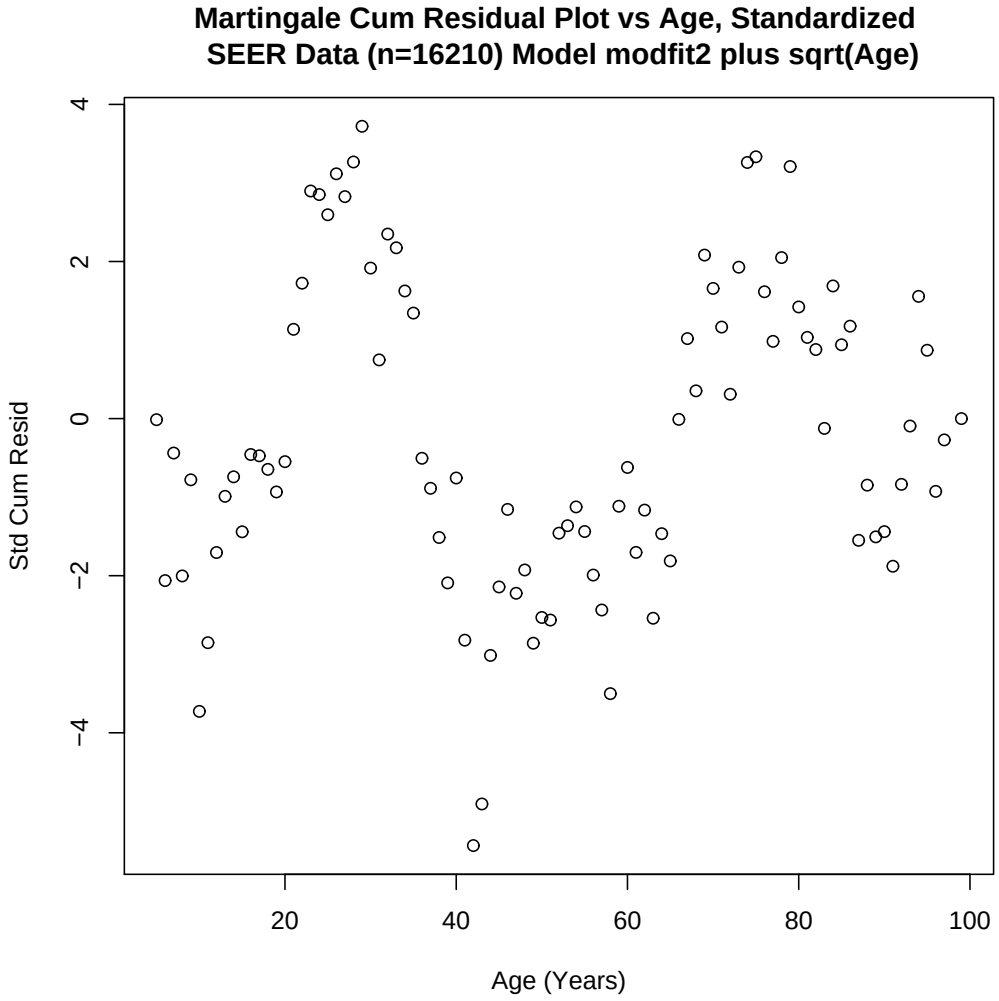

Figure 15: Martingale residual plot for `modfit2` Cox-model fit, augmented by `sqrt(Age)` variable, fitted to full (n=16210) Hodgkins SEER White dataset. Plot should be compared with standard normal pointwise if the model fits.

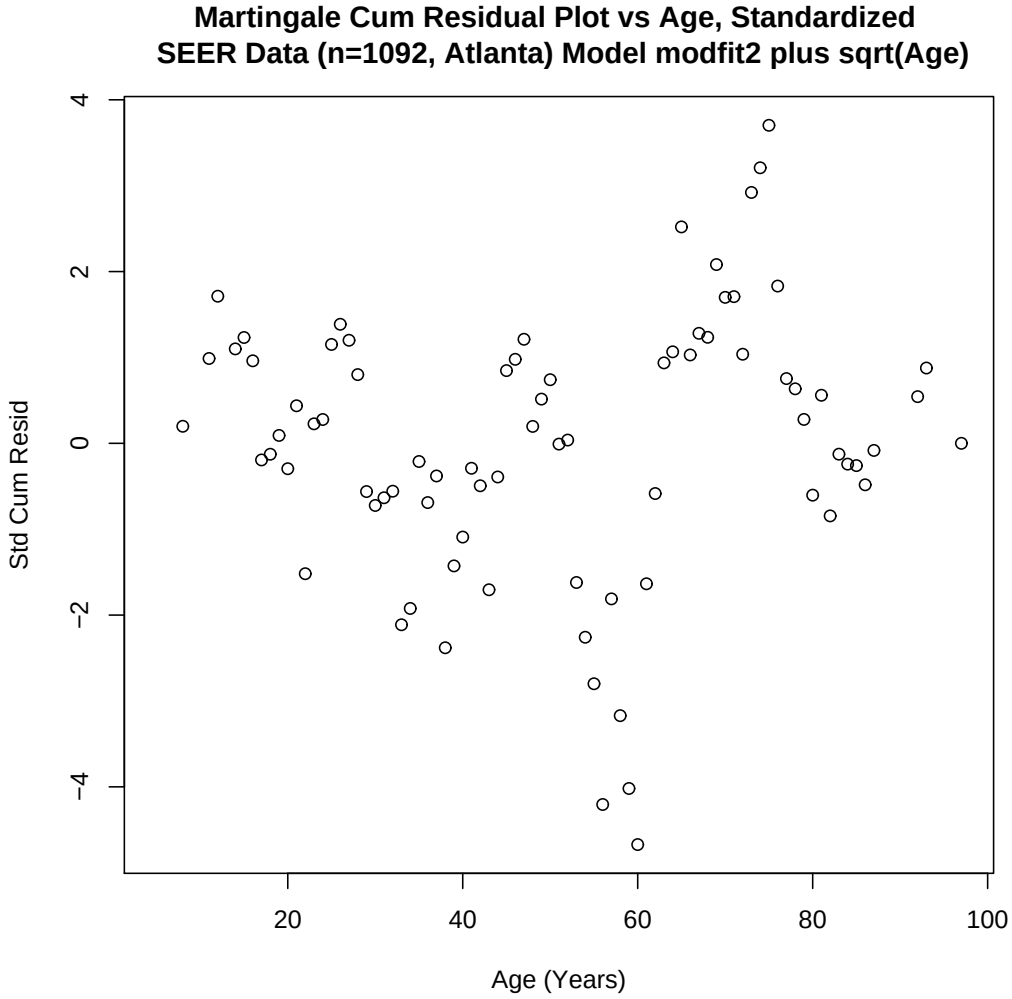

Figure 16: Martingale residual plot for `modfit2` Cox-model fit, augmented by `sqrt(Age)` variable, fitted to partial (n=1092, Atlanta only) Hodgkins SEER White dataset. Plot should be compared with standard normal pointwise if the model fits. Lack of fit is much less clear in smaller dataset.

Cox-Model Inadequacy: Further Display

Martingale residuals plots used most often to check for lack of fit of Cox models. In this setting, we saw that different data subsets seemed to follow Weibull models with different shape- parameters γ. This means among other things that different subjects have *nonproportional hazards*.

One way to check this: re-fit a model using TIME-DEPENDENT COVARIATE $V_i(t) = Z_i \log(t)$
("interaction" between $\log t$ and baseline covariates).

Another approach: refit so-called **frailty model** (described below).

Here we use pure diagnostic display to show Cox model inadequate: plot of cumulative martingale residuals plotted over time, and standardized.

Note: the very large standardized cumulative martingale residuals versus time suggest strongly that hazard ratios among subjects are time-varying.

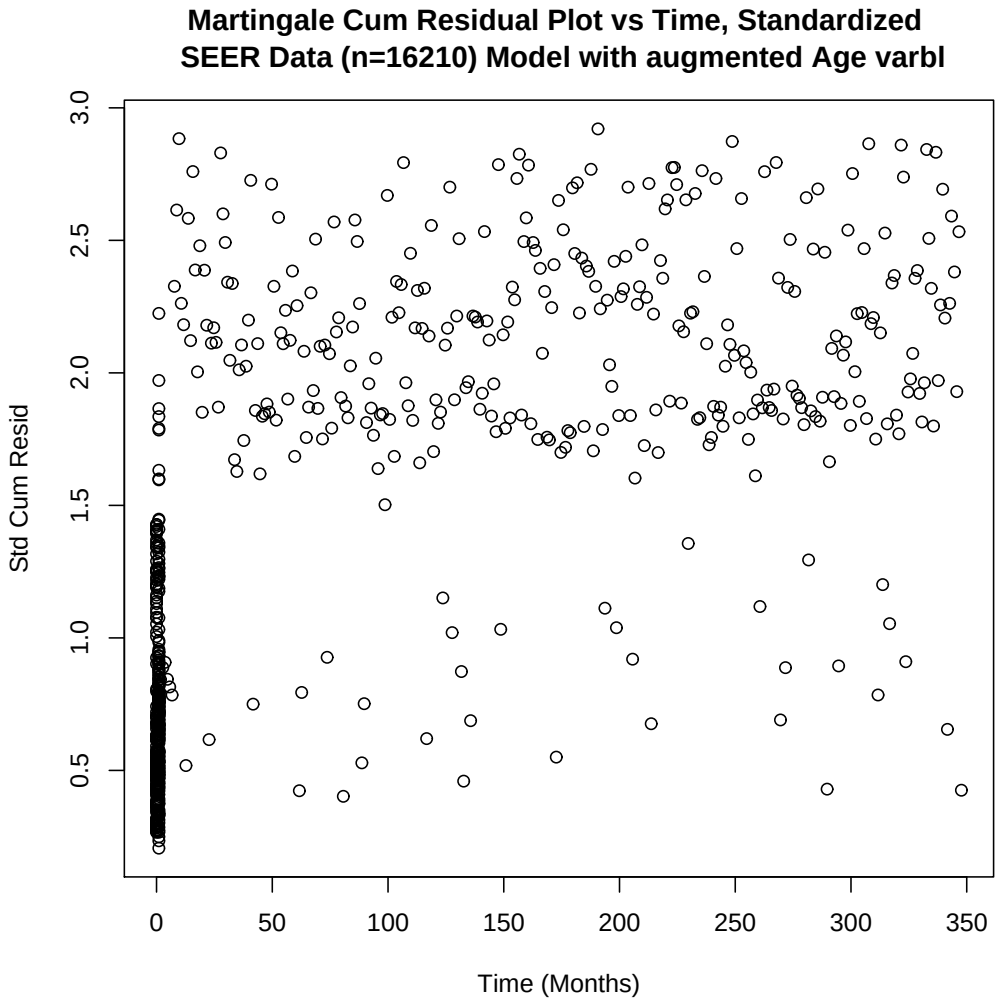

Figure 17: Martingale residual plot versus Time for `modfit2` Cox-model fit, augmented by `sqrt(Age)` variable, fitted to full (n=16210) Hodgkins SEER White dataset. Plot should be compared with standard normal pointwise if the model fits. **Note**: plot required random jittering of event-times to break ties, before model-fitting and residual calculation.

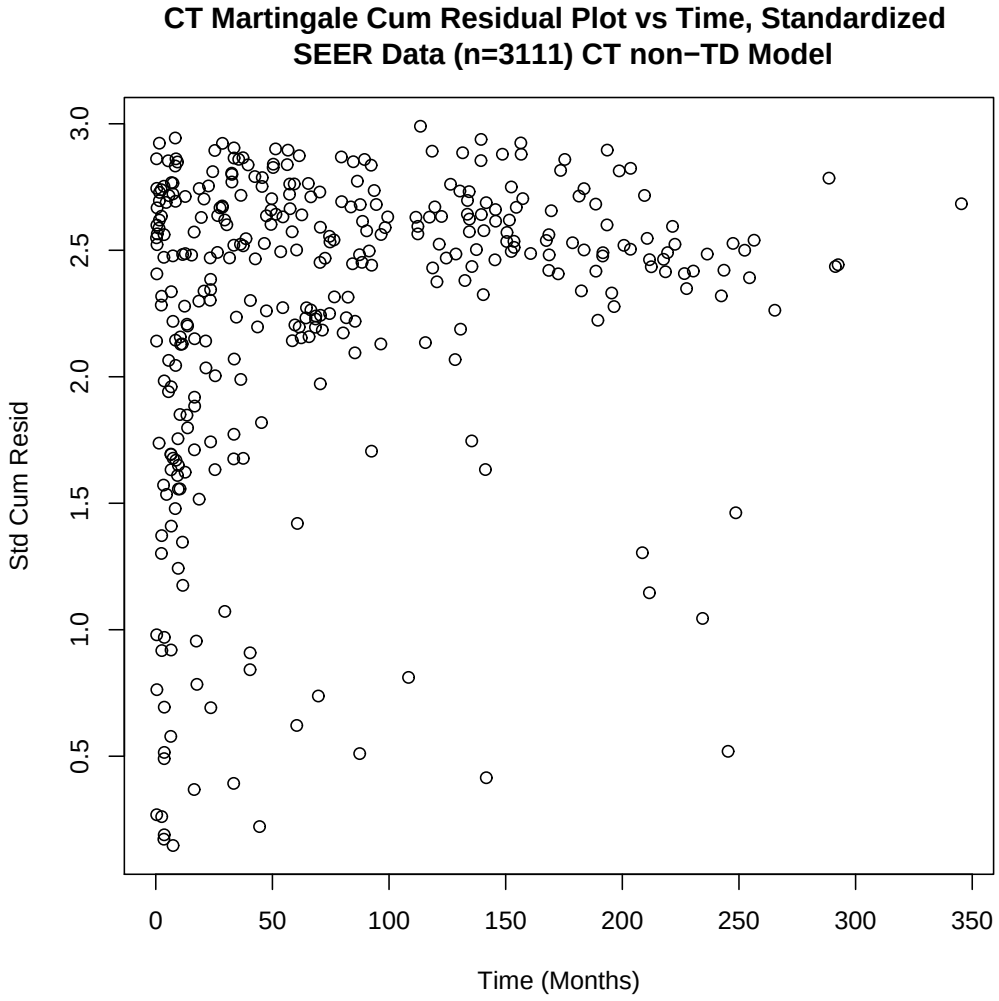

Figure 18: Martingale residual plot versus Time for `modfit2` Cox-model fit, augmented by `sqrt(Age)` variable, fitted to partial (n=3111, CT) Hodgkins SEER White dataset. Plot should be compared with standard normal pointwise if the model fits. **Note** correction was needed as in previous time-sequence residual plot to adjust handling of residuals with tied death-times. Note also that the lack of fit is still seen even on the smaller dataset.

Discussion of R Model-fitting Functions

We already saw the Cox model fitting syntax in **R**:

```
modfit5 <- coxph(Surv(Tim, Dth) ~ Age + Sex +
Rad + Stag + I(sqrt(Age)), data=Dcox4)
```

Interactions are written in formula as e.g. `Sex:Stag`, and stratifying variables (one or more) are given in form `strata(Sex, Reg)`. Quantities in output list:

```
round((rbind(modfit5$coef,
             sqrt(diag(modfit5$var))),4)
 Age SexM  Rad  Stag2 Stag4 Stag9 I(sqrt(Age))
.095 .291 -.373 .169  .503  .720    -0.631
.005 .026  .028 .053  .047  .042     0.072
```

martingale residuals given in `modfit5$residual`, logPL in `modfit5$loglik`. To plot summary survival curves: `plot(survfit(modfit5))`, with CI, but to work directly with baseline hazard,
`coxhaz <- basehaz(modfit5, FALSE)`.

`FALSE` *refers to changed centering of risk-factor exponents.* Output list has components `"hazard", "time", "strata"` for stratified model (with hazards and times concatenated), just "hazard" and "time" otherwise.

R Model-fitting, cont'd

Model fitting syntax above is for **non** time-dependent covariates in Cox model. Time-dependent covariate Cox model is also available, but with slightly different data structure: data-frame or list no longer has `time` and `status` columns, but instead `start, stop, event`.

Here view covariates `x` as piecewise constant on interval of 'exposure' $(a, b]$ when `start=a, stop=b`, and single subject i may appear in many such intervals, with `event` $= \Delta_i$ when $a < b = T_i$.

This is a counting-process data-structure introduced by Andersen and Gill (1982).

Syntax of model-fitting becomes:

```
coxph(Surv(start,stop,event) ~ x + strata(Reg),
data=Dfram)
```

For other R scripts, see attachment in Appendix along with listings of some special R functions described here.

Adjusted Logrank Test Statistics

Viewed two-sample **weighted logrank test** as a locally optimal score test with hazards parameterized as $\lambda(t)\, e^{\vartheta\, z\, b(t)}$ with z the treatment-group indicator. Approach also works when Cox-type model is fitted to adjust (eg, because of randomization imbalances) for the effect of other risk-factor covariates. *Treatment effectiveness* assessed through significance of treatment-gp coeff.

Easiest way to do this test is as **Wald test** (using Max Partial Likelihood estimated coefficient), standardizing by estimated standard error.

Example using previous data, testing for Sex differences in survival *in presence of other covariates*:

```
> modfit5 <- coxph(Surv(Tim, Dth) ~ Age +
     Sex + Rad + Stag + I(sqrt(Age)), data=Dcox4)
> modfit5$coef[2]/sqrt(modfit5$var[2,2])
[1] 11.00          #### wildly significant
> modfit5AT <- coxph(Surv(Tim, Dth) ~ Age + Sex + Rad +
     Stag + I(sqrt(Age)), data=Dcox4, subset= Reg==27)
> modfit5AT$coef[2]/sqrt(modfit5AT$var[2,2])
    4.13    #### highly significant normal deviate !!
```

Test can also be implemented as a score test (Tsiatis, Rosner and Tritchler 1985; Kong & Slud 1997 Biometrika).

Frailty Models

Cox models incorporating random effects are called *frailty* models: view survival conditionally given unobserved random 'covariate' ξ_i for subject i, as

$$S_{X|Z,\xi}(t|z,\xi) \;=\; \exp\left(-\,\xi\, e^{\beta' z}\Lambda(t)\right)$$

Here must assume the distribution of ξ from known parametric family, with mean fixed at one to avoid scale nonidentifiability in Λ.

Main example: 'Clayton-Cuzick' or 'semiparametric Pareto') model with $\xi \sim \text{Gamma}(1/\vartheta,\, 1/\vartheta)$: integrate out unobserved ξ to get

$$S_{X|Z}(t|z) \;=\; \int e^{-\xi y}\, dF_\xi(\xi)\,\Big|_{y=e^{\beta' z}\Lambda(t)} \;=\; e^{-G(e^{\beta' z}\Lambda(t))}$$

with $G(y) \;=\; \vartheta^{-1}\log(1+\vartheta y)$. (Contrast identity function for G in Cox-model, limiting case when $\vartheta \to 0$.)

Now-standard EM-based fitting idea integrates ξ_i out of `logPL` iteratively.

Additional References

General ref. for theoretical material up through 1993:

Andersen, P., Borgan, Ø., Gill, R., and Keiding, N. (1993) **Statistical Models based on Counting Processes.** Springer.

Key Cox-model theoretical papers are Cox, D.R. (1972, JRSSB), Andersen and Gill (1982, *Ann. Statist.*).

Martingale residuals in D. Schoenfeld (1982, *Biometrika*), discussed in Splus and R materials (incl. Springer books) by Therneau & Grambsch. Formal hypothesis tests of fit available but not the main point !

Frailty models discussed by Hougaard, P. in mid-80's, Clayton and Cuzick (1986, "gamma frailties"), G. Nielsen et al. (1992, *Scand. Jour. Statist.* including EM fitting algorithm implemented in later papers of Klein and others. Later theoretical papers by Murphy (1994, 1995, *Annals*), Parner (1998, *Annals*), Slud and Vonta (2004, *Scand. Jour.*) with different fitting method, current papers of Kosorok et al. (2004, *Annals*) and others.