1 "Survival" or "Time-to-Event" Data

- Examples (events not necessarily 'bad')
- Play down 'time-to'; emphasize its reciprocal (event rates, hazard function) & cumulative incidence
- Why such data need special techniques
- Types of censored data
- Distinction between censoring and truncation
- (equivalent) Functions: S[t] , hazard h[t] , pdf[t]
- Links: e.g. $S[t] = \exp\left[-\int_0^t h[u]du\right]$
- Summaries of these functions
- "Cause-specific" Survival; Competing Risks

(NON-PARAMETRIC / SEMI-PARAMETRIC)

- **Estimation** (point &interval) of $S[t], h[t]$ and $pdf[t]$

  - Lifetable [fixed interval] - Kaplan-Meier [data-determined] - - Nelson-Aalen [data-determined]
- **Comparisons**
- **Risk Sets**
- **Adjusted comparisons** (non-regression methods)

  Contrasts in unexposed and exposed person-time ("Time-dependent" exposure-)

SOFTWARE / GRAPHICAL DISPLAYS

APPLICATIONS

READINGS (* = most relevant)

`http://www.epi.mcgill.ca/hanley/c634/survival_analysis *`

- * Survival Analysis Sections 1 and 2 [Intro and Lifetables] Ch 17 of Armitage et al 4th ed.
- * Lifetables [and Survival after Treatment..] pp 199-205 of Ch 18 of Bradford Hill
- Survival Analysis Chapter 12 from Statistics at Square One [bmj online]
- Survival Analysis Chapter 11 from Statistical Methods for Comparative Studies by Anderson et 5 al.

OTHER RESOURCES

- `http://www.epi.mcgill.ca/hanley/c634/survival_analysis`
- Textbooks devoted to Survival Analysis by ...

  ...Hosmer & Lemeshow
  ...Collett
  ...Kleinbaum & Klein

"SURVIVAL" or **"TIME-TO-EVENT[1]" DATA**

- Examples (events not necessarily 'bad')
  - *women/couples* : becoming pregnant; fetuses: being born (gestational age)
  - *infants*: first sleep through the night, word uttered, walk, tooth, mosquito bite after application of (sham or real) prophylaxis, tooth eruption, caries

[1] Merriam-Webster `http://www.m-w.com/cgi-bin/dictionary`

Main Entry: EVENT. *Pronunciation*: i-'vent. *Function*: noun. *Etymology*: Middle French or Latin; Middle French, from Latin eventus, from evenire to happen, from e- + venire to come – *Date*: 1573 1 a archaic : OUTCOME b : the final outcome or determination of a legal action c : a postulated outcome, condition, or eventuality ¡in the event that I am not there, call the house¿ 2 a : something that happens : OCCURRENCE b : a noteworthy happening c : a social occasion or activity 3 : any of the contests in a program of sports 4 : the fundamental entity of observed physical reality represented by a point designated by three coordinates of place and one of time in the space-time continuum postulated by the theory of relativity 5 : a subset of the possible outcomes of an experiment.
JH would add an 'epi' definition: a **TRANSITION** from one state to another.

  - *infants*: last breast feeding, diaper (and the 'flip side' thereof *)
  - *adolescents:* first beer, cigarette, sexual intercourse, driving licence,
  - *then:* job, motor vehicle accident; university degree, marriage/cohabitation
  - *then*: first gray hair; Ph.D.; divorce; lose job; offspring born; grandchild, cancer diagnosis, menopause, bph, etc....
  - *new (transient) condition*: (headache, rash, cold,) ... resolution *(removed??) condition,* e.g. cancer: re-appearance ; death from *life threatening situation*, eg buried by avalanche: how long survive?

- Play down 'time-to'
  - emphasize its reciprocal (event rates, hazard function) & cumulative incidence at issue is exit from a state (to another), and the exit rates

- Why such data need special techniques
  - not everyone will experience event (no matter how long followed)
  - some haven't been followed for full length of time (enrolled late)
  - some 'lost to view'
  - some die (of unrelated causes) or have the "target" removed [NB "data not symmetrically/normally distributed" not reason per se] [likewise, absence of censored data doesn't mean one can't use survival analysis techniques.. see fruitfly survival data]

- Other types of censored data (besides *right*-censored & time)
  - **left** censored
    ... HepC +ve now, but *since when*?
    ... PSA level post prostatectomy 'undetectable' .. limit of detection
    ... Thermometer stops at -10C
  - **interval** censored
    ... onset of puberty / caries / when hiv+ : periodic examinations
    ... rounded or grouped measurements (eg age, income)
  - **right** censored
    ... measurement off the upper end of instrument scale
    ... open-ended category
    ... thermometer stops at +40C

- Distinction between censoring and truncation
  - censoring: **every** (or representative sample of) person(s)/object(s) is observed; have some bounds on the quantity
  - truncation: **some** person/objects not observed / excluded, and probability of in/exclusion has to do with the very quantity of interest.. the length of time ... , their size, etc. [length-biased sampling, deliberate exclusions, ..]

    Examples of truncated data...

    cross-sectional survey misses those who exit quickly

    * ask in 2004 for list of all Ph.D. students 'on the books' i.e. active in 1994 and determine in which year (Ph.D. 3, 4, ..) these students got the degree

    * survival of Alzheimer pts [Wolfson]

    * ask in 2004 for list of all patients on hospital census on randomly selected days in 2002; calculate average length of stay.

    sampling design misses objects of short sizes

    * select words by sticking a pin at random on page; measure average length of the words selected.

    * select inter-arrival times of buses via cross-sectional sampling design

    measuring instrument misses objects of short sizes

    * select fish using a given size mesh of net ;

    * lose rapid onset events if counter takes time to reset after previous event .. cars, radioactive disintegrations etc.

    exclude pts who die early, before 'an adequate trial of tx or

    [for 5-year survival, yes/no], include patients who entered study less than 5 years ago if they already died, but exclude those who entered less than 5 years ago but who have not died.

[EQUIVALENT] FUNCTIONS: $S[t]$ **,** **hazard** $h[t]$ **,** $pdf[t]$

T: random variable (duration, time to, time from T_0, etc..)

t: a specific point on T scale (eg 7 days / 5 years post-op)

- $S[t]$ (survival function)
  - $S[t] = Prob[T > t]$ is **unconditional**.
  - can debate whether to use $>$ or $\geq$; by convention in mathematical statistics, we define the complement of the $S[t]$ function, namely $1 - S[t]$, as $F[t] = Prob[T \leq t]$, so I will use $S[t] = Prob[T > t]$. In practice, since we measure time in discrete amounts, it is not an issue; survival textbooks are divided on this fine point. $F[t]$ is often called the cdf or cumulative distribution function (maybe that's where the silly term 'cumulative' survival comes from!)
- $h[t]$ (hazard function)
  - $h[t] = \lim_{\delta t \to 0} \frac{Prob[t<T\leq t+\delta t|T>t]}{\delta t}$ is **conditional**.
  - Can think of $h[t]$ as a short-term ('instantaneous') rate, in epi sense, with time denominator. To see why, consult page 12, section 1.3 of Collett, or consider the diagram in the next column.
  - Before taking limit, can see that the conditional probability

$$Prob[t < T \leq t + \delta t | T > t]$$

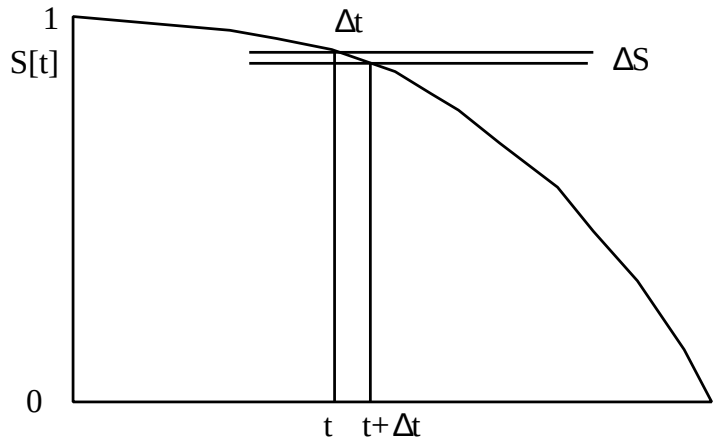
in the top part of the above expression can be re-written as

$$\frac{Prob[t < T \leq t + \delta t]}{Prob[T > t]}$$

The Numerator of this expression is proportional to the number of deaths in the interval, just like d_x in a lifetable. i.e., it is the amount by which S (or lower-case l in lifetable) changes during the interval. The Denominator of this expression is $S[t]$ and, in lifetable terms, is proportional to the number alive at $T = t$, and so has dimension 'persons'. Divide the top of the first expression by δt to get a quantity proportional to

$$\frac{d}{S[t] \times \delta t} = \frac{\text{number of deaths}}{\text{Person-time}}$$

Think of the rectangle standing on the base $(t, t + \delta t)$ as a person time denominator, and the $\delta S = d$ as the 'persons' numerator. As one narrows the δt, the rate hardly changes if the curve is smooth.



In mathematical-statistical terms, we replace d by the product of the probability density function $f[t]$ and the δt, so that the limit, after the δt cancels out, $h[t]$ becomes

$$h[t] = \frac{f[t]}{S[t]}$$

$f[t]$ is the negative of the derivative of $S[t]$, so can rewrite as

$$h[t] = \frac{-dlogS[t]}{\delta t}.$$

Rothman1986[2] says we can solve this differential eqn. to get

$$S[t] = \exp\left[- \int_0^t h[u]du \right].$$

Bottom line.. can reconstruct $h[t]$ from $S[t]$ & vice versa – or from $f[t]$ (see alternative derivation Incidence $\longleftrightarrow$ cumulative incidence, survival function Notes by JH in Resources for Lifetables.)

Survival analysis packages plot the negative of the log of the $S[t]$ curve against t, since it allows us, when comparing two curves, to judge more easily whether the hazard functions are proportional to each other at all values of t The integral of $h[u]$ up to t is called (not surprisingly) the 'integrated hazard'.

[2] The $PoissonProb(0\ events \mid \mu = integral]$ is an easier way to see this.

SUMMARIES of (3 equivalent) functions $S[t]$, $h[t]$ & $f[t]$

- median: the value of t at which S[t] = 1/2 (half-life" or t50)
- mean: the area under the (complete) S[t] curve (if available) equivalent to e_0 in life table
- quantile/fractile/percentile: the value of t at which S[t] equals some proportion or %
- x-year survival (or cumulative mortality): the value of S[t] at specified value of t

"CAUSE-SPECIFIC" SURVIVAL; COMPETING RISKS

Treat time of death from another cause (not of interest) as a censored observation (used a lot in cancer statistics)

- can give misleading answers if substantial other forces of mortality

  ... see material on prostate cancer on 626 web page.

- it is possible to have survival curves with 3 categories (alive, dead of target condition, dead of something else).

  ... Again, see 3-ply curves in Albertsen Hanley et al JAMA Sept 1998.

  same would apply to outcomes of starting a Ph.D.. e.g. at 5 years..
  ..... xx% have obtained a Ph.D.
  ..... yy% have decided it is not for them
  ..... zz% are still pursuing it

- used (sometimes naively) to calculate 'lifetime probability' of *developing* cancer or other condition.

  Should ask: does the calculation allow for the possibility that one might die of another cause before one could develop the target condition?

INFERENCE (Non-Parametric / Semi-Parametric)

- Estimation (point & interval) of S[t] , h[t] and pdf[t]
  - Lifetable (*fixed interval*) [Bradford Hill or Armitage]
  - Kaplan-Meier (*data-determined*) [cf. Armitage]
  - Nelson-Aalen (*data-determined*) [cf. Collett or Clayton/Hills]

- Comparison of Survival Data/Curves
  - *x-year* (e.g. 5-year) survival (or cumulative mortality)

    Use $SE[\hat{S}_{index-cat.}\hat{S}_{ref.-cat.}]$

    SE for each determined by formula of
    ... Greenwood (Armitage eqn 17.7 p 576)
    ... Kalbfleisch & Prentice (Armitage eqn 17.8 p 575)
    ... Peto (Armitage eqn 17.9 p 575)

  - *entire curves*

    log-rank test [M-H ; one 2×2 table / distinct event-time]
    Armitage section 17.6 p 576)

    note that it is a test;
    can be used to obtain 'relative death rates' (Armitage p578)

    Wilcoxon (Gehan) test; Peto test
    [Kleinbaum Chapter 2]

    all of these tests have the log-rank format,
    but weight the $(a - E[a_{null}])$ differences differently

    log rank : gives equal weight to each failure time
    Peto : gives more weight to early failure times

  - *Software:* SAS / Stata / R : see examples in Resources

2 Fitting Rate/Hazard/ID Functions via Regression Methods

2.1 Déjà

- **1 (homogeneous) sample**: "Survival" / "Time-to-event" data:
  - (equivalent) Functions: $S[t]$, hazard $h[t]$, $pdf[t]$
  - Links: e.g. $S[t] = \exp\left[-\int_0^t hu(du)\right]$
  - Summaries of these functions (e.g. $T_25, T_{50}, S[T]$)
  - Non-Parametric / Semi-Parametric Estimation (point & interval) of $S[t], h[t]$ & $pdf[t]$
    — Lifetable [fixed-ΔT's] & K-M/N-A [data-determined ΔT's]
  - Censored data not necessarily "time - to - event":
    Y = PSA levels < detection limit, salaries in intervals, distance travelled on set of tires, pages on single ink cartridge, etc.
  - '1 (homogeneous) sample' structure → think of as "intercept-only" regression model
- **Comparison of 2** Survival/Hazard Curves or Distributions
  - think of as regression model with single binary X
  - Risk Sets (match on time of event)
  - Adjusted comparisons (non-regression methods, e.g. standardization/MH)
- **Not covered**: Parametric models for *Lifetime* Distributions

  `SAS LIFEREG` procedure fits parametric models to failure time data that can be right, left, or interval censored. The models for the response variable consist of a linear effect composed of the covariates and a random disturbance term. The distribution of the random disturbance can be taken from a class of distributions that includes the extreme value, normal, logistic, and, by using a log transformation, the exponential, Weibull, lognormal, loglogistic, and gamma distributions.

  `Stata streg` performs maximum likelihood estimation of parametric regression survival-time models. Survival models currently supported are exponential, Weibull, Gompertz, lognormal, log-logistic and generalized gamma. Also see help stcox for estimation of proportional hazards models.

  `R survival package`: Regression for a Parametric Survival Model: These are all time-transformed location models, with the most useful case being the accelerated failure models that use a log transformation.

- **(Parametric) Regression Models for Rates**

  Model (event) rates or hazards

  Work in 'inverse-time' (t^{-1}) scale, rather than time-scale:

  $Rate = \frac{\textit{no. of events}}{\textit{amount of person-time}}$; $\quad TimeToEvent = \frac{\textit{amount of person-time}}{\textit{no. of events}}$

  i Models with '***multiplicative***' rates/hazards, e.g.,

  [β_V denotes 'regression coefficient associated with variate V']

  $$\log[rate] = \log[h] = \log[\lambda] = \beta_0 + \beta_t \times t + \beta_{X_1} \times X_1 + \beta_{X_2} \times X_2 \ldots$$
  $$rate = h = \lambda = e^{\beta_0 + \beta_t \times t + \beta_{X_1} \times X_1 + \beta_{X_2} \times X_2 \ldots}$$
  $$rate = h = \lambda = \underbrace{e^{\beta_0} \times e^{\beta_t \times t}} \times \underbrace{e^{\beta_{X_1} \times X_1} \times e^{\beta_{X_2} \times X_2} \times \ldots}$$

  Rates/hazards are PROPORTIONAL (rate ratio parameter constant over time-bands and covariate patterns...) if no product terms for 'effect modification'/'interaction').
  In Generalized Linear Model, we model the numbers of events, with log link .. and log(PT) as offset.

  $\underline{\exp[\hat{\beta}_0]}$: Rate/incidence/ID at $t = 0$, $X_1 = 0$, $X_2 = 0$, *etc.*.

  $\underline{\exp[\hat{\beta}]}$: Rate ratio/IR/IDR/HR:
  contrasting rates for two X (or t) values 1 unit apart

  ii Models with ***additive*** rates/hazards, e.g.,

  $$rate = h = \lambda = \beta_0 + \beta_t \times t + \cdots + \beta_{X_1} \times X_1 + \beta_{X_2} \times X_2 \ldots$$

  Not as 'natural'. See pp59– from Ch. 2 Vol. I of Breslow & Day (in Resources) for empirical evidence for better fit of proportional rate models (constant rate ratio models) than additive rates models (constant rate difference models) in cancer epidemiology.
  N.B.: B&D use the term "relative Risk" very loosely, when in fact they mean "relative Rates" or rate Ratios.
  In GLM, model the no.'s of events, with identity link .. no intercept (no cases if denominator is zero) & (as regressors) product of PT denominator with each regressor in the rate model.
  $\underline{\beta}$: rate difference; contrast of 2 rates one X (or t) unit apart

NOTE: Models contain terms for t, i.e., **TIME**, measured in suitable scale. Models are '***parametric-in-t***' or '***smooth-in-t***'

2.2 Semi-Parametric Models for Rate/Hazard functions

2.2.1 First, the basics...

Again, '***multiplicative***' rates/hazards, but now...

Split Model into 2 distinct parts:

$$\log[rate] = \log[h] = \log[\lambda] = \underbrace{\log[?(t)]}\ +\ \underbrace{\beta_{X_1} \times X_1 + \beta_{X_2} \times X_2 \ldots}$$

$$rate = h = \lambda = \underbrace{e^{\log[?(t)]}}\ +\ \underbrace{\beta_{X_1} \times X_1 + \beta_{X_2} \times X_2 \ldots}$$

$$rate = h = \lambda = \underbrace{?[t]}\ \times \underbrace{e^{\beta_{X_1} \times X_1} \times e^{\beta_{X_2} \times X_2} \times \ldots}$$

where $?(t)$ is an *unspecified* hazard function for the rates over time in the reference cell or profile (each of the X's = 0).

2.2.2 A few more details...

Shorthand: You will often see the h (or ID or $rate$ or λ) model written, with underlined $\underline{X}$ as shorthand for a vector of variates, and $\underline{\beta}$ as the corresponding vector of regression coefficients, as

$$hazard[t, \underline{X}] = h[t, \underline{X}] = \underbrace{h_0[t]} \times \underbrace{e^{\underline{\beta} \underline{X}}}$$

JH prefers to write it as

$$hazard[t, \underline{X}] = h[t, \underline{X}] = \underbrace{\{? \ h_0[t] \ ?\}} \times \underbrace{e^{\underline{\beta} \underline{X}}}$$

to emphasize the fact that the form of $h_0[t]$ is left unspecified.

The word "baseline": Statisticians often refer to the $h_0[t]$ function as the '*baseline*' hazard function. In this context, the word "baseline" does not refer to measurements (covariates) recorded at $T = 0$. A better name for it might be the '*hazard function for the reference profile*': the subscript ($_0$) means that it refers to the hazard function for the profile where all X variates are set to zero, against which all other profiles are compared. Thus, it has the same meaning as the "corner" or "point of departure" category used by Clayton and Hills (eg. "40-49 year olds, unexposed" in the regression example in Table 22.6 p 221 of Clayton and Hills.

You could also think of the entire curve $h_0[t]$ is the "intercept curve".

"Proportional" hazards: Rates/hazards are PROPORTIONAL (rate ratio parameter constant over time-bands and covariate patterns...) if no product terms for 'effect modification'/'interaction').

"How often are hazards "Proportional" ?: Often, one can predict, based on the biology of the situations, whether they might be. See examples, on earlier handouts, from JUPITER trial, COMPARE trial, SHEP trial, cancer screening for cancer or abdominal aortic aneurysms, weekend vs weekday admissions for MI, role of circumcision in prevention of HIV infection, etc. See if you can recognize which is which in the schematic examples in the Figure at end of these notes.

Where does the 'semi-parametric' come into it? The model is called semi-parametric because it only models a portion of the hazard function using the smooth parametric component $e^{\underline{\beta} \underline{X}}$ and avoids modelling the nuisance ("t") part. We don't fit parameters that (a) are not our focus (b) waste "degrees of freedom".

How does one 'get rid of' the nuisance part $h_0[t]$? We use risksets & conditioning to get rid of the $h_0[t]$, and thus focus on the the β parameters.

Risk sets are always ordered in time: sometimes, there are different possible choice for a time-scale: e.g. 'calendar tim, or age, or time since entry': How to choose which one is used to define the risksets? Cox and Farewell say '*use the the one over which the hazard function is the most difficult to model .. avoid this challenge: match risksets on this scale.*'

The Figure at end of the notes shows Framingham data analyzed with two different times scales, namely age and 'grant-year' (which started in 1948).

Links to analysis of matched case-control studies: These same multiplicative models, and the strategy of conditioning as a way of eliminating parameters, applicable to matched case-control studies and even to c-c and other (e.g. consumer choice*) studies with no 'time' element ["conditional logistic regression"]

(* Daniel McFadden shared the Nobel Prize for his development of theory and methods for analyzing discrete choice in Economics: `http://www.nobel.se/economics/laureates/2000/mcfadden-autobio.html`)

2.2.3 Once we fit the β's, how do we get survival curves or CI curves for different profiles?

See Julien M and Hanley JA. Profile-specific survival estimates: Making reports of clinical trials more patient-relevant Clinical Trials 2008; 5: 107-115. Under `r e p r i n t s` *on JH's main page*

$S_x[t]$ in terms of $S_0[t]$

- Remember general law: $S[t] = \exp[-H[t]]$, where $H[t]$ is the integral or integrated or "cumulative" hazard.
- (simplest case) Relationship between $S[t]$ curve for $x = 1$ and $S[t]$ curve for $x = 0$ ["corner"]

  $h[t|x = 1] = h_0[t] \times e^{\beta \times 1} = h_0[t] \times HazardRatio.$ So, ...

  integral of $h_1[t]$ = integral of $h_0[t] \times HazardRatio.$ So, ...

  $$\begin{aligned} S[t|x=1] &= \exp\{-H[t|x=0] \times HazardRatio\} \\ &= \{\exp\{-H[t|x=0]\}^{HazardRatio} \\ &= S_0[t]^{HazardRatio}, \end{aligned}$$

  In general, the HazardRatio would involve all of the variates on which the profile in question differered from the reference profile – here we just had a 1-dimensional profile, with just 2 levels, $x = 1$ (index) and $x = 0$ (reference).

S curve for a profile is constant power of curve for ref. profile.

But need to fit the β's first: obtaining the S or CI curves for various profiles is a "post-processing" option – most people seem to be unaware it exists.

2.2.4 Test of Proportionality

Two $log[-log[S]]$ functions (for $x = 1$ & $x = 0$) should be parallel

- H[t] is the integrated or "cumulative" hazard
- $-\log[S] = H[t]$, so $-\log[S_1[t]] = HR \times \{-\log[S_0[t]]\}$
- Two $-\log[S]$ curves should be proportional (easier to judge if these are parallel than that hazards are proportional)
- use as test of proportionality assumption
- hazard functions not stable enough to assess if h[t] curves are prop'nl)
- See textbooks for more details on tests of residuals, etc.

2.2.5 Readings

[`http://www.epi.mcgill.ca/hanley/c681/cox`]
Clayton&Hills, Ch 30, sections 4-6
Collett Textbook, Chapter 3/4
Kleinbaum's 'Self-Learning' textbook, Chapter 3/4
Pair of expository articles by JH

2.2.6 Exercises, Ch30, Cox'sRegressionAnalysis, Clayton & Hills

Table 30.1. A cohort of 10 subjects

		Entry to Study		End of Study	
Subject	Sex	Date	Age	Date	Age
A	F	13/ 6/65	29.3	31/12/89	53.8
B	M	23/10/72	25.2	31/12/89	42.4
C	M	3/3/59	22.1	31/12/89	52.8
D	F	10/10/67	32.2	31/12/89	54.4
E	M	2/ 1/60	33.1	4/ 7/79	52.6
F	M	9/ 1/75	42.1	31/12/89	57.1
G	F	5/8/53	35.2	3/10/68	50.4
H	M	10/10/69	27.0	31/12/89	47.2
I	M	2/3/72	44.8	31/12/89	62.7
J	F	1/11/70	51.5	31/12/89	70.6

Exercise 30.1. The data set out in Table 30.1 refer to 10 subjects from a cohort study. Subjects E and G died at the second date while the remaining eight subjects survived until the date of analysis (31/12/89). List the members of the rIsk sets for both deaths when the appropriate time scale is (a) calendar date (b) age (c) time since entry into the study.[3]

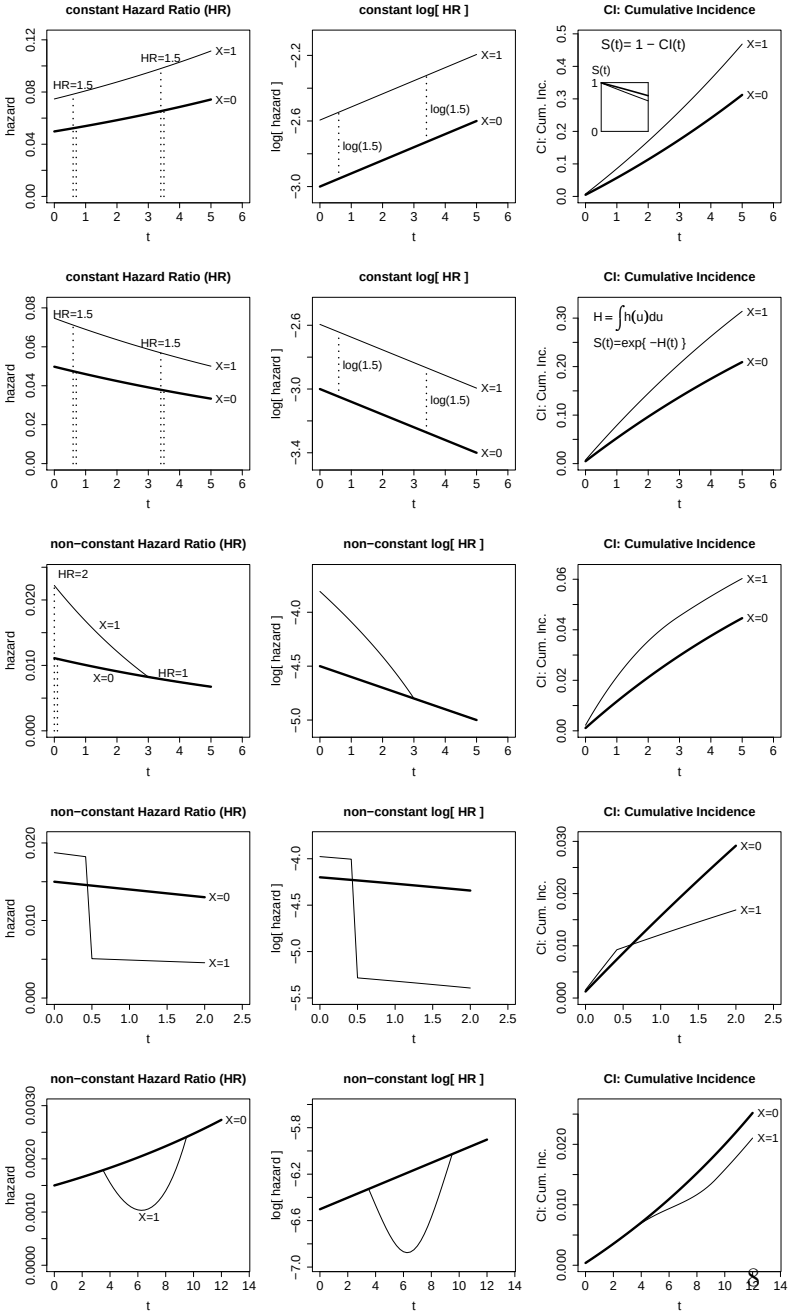
The difference between these analyses is that they represent three different models. In each case the model parameters represent variation of baseline rates along different time scales.

Exercise 30.2. Repeat Exercise 30.1 for an analysis which is to be stratified by sex.

[3] *Hint(JH): 10 'lifelines' drawn on a Lexis diagram make it easy to see who is in which.*

Schematic examples of PH and non-PH models

Course EPIB634: Proportional (and Non-proportional) Hazards Models {version 2008.03.28}



Comments

Fitting proportional hazards model: Risksets

Our prime interest is in estimating the parameters of HR; we will also, as a secondary objective, estimate $h_0(t)$. The keys to the estimation are the Risk Sets, the collections of candidates for (individuals at risk just before) each distinct failure time (event)

Simplest case (1 covariate z, 2 levels or Tx groups which we will distinguish using indicator variable z= 0 and z=1). In e.g. below, a • denotes a failure (event), a + denotes a censored observation; and time runs from left to right [*note*: to estimate HR function we do not need the failure & censoring times themselves, only their *order* with respect to z].

Raw data..(7 individuals).

```
z=1   •   +       •           •
z=0           •       +           •
```

It is easier to lay them out as separate time lines [in the 'early days' before computers, some investigators would represent survival data on their patients using lines of thread along a wall].

```
z=1 --•
z=1 ----+
z=1 ----------•
z=1 ----------•----------•

z=0 ----------•
z=0 ----------•----------+
z=0 ----------•----------•----•
Riskset #  1       2   3          4    5
```

Cox argued that since there are no failures (events) between the •'s, we do not know much about the hazards in these gaps [unless we want to posit parametric form for $h_0(t)$ or $S_0(t)$]. In any case our prime interest is in HR, and so we will concentrate just on these risk sets.

Estimating HR by (Partial) Likelihood approach

It helps to lay it out the 5 risk sets as follows (note that in the 5th riskset there is 'no contest') ...

$o=d_1$	**1**	0	**1**	**1**	-
s_1	3	2	1	0	-
n_1	4	2	2	1	
d_0	0	**1**	0	0	**1**
s_0	3	2	2	1	0
n_0	3	3	2	1	1

In the Maximum Likelihood method, we find that value of the HR which maximizes the likelihood of the **observed data pattern** (the **sequence** is indicated in **bold** above) The likelihood is a function of HR.

To construct the Likelihood function, we need a probability model for each table (i.e., for the outcome in each riskset) and an assumption regarding the separate tables. In the calculation of a variance for the MH statistic (log rank test) we already assumed that the 2x2 tables were realizations of hypergeometric (urn sampling) models and that the tables could be treated as if they were independent of each other. We could do the same here to set up a likelihood.

For each risk set, we ask

> ***"Given that the event occurred, what is the chance that it ocurred to the individual it happened to, rather than to someone else in the risk set?"***

Consider a risk set where the event happened at t to a person with z=1.

If the hazard for persons with z=1 is $HR \times h_0(t)$ and $1 \times h_0(t)$ for those with z=0, and if in the risk set there are n_1 and n_0 persons respectively, then the [conditional] probability that the event happened to that particular person with z=1 out of the n_1 and n_0 'at risk' is

$$\frac{HR \times h_0[t]}{n_1 \times HR \times h_0[t] \quad + \quad n_0 \times 1 \times h_0[t]}$$

which simplifies to

$$\frac{HR}{n_1 \times HR \quad + \quad n_0 \times 1}$$

Conversely, in a risk set where the event happened to a person with z=0. then the [conditional] chance that the event happened to that particular person with z=0 out of the n_1 and n_0 'at risk' is

$$\frac{1}{n_1 \times HR \quad + \quad n_0 \times 1}$$

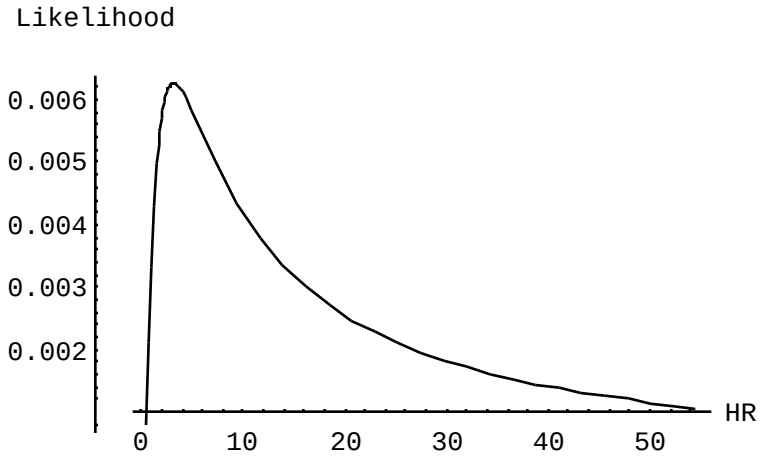
Thus, for the example above, the product of the probabilities of the observed outcome (likelihood) in each of the 4 informative risksets is

$$L = \frac{HR}{4HR+3} \times \frac{1}{2HR+3} \times \frac{HR}{2HR+2} \times \frac{HR}{HR+1}$$

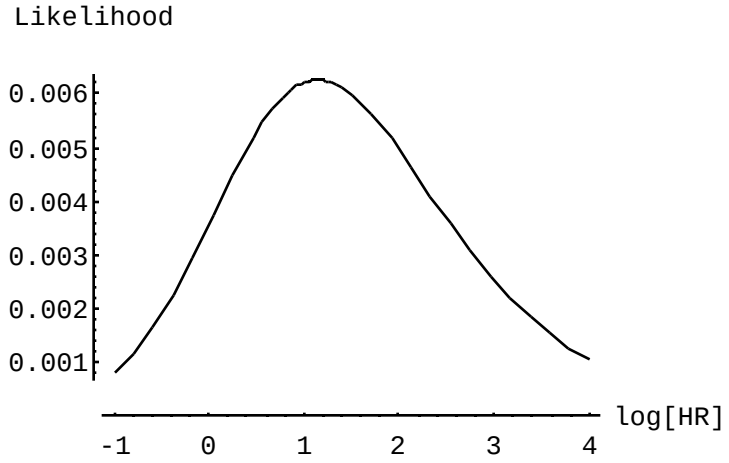
This likelihood L(HR)= prob(data | HR) can be evaluated for a range of HR values in order to find the value $\hat{HR}_{ML}$ which mazimises L. e.g.

HR	1/2	1	2	4	8	16
$Lx10^3$	1.4	3.6	5.8	6.1	4.8	3.0

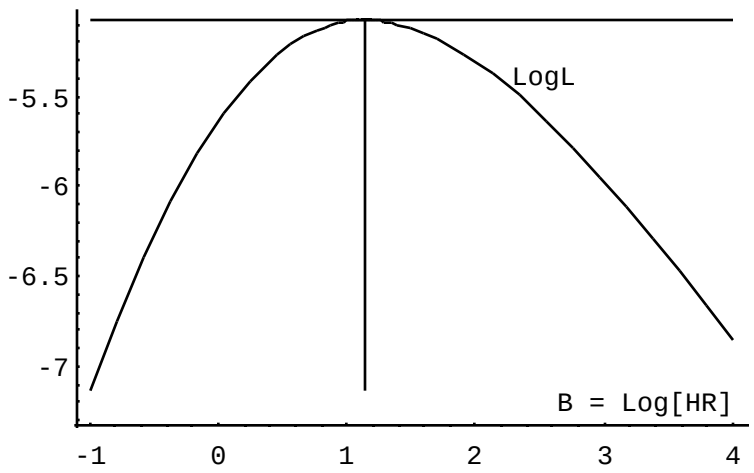
The function L & derived functions are shown graphically on next page.



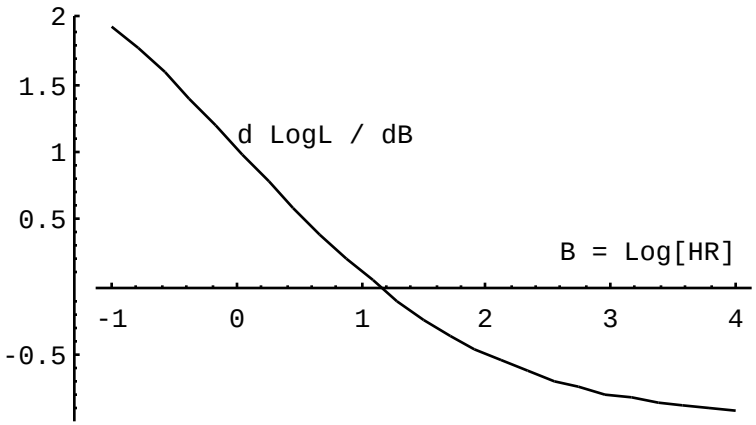
Or with the parameter B = Log[HR] ...



or in the log Likelihood scale...



The Derivative of the log Likelihood ...



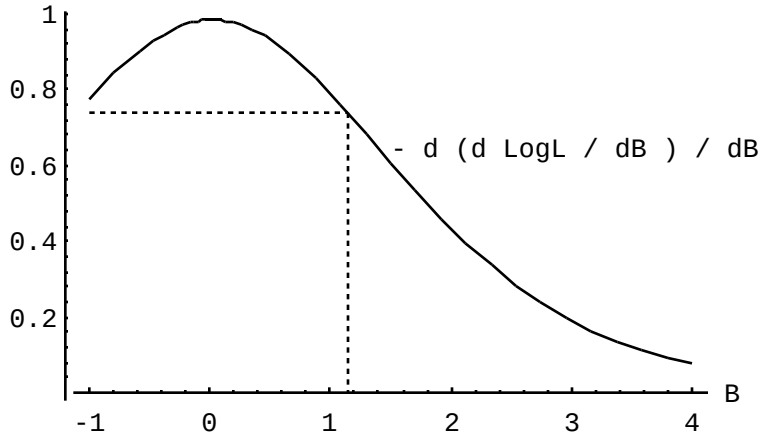
Tangent to logL curve is zero at B = 1.14 (we call this B_hat or b];

So... $\hat{HR}_{ML} = \exp[b] = 3.14$.

Uncertainty / Information concerning log [HR]

The 'sharpness' or 'flatness' of the logL(HR) curve in the vicinity of B = 1.14 gives an indication of how sensitive logL is to changes in log[HR] i.e. of how well or badly other values of log[HR] would do in producing a large likelihood. This can be measured by the 2nd derivative of logL (or if you like by the tangent to the 1st derivative curve) with respect to B. Note that the L curve increases until B = 1.14 then decreases. Thus the slope dlogL/dB goes from positive to negative over this range. ie the 2nd derivative is negative. Since we are simply interested in the curvature we use the negative of the 2nd derivative; it will be a big positive quantity when the curvature is very sharp, and a small positive quantity when the curvature is very slow.

The plot below shows that the curvature of logL is quite small (approximately 0.7412 at B = 1.14). This negative of the 2nd derivative of the log likelihood, evaluated at the ML estimate, is called the "**Information**" in the data. Its reciprocal is a good measure of the variance of the ML estimate of B.



We usually work with B =log[HR], since the sampling variability of b is more symmetric. The I[B] calculated at b = 1.14 is approximately 0.7412, yielding SE[b] = $\sqrt{(1/0.7412)}$ = 1.16, yielding a 95% CI for HR=exp[B] of {0.3 to 31}.The 4 informative risk sets provide just a small amount of information about log[HR] and our confidence in values near the ML estimate is low.

Estimating HR via SAS PROC PHREG

```
DATA a;
INPUT event  time   tx ;  /* Note arbitrary times             */
LINES;                    /* only ORDER matters               */
       1     2     1     /* event=0 stands for censored obsn. */
       0     4     1
       1     6     0
       1     8     1
       0    10     0
       1    12     1
       1    14     0
;
title null model; proc phreg ; model time*event(0) = ;
```

```
Dependent Variable: TIME       Number of Event & Censored Values
Censoring Variable: EVENT
Censoring Value(s):            Total   Event  Censored  %Censored
Ties Handling:      BRESLOW        7       5         2    28.57
```

NOTE: No explanatory variables in this model. -2 LOG L = 11.27

JH: LOG L = log[{1/7} x {1/5} x {1/4} x {1/2}] = LOG[1/280] = -5.63

```
title model with tx; proc phreg data=a;
                     model time*event(0) = tx / RISKLIMITS;
```

```
    Testing Global Null Hypothesis: BETA=0

        Without   With Covariates        Model Chi-Square

-2 LOG L  11.27   10.15              1.12  with 1 DF (p=0.29)

ML Estimates
         Parameter Standard  Wald    Pr >     Risk*      95% CL
Variable  Estimate   Error  Chi-Sq  Chi-Sq    Ratio    Lower  Upper
TX          1.14     1.16** 0.9685   0.33     3.14     0.32    30.6
```

* Technically speaking, should be called Hazard Ratio; Obtained as exp[1.14]

** See 2nd Derivative graph on left: SE[b] = Sqrt[var] = sqrt[1/Information]

Estimating HR via Stata

```
1.input event time tx
          1     2   1
          0     4   1   2. stset time , failure(event)
          1     6   0
          1     8   1
          0    10   0
          1    12   1
          1    14   0
   end
```

```
7 obs., representing 5 failures in single record/single failure data
56  total analysis time at risk, at risk from t =  0
earliest observed entry t = 0 last observed exit t =  14
.
* null model
stcox, estimate
         failure _d:  event
   analysis time _t:  time
Iteration 0:   log likelihood = -5.6347896

Log likelihood  =   -5.63   Prob > chi2     =         .
* model with tx.. gives beta_hats, not HR_hats
stcox tx, nohr
Iteration 0:   log likelihood =  -5.074435
LR chi2(1)      =     1.12
Log likelihood  =    -5.07   Prob > chi2     =    0.2898
 _t |    Coef. Std. Err.   z    P>|z|  [95% Conf. Int]
-------------------------------------------------------------
  tx |   1.143    1.161  0.98   0.325    -1.13   3.41
-------------------------------------------------------------
. * model with tx.. gives HR_hats, not beta_hats
. stcox tx
 _t | HazRatio Std. Err.  z    P>|z|   [95% Conf. Int]
-------------------------------------------------------------
 tx |  3.14      3.64   0.98   0.325     .32    30.55
-------------------------------------------------------------
```

Estimating HR via survival package in R

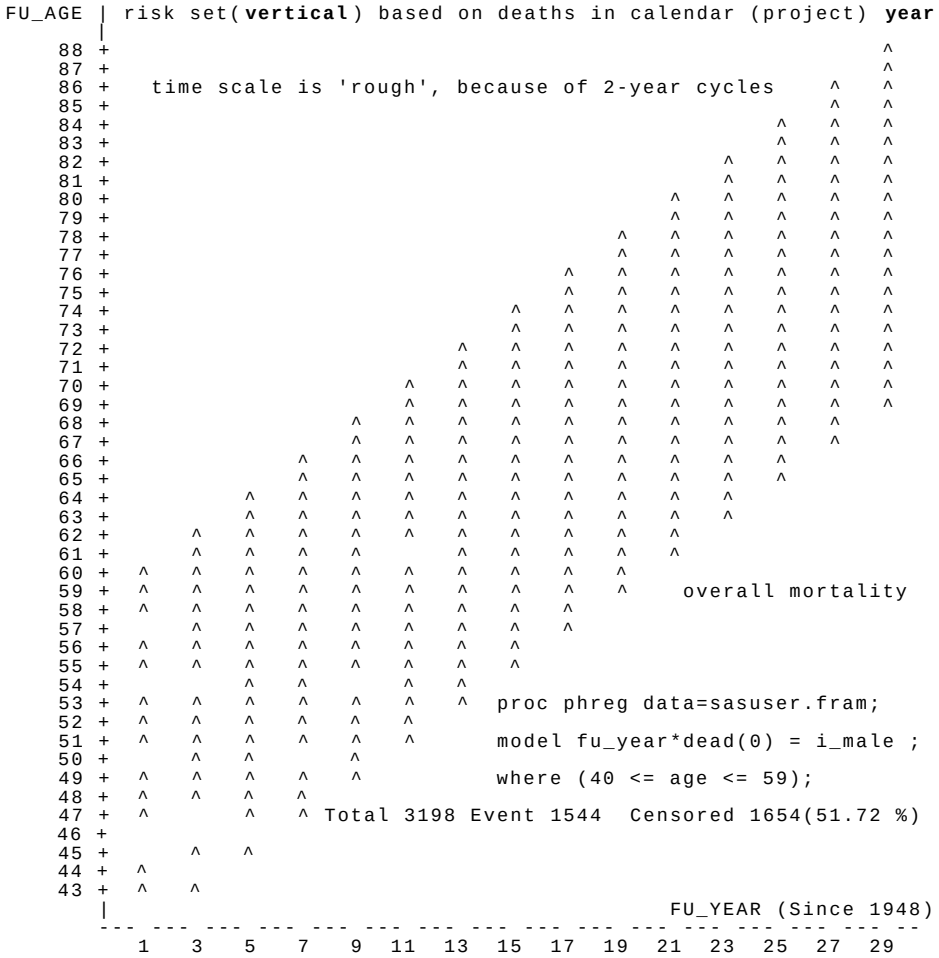
```
require(survival);      event=c(1,0,1,1,0,1,1);
time=c(2,4,6,8,10,12,14); tx =c(1,1,0,1,0,1,0);
fit=coxph( Surv(time, event) ~ tx); summary(fit)
```

```
   coef exp(coef) se(coef)     z    p
tx 1.14      3.14     1.16 0.984 0.33
               hr

   exp(coef) exp(-coef) lower .95 upper .95
tx      3.14      0.319     0.322      30.6

Likelihood ratio test= 1.12  on 1 df,   p=0.29
Wald test            = 0.97  on 1 df,   p=0.325
Score (logrank) test = 1.07  on 1 df,   p=0.3
```

Framingham study: **TIME Scale = YEAR_of_research_grant** ...



Testing Global Null Hypothesis: BETA=0

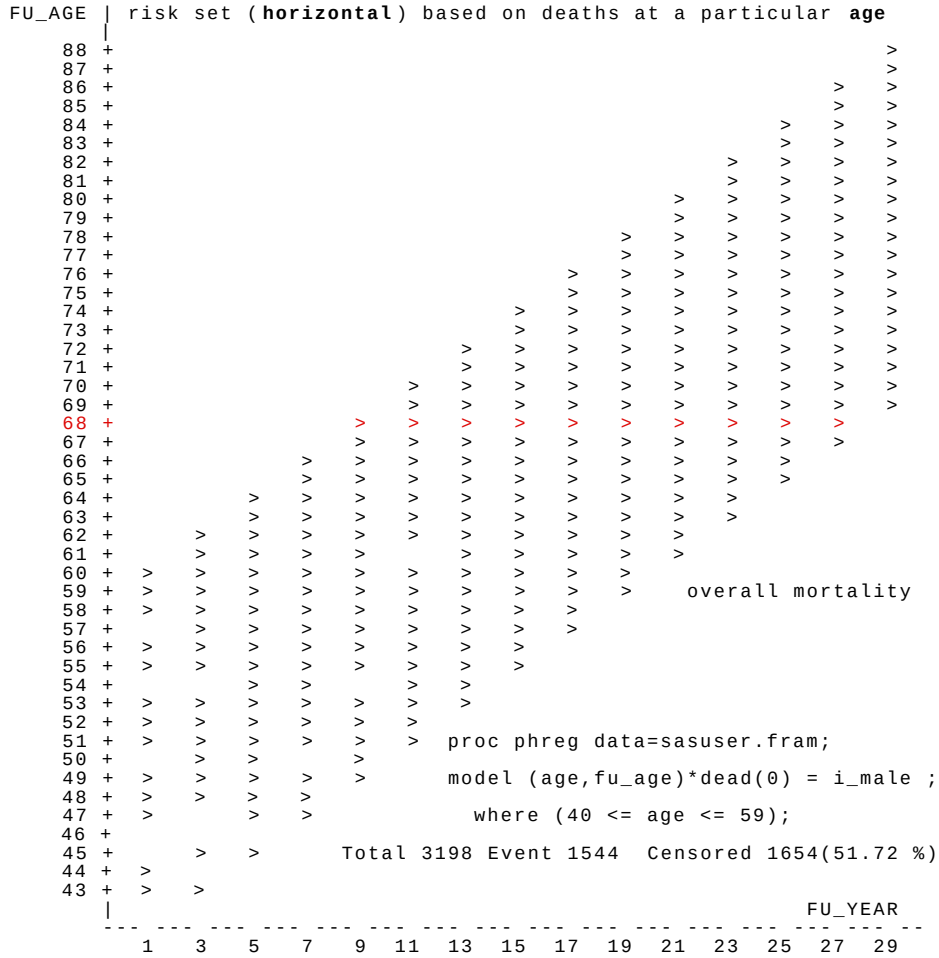
-2LOGL W/out:24098.1 With:23968.1 Covariates; Chi-Sq(1) 130; p=0.0001

Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Sq	Pr > Chi-Sq	Risk (hazard) Ratio
I_MALE	1	0.583	0.051	129.3	0.0001	1.79

Risk-sets "1" "3", ... candidates for deaths in FU-YEAR "1" "3" ...
(each set has persons with a range of ages)

TIME Scale = AGE .(NOTE how delayed entry is specified)



Testing Global Null Hypothesis: BETA=0

-2LOGL W/out:22819.8 With:22662.7 Covariates; Chi-Sq(1) 157; p=0.0001

Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Sq	Pr > Chi-Sq	Risk (hazard) Ratio
I_MALE	1	0.643	0.051	156.3	0.0001	1.90

Risk-sets "68" "69" ... ... candidates for death at age 68, 69,...

Mortality rates vary much more (and in more complex way) over 20 years of age, than over 20 calendar years **=> "t"=age**