$\pi\mu\epsilon$ Talk , Wed. March 11, 2015, 3-4pm, MTH0101

Eric Slud, UMCP & Census Bureau

Mathematics and Careers Related to Waiting Times, in Insurance, Biomedical Statistics, and Reliability

1. Introduction to Waiting Times – Clinical Trials and Insurance

2. Introduction to Life Tables

3. Description of Related Research Areas and Careers

I. What are Duration or Waiting Time Data ?

With i indexing individual (subject)

$E_i =$ Entry time
$X_i =$ time from entry to failure (or other event of interest)
$C_i =$ time from entry until withdrawal ("loss to followup")

Statistical interest is in X_i, in questions like,
"if no failure has occurred by time t, what is the change there will still be no failure by time s+t" ?

Examples:

(a) $E = 0$ (birth), X follows some lifetime distribution,
$C =$ time loss to followup (emigration or end of study)

More Examples

(b) $E_i =$ calendar time of surgery to remove diagnosed tumor (e.g., breast cancer, colon cancer) from specific site

$X_i =$ post-surgery survival time
$C_i =$ end of clinical trial or of withdrawal from study

(c) Lifetimes may be measured from vital statistics (all-cause or cause-specific) or insurance portfolios: E_i may be time of arrival of insured life.

(d) "Time" in the case of devices may be not calendar time but "operational time" based on loading or stress.

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.

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

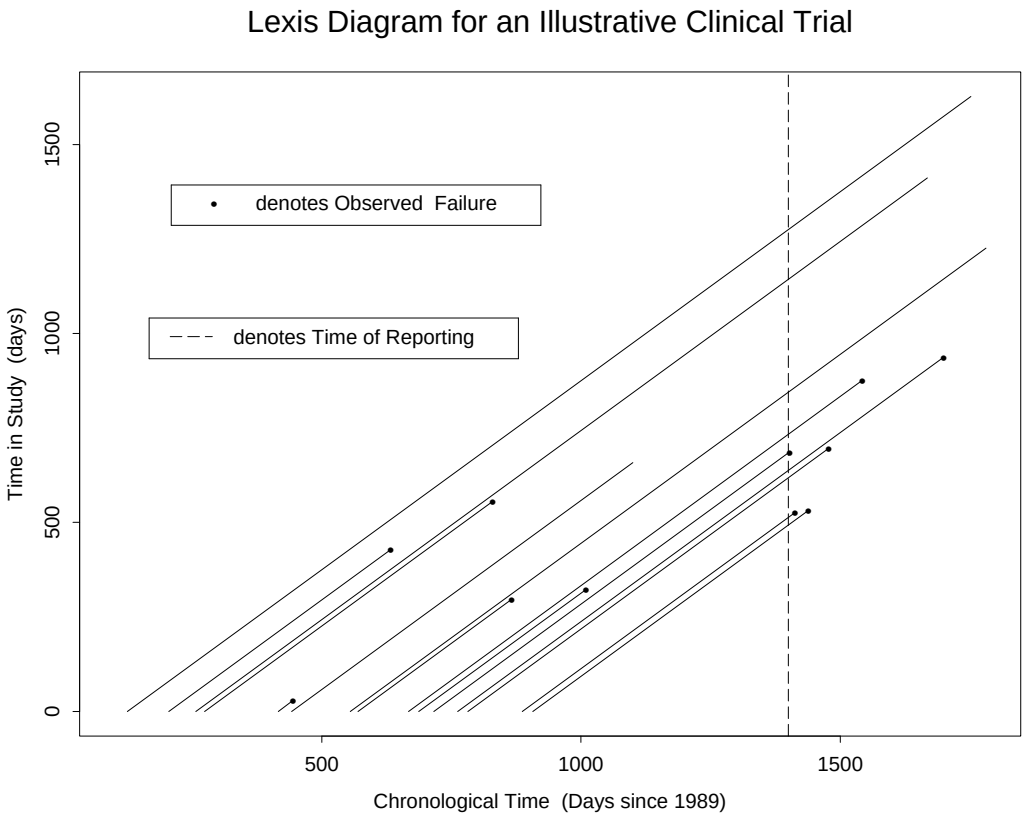

Figure 1: "Lexis Diagram" (from contributed article to Encyclopedia of Biostatistics): from entry, patients' followup is pictured as 45° line: solid dot represents death, line not ending in dot representes censoring.

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:

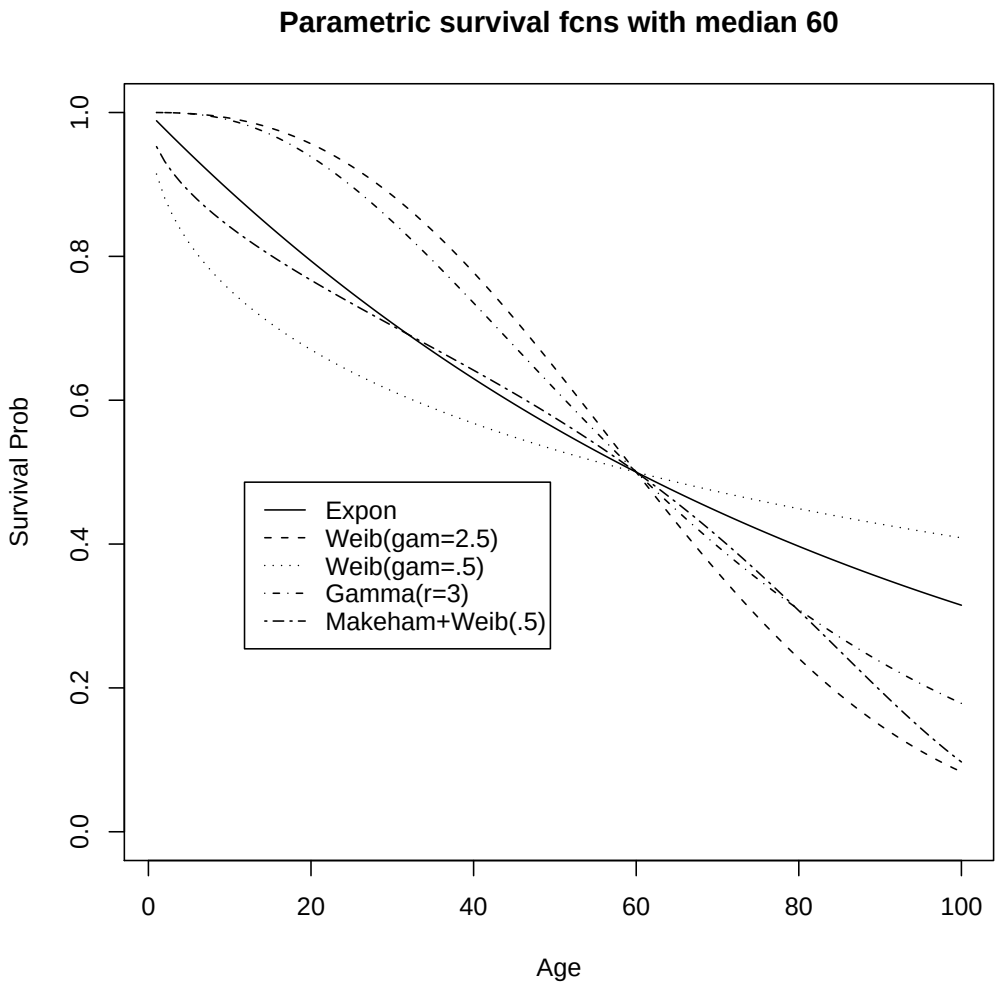

Figure 2: Graphs of survival functions from several parametric models designed to have common median 60.

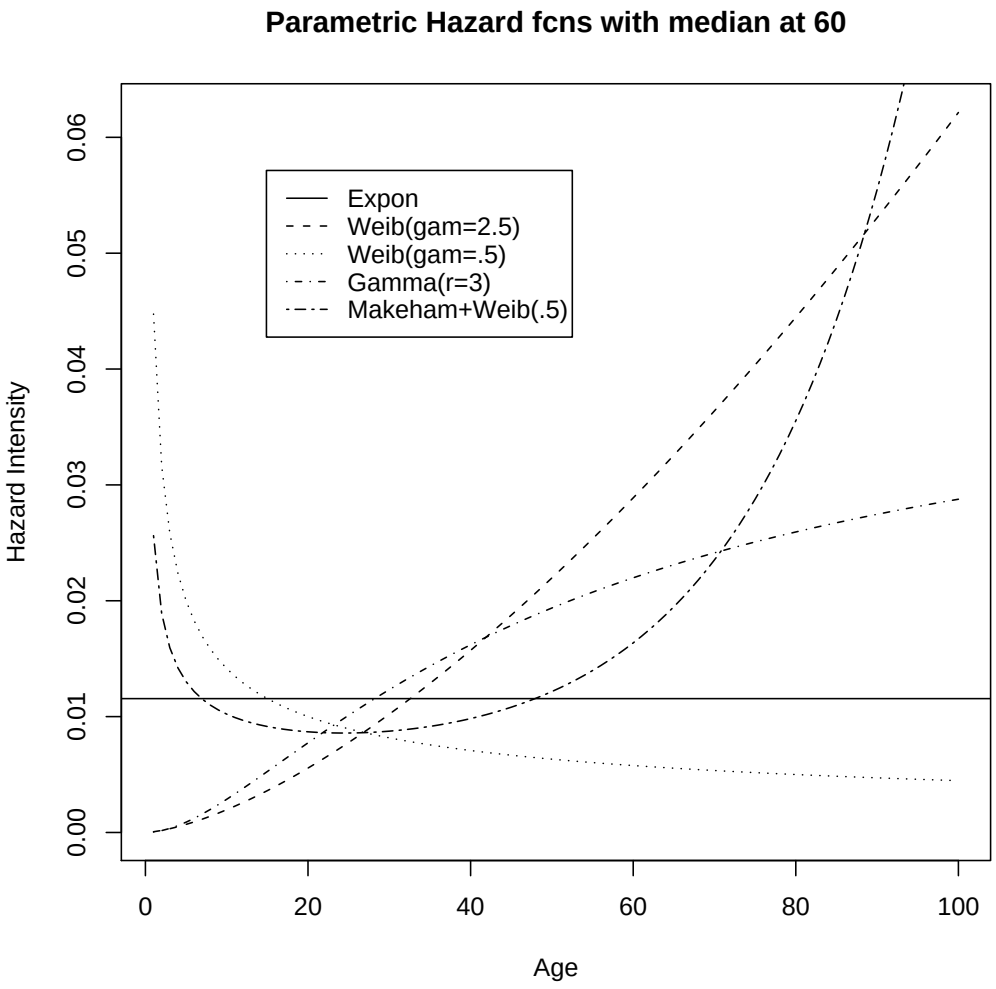

Figure 3: Graphs of hazard intensity functions for several parametric models designed to have common median 60.

II. Definition of Life Table

Refer all lifetimes [or specific cohort of lifetimes] to same origin.

Group into intervals of age $[x, x+1)$, with

$l_x =$ number of lives under observation ("at risk") at age x

$d_x =$ number of lives dying within age interval $[x, x+1)$

$c_x =$ number of lives removed (withdrawn, censored, lost to followup) during age interval $[x, x+1)$

$i_x =$ number of lives added to risk set during age interval ("immigrants")

Think of insurance portfolio. Actuaries derive from data like this a death-rate

$q_x =$ P(life aged x will die before age x+1)

which we interpret in terms of a death-time random variable T and its (continuous-time) probability distribution as

$$P(T < x+1 \,|\, T \geq x) \;=\; 1 - S(x+1)/S(x), \quad S(t) = P(T > t)$$

Typical "actuarial" estimate treats d_x, c_x, and i_x as happening at uniformly distributed times during year of age, approximating

$$q_x \;\approx\; d_x/(l_x + 0.5 \cdot (i_x - c_x))$$

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.

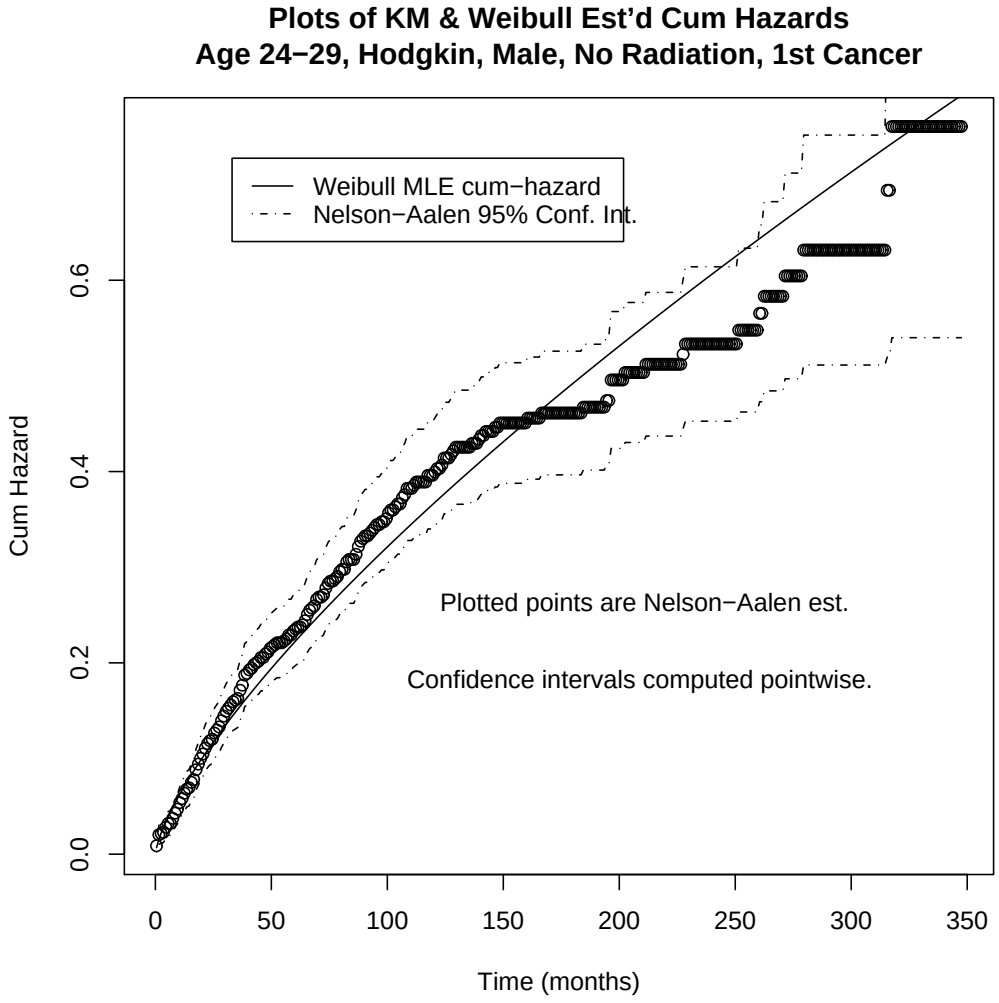

Figure 5: Comparisons of Weibull versus Nelson-Aalen estimated cumulative hazard functions for SEER data, Hodgkin Males.

III. Types of Statistical Professionals Analyzing Data Like This – Applications

Actuaries: estimate survival distributions for their portfolios of insured, and expected present values (discounted by inflation/interest) of future payouts under insurance and annuity contracts.

Biostatisticians: analyse survival data from epidemiologic and clinical survival studies, creating predictive models for hazards in terms of risk-factor combinations of "prognostic covariates"

Reliability Analysts: reliability testing of devices, modeling networks of devices and components

Economists ...

Survey Statisticians ...