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.

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

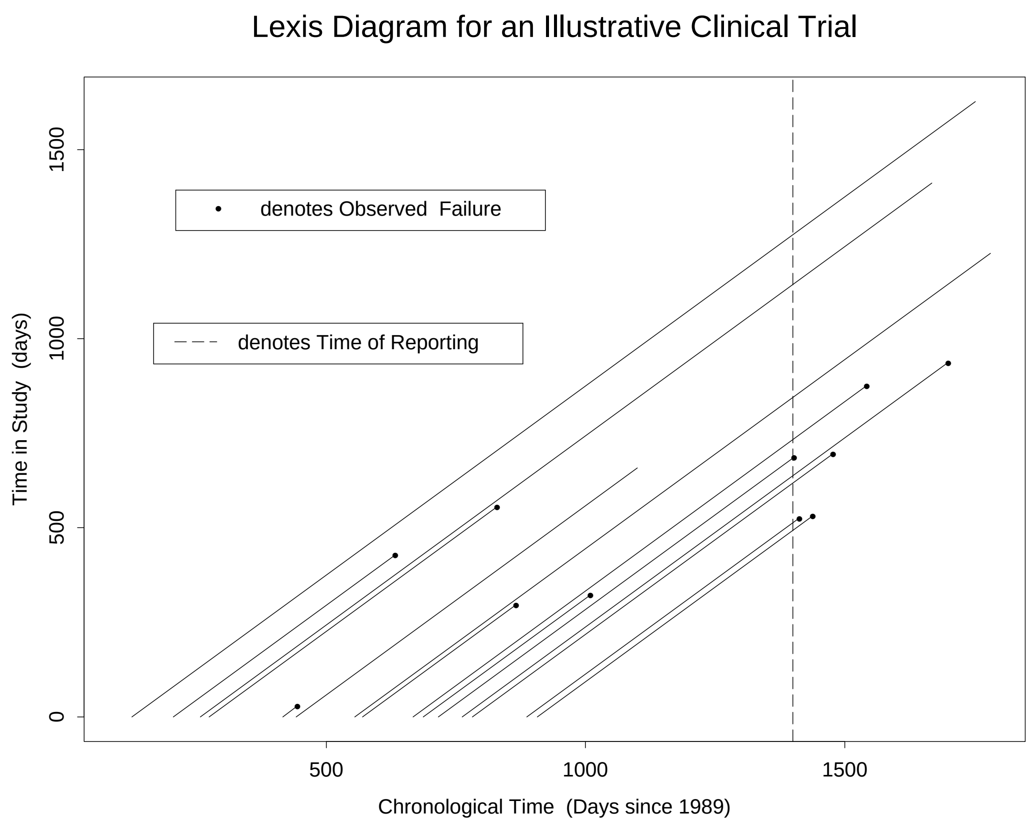

Figure 1: "Lexis Diagram" (from contributed article to Encyclopedia of Biostatistics): from entry, patients' followup is pictured as 45° line: solid dot represents death, line not ending in dot representes censoring.

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' $X \sim$ Expon(τ) , $\tau \sim G$ 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 $V \sim$ Expon(1) :

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

  $h(t) \; = \; A + B e^{ct} \qquad (A, B, c > 0)$

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

Pictures follow:

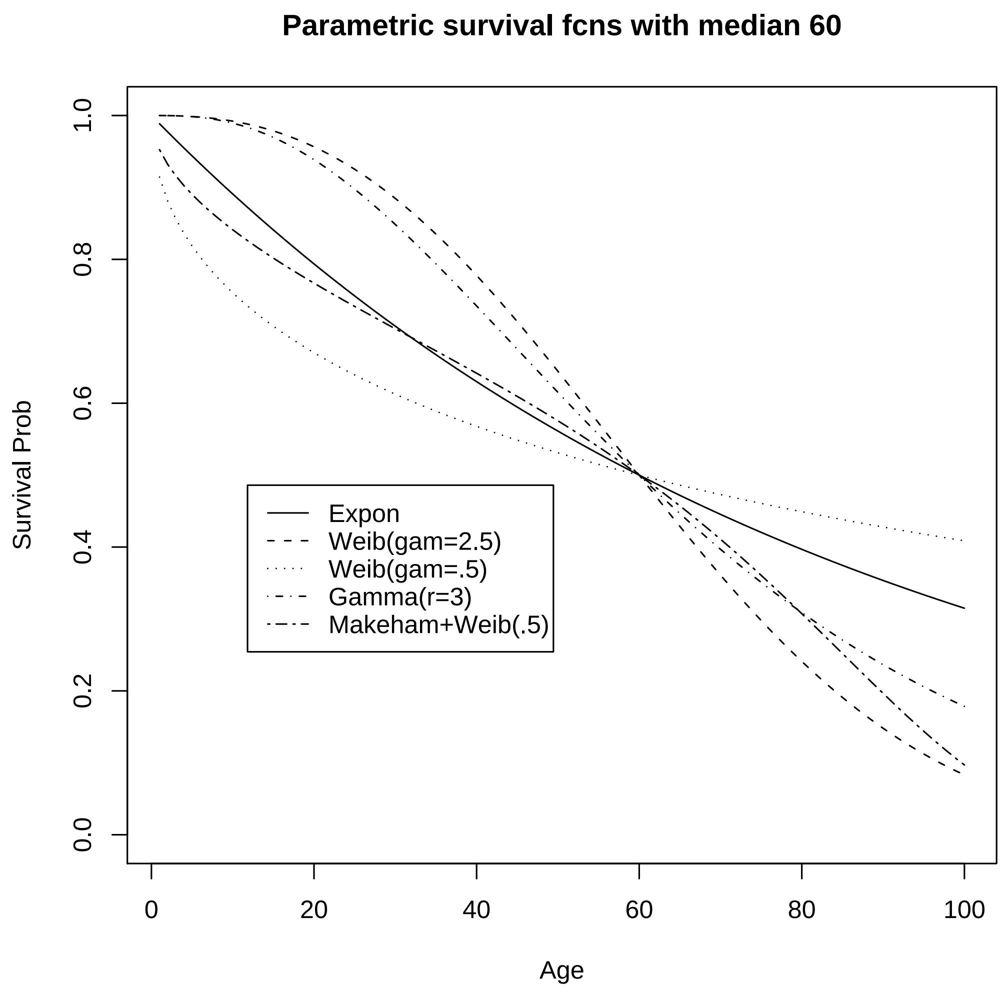

Figure 2: Graphs of survival functions from several parametric models designed to have common median 60.

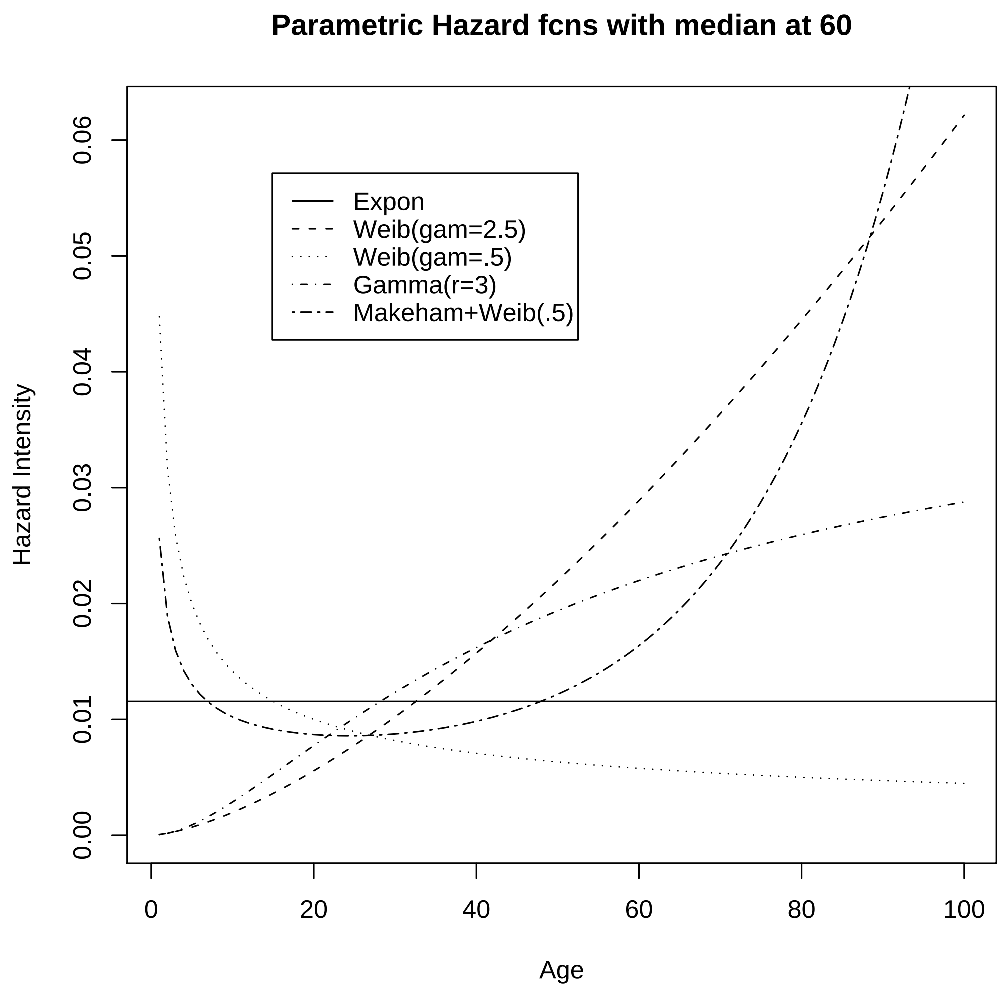

Figure 3: Graphs of cumulative hazard functions of several parametric models designed to have common median 60.