

A Conceptual Approach to Survival Analysis

Laura Lee Johnson, Ph.D.

Statistician

**National Center for Complementary and Alternative
Medicine**

Fall 2008

johnslau@mail.nih.gov

Objectives

Vocabulary used in survival analysis

Present a few commonly used statistical methods for time to event data in medical research

The Big Picture

Take Away Message

**Survival analysis deals with making inference about
EVENT RATES**

Rate at t = Rate among those at risk at t

Look at Median survival (50%) not Mean survival

Mean: need everyone to have an event

Outline

How to Measure Time and Events

Truncation and Censoring

Survival and Hazard Functions

Competing Risks

Models and Hypothesis Testing

Example

Conclusions

What is a Model?

- **Basic**

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Y = outcome or response variable

β = coefficient

X = covariate, variable

Survival

$$\Lambda(t) = \lambda_0(t) \exp\{ \beta_1 X_1 + \dots + \beta_p X_p \}$$

$\Lambda_0(t)$ = baseline hazard

β_1, \dots, β_p = regression coefficients

X_1, \dots, X_p = prognostic factors

Vocabulary

Survival vs. time-to-event

Outcome variable = event time

Examples of events:

Death, infection, MI, hospitalization

Recurrence of cancer after treatment

Marriage, soccer goal

Light bulb fails, computer crashes

Balloon filling with air bursts

Time Notation

t: for time axis

t = 0 is the time origin

T: random outcome variable

time at which event occurs

Vocabulary

t = time

Baseline = 0 months

6, 12, 18, 24 months, etc.

S(t) = Survival at time t

$P[T \geq t]$ = Probability Time of event is greater than time t

Define the Outcome Variable

What is the *event*?

Where is the *time origin*?

What is the *time scale*?

Could do a *logistic regression* model

Yes/No outcome

Not focus of lecture

Choice of Time Scale

| Scale | Origin | Comment |
|--------------|-----------------|-----------------|
| Study time | Dx or Rx | Clinical Trials |
| Study time | First Exposure | (Occupational) |
| | Epidemiology | |
| Age | Birth (subject) | Epidemiology |

Treatment for a Cancer

Event = death

Time origin = date of surgery

Time scale = time (months)

T = time from surgical treatment to death

Graph = $P[T \geq t]$ vs t

Graph with Proportion Alive and Months Since Surgery

Y Axis: Proportion Alive - 1.0

X Axis: Months Since Surgery - 9

Example Numbers

$$S(9) = P[T \geq 9] = 0.25$$

25% is the probability the time from surgical treatment to death is greater than 9 months

“9 month post-resection survival is 25%” = Plain English

$$0 \leq S(t) \leq 1$$

Herpes Example

Recurrence of Herpes Lesions After Treatment for a Primary Episode

Event = recurrence

needs well defined criteria

Time origin = end of primary episode

Time scale = months from end of primary episode

T = time from end of primary episode to first recurrence

Toxin Effect on Lung Cancer Risk

Occupational exposure at nickel refinery

Event = death from lung cancer

Origin = first exposure

Employment at refinery

Scale = years since first exposure

T = time: first employed to death from LC

Population Mortality

Event = death

Time origin = date of birth

Time scale = age (years)

T = age at death

Volume of Air a Balloon Can Tolerate

Event = balloon bursts

t = ml of air infused

Origin = 0 ml of air in the balloon

T = ml of air in balloon when it bursts

Unique Features of Survival Analysis

Event involved

Progression on a dimension (usually time) until the event happens

Length of progression may vary among subjects

Event might not happen for some subjects

Sample Size Considerations

Event may not ever happen for some subjects

Sample sizes based on number of events

Work backwards to figure out # of subjects

Covariates must be considered (age, total exposure, etc)

Notation

T = event time

T^* = observation time

T if event occurs

Follow-up time otherwise

δ = failure indicator

1 if $T^* = T$

0 if $T^* < T$

“censor” or “censor indicator”

Outline

How to Measure Time and Events

Truncation and Censoring

Survival and Hazard Functions

Competing Risks

Models and Hypothesis Testing

Example

Conclusions

Truncation and Censoring

Truncation is about *entering* the study

Right: Only sample those with Event of interest (cancer registry)
(underestimate)

Left: “staggered entry”, >65 years of age

Censoring is about *leaving* the study

Right: Incomplete follow-up (common)

Left: Observed time > survival time (know the subject exists)

Independence is key

Left Truncation

Mention more in epi vs medical studies

Medical: zero-out at time of dx/tx

Key Assumption

Those who enter the study at time t are a random sample of those in the population still at risk at t

Allows one to estimate the hazard function $\lambda(t)$ in a valid way

Censoring

Incomplete observations

Right

Incomplete follow-up

Common and Easy to deal with

Left

Event has occurred before T_0 , but exact time is unknown

Not easy to deal with

Left Censoring

Age smoking starts

Data from interviews of 12 year olds

12 year old reports regular smoking

Does not remember when he started smoking regularly

Study of incidence of CMV infection in children

Two subjects already infected at enrollment

One Form of Right Censoring: Withdrawals

**Must be unrelated to the subsequent risk of event for
'independent censoring' to hold**

Accidental death is usually ok

Moves out of area (moribund unlikely to move)

Right Censoring

Graph showing right censoring

Types of Censoring

Type I censoring

T^* same for all subjects

Everyone followed for 1 year

Type II censoring

Stop observation when a set number of events have occurred

Replace all light bulbs when 4 have failed

Random censorship

Our focus, more general than Type I

**Key Assumption:
Independent Censoring**

Those still at risk at time t in the study are a random sample of the population at risk at time t , for all t

This assumption means that the hazard function ($\lambda(t)$) can be estimated in a fair/unbiased/valid way

Independent Censoring: If you have Covariates

Censoring must be independent *within* group

Censoring must be 'independent' given X

Censoring can depend on X

Among those with the same values of X, censored subjects must be at similar risk of subsequent events as subjects with continued follow-up

Censoring can be different across groups

Age Example

Early in trial older subjects are not enrolled

Condition on age: ok

**Do not condition on age: the estimates will be biased
because censoring is not independent**

Take Away: Study Types

Clinical studies

Time origin = enrollment, treatment begins

Time axis = time on study

Right censoring common

Epidemiological studies

Time axis = age

Right censoring common

Left truncation common

Bottom Line

Standard methods to deal with right censoring and left truncation

Key assumption is that those at risk at t are a random sample from the population of interest at risk at t

Outline

How to Measure Time and Events

Truncation and Censoring

Survival and Hazard Functions

Competing Risks

Models and Hypothesis Testing

Example

Conclusions

Survival Function

$$S(t) = P[T \geq t] = 1 - P[T < t]$$

Plot: Y axis = % alive, X axis = time

Proportion of population still without the event *by time t*

Survival Curve

Survival Function in English

Event = death, scale = months since Rx

“ $S(t) = 0.3$ at $t = 60$ ”

“The 5 year survival *probability* is 30%”

“70% of patients die within the first 5 years”

Everyone dies $\rightarrow S(X) = 0$

Hazard Function

Incidence rate, instantaneous risk, force of mortality

$\Lambda(t)$ or $h(t)$

Event rate *at t* among those at risk for an event

Key function

Estimated in a straightforward way

Censored

Truncated

Hazard Function in English

Event = death, scale = months since Rx

“ $\lambda(t) = 1\%$ at $t = 12$ months”

“At 1 year, patients are dying at a *rate* of 1% per month”

“At 1 year the chance of dying in the following month is 1%”

Hazard Function: Instantaneous

120,000 die in 1 year

10,000 die in 1 month

2,500 die in a week

357 die in a day

Instantaneous: move one increment in time

Survival Analysis

Models mostly for the hazard function

Accommodates incomplete observation of T

Censoring

Observation of T is 'right censored' if we observed only that $T >$ last follow-up time for a subject

Typical Intervention Trial

Accrual into the study over 2 years

Data analysis at year 3

Reasons for exiting a study

Died

Alive at study end

Withdrawal for non-study related reasons (LTFU)

Died from other causes

Outline

How to Measure Time and Events

Truncation and Censoring

Survival and Hazard Functions

Competing Risks

Models and Hypothesis Testing

Example

Conclusions

Competing Risks

Multiple causes of death/failure

**Special considerations of competing risk events
described in the literature**

Example:

event = cancer

death from MI = competing risk

No basis for believing the independence assumption

Competing Risks

Interpretation of $\lambda(t)$ = “risk of cancer at t when the risk of death from MI does not exist” isn’t practically meaningful

Rather, interpret $\lambda(t)$ = “risk of cancer among those at risk of cancer at t”

This will exclude MI deaths (if you are dead from an MI you are not at risk of cancer) and that is ok

Polar Bear Club Death Rates (fiction)

Annual death rates

3% taking dip 1Jan in Lake Michigan

2% Males all other causes

1% Female all other causes

Over a decade

25% of women died from taking a dip in Lake Michigan 1

Jan

24% of men died from taking a dip in Lake Michigan 1

Jan

Polar Bear Club Death Rates (fiction)

Why does it harm women?

Over a decade

33.5% of women died from all other causes

40% of men died from all other causes

There are more women to harm

People die of something

Which means they cannot die from something else

Bottom Line

We make inference about $\Lambda^{\text{obs}}(t)$ = event rate among subjects under observation at t

We can interpret it as $\Lambda(t)$ = event rate among subjects with $T \geq t$, if censoring is independent

Outline

How to Measure Time and Events

Truncation and Censoring

Survival and Hazard Functions

Competing Risks

Models and Hypothesis Testing

Example

Conclusions

Kaplan Meier

One way to estimate survival

Nice, simple, can compute by hand

Can add stratification factors

Cannot evaluate covariates like Cox model

No sensible interpretation for competing risks

Kaplan Meier

Multiply together a series of conditional probabilities

Chart

| Time t_i | # at risk | 3 events | Est. survival = |
|------------|-----------|----------|------------------------|
| 0 | 20 | 0 | 1.00 |
| 5 | 20 | 2 | $[1-(2/20)]*1.00=0.90$ |
| 6 | 18 | 0 | $[1-(0/18)]*0.90=0.90$ |
| 10 | 15 | 1 | $[1-(1/15)]*0.90=0.84$ |
| 13 | 14 | 2 | $[1-(2/24)]*0.84=0.72$ |

Kaplan Meier Curve

Graph:

X Axis: Proportion Surviving (95% confidence)

Y Axis: Survival Time

Kaplan Meier Estimator

One estimate of $S(t)$

Need independent censoring

If high risk subjects enter the study late then early on the K-M curve will come down faster than it should

Censored observations provide information about risk of death while on study

Kaplan Meier

Just the outcome is in many models

One or more stratification variables may be added

Intervention

Gender

Age categories

Quick and Dirty

How to Test? At a Given Time

$H_0: S_1(t) = S_2(t)$

Form test statistic

“Arbitrary time” – choosing t *post hoc*

Not using all of the data

Inference

For single event data inference about rates → inference for $S(t)$

No time dependent covariates, no recurrent events, no competing risk events

Logrank statistics compare event rates and allow the same generality as right censoring, left truncation

Log Rank

$H_0: S_1(.) = S_2(.)$

Test overall survival

2 independent samples from the same population

Observed # events vs. Expected #

Software; statistician should check

Some variations and some assumptions

Log Rank

Confounding

Are prognostic factors balanced between treatment groups?

Can see a difference using logrank, but just bias

Stratified Log Rank

Compare survival within each stratum

Essentially perform test within each stratum

Can prognostic factor be categorized?

Enough people per stratum?

Loss of power

Significance test, no estimates of difference

Proportional Hazards: Cox

Cox Proportional Hazards model

$$\Lambda(t) = \Lambda_0(t) \exp\{\beta_1 X_1 + \dots + \beta_p X_p\}$$

$\Lambda_0(t)$ = baseline hazard

β_1, \dots, β_p = regression coefficients

X_1, \dots, X_p = prognostic factors

$\beta = 0 \rightarrow$ hazard ratio = 1

Two groups have the same survival experience

Cox Proportional Hazards Model

Add covariates to the model

No need to stratify

Change in a prognostic factor -> proportional change in the hazard (on the log scale)

Statistical software

Can test the effect of the prognostic factor as in linear regression - $H_0: \beta=0$

Cox Model for Event Rates

Provides a framework for making inference about covariate effects

Semi-parametric

$\Lambda_0(t)$ completely unspecified

Multiplicative - $e^{\beta x}$

Effect of covariate is to multiply the rate by a factor

Cox cont.

Requires either that

RR is constant over time (proportional hazards), or

That we model RR over time

Allows time-dependent covariates and stratification factors

Age Example

Early in trial older subjects are not enrolled

If age is not in the Kaplan Meier then the KM estimate is biased because censoring is not independent

Put age in the Cox model – conditioned on age; ok

Age Example (cont.)

If I follow everyone for 1 year, am I ok?

Not necessarily

The study is not proportional by age to the population risk set

Could try to over sample older people later in the study to make the final study more correctly proportional

Easier to condition on age?

Testing Proportional Hazards

$$\Lambda(t) = \Lambda_0(t) \exp\{ \beta_1 \text{ age} + \beta_2 \text{ drug} \}$$

$$\exp\{ \beta_1 \text{ age} + \beta_2 \text{ drug} + \beta_3 \text{ age} * \ln(t) + \beta_4 \text{ drug} * \ln(t) \}$$

Look at p-values associated with β_3 and β_4 (Wald tests)

Do a partial likelihood ratio test comparing the two models

Look at Schoenfeld residual plots

Testing Proportional Hazards

| Variable | Coef | SE | P-value | 95%CI |
|------------|-------|------|---------|---------------|
| Drug | 0.58 | 0.25 | 0.020 | (0.09, 1.1) |
| Age | 0.18 | 0.03 | <0.001 | (0.12, 0.25) |
| Drug | 0.57 | 0.25 | 0.023 | (0.08, 1.1) |
| Age | 0.19 | 0.03 | <0.001 | (0.12, 0.26) |
| Drug*ln(t) | 0.002 | 0.16 | 0.988 | (-0.32, 0.31) |
| Age*ln(t) | 0.007 | 0.02 | 0.716 | (-0.03, 0.05) |

Testing Proportional Hazards

| Variable | Coef | SE | P-value |
|------------|------|------|---------|
| Drug | 4.24 | 0.61 | <0.001 |
| Age | 0.17 | 0.03 | <0.001 |
| Drug | 8.98 | 1.88 | <0.001 |
| Age | 0.19 | 0.03 | <0.001 |
| Drug*ln(t) | 2.71 | 0.84 | 0.001 |
| Age*ln(t) | 0.01 | 0.02 | 0.60 |

Time-Dependent Survival Curves

Failure to account for change in exposure/treatment over time

Usually assume there is no change

Think about HAART example

Stanford Heart Transplant Study (1971)

End-stage heart disease

Not responding

Seeking transplant

Heart Transplant Study

N=100

27 / 31 (87%) without transplant died

45 / 69 (65%) with transplant died

Exposure: Transplant yes/no

Outcome: time to death

Time origin: study entry

Fixed-Effect or Time Independent

Patients classified as ever/never receiving transplant during study

Chart:

Y axis: Probability

X Axis: Follow-up Time (days)

Kaplan Meier

Chart:

Y Axis: Proportion Surviving

X Axis: Survival Time

Timing of the Transplant?

Sample of Patients from Stanford Heart Transplant Study

Problem: Time Dependent Dataset

Total follow-up time (days)

Time of transplant (days)

Missing = no transplant

Transplant status (0=no, 1=transplant)

End of time interval for given transplant status (days)

Censoring (0=alive, 1=dead)

Patient ID

Effect of Transplant on Survival?

Graph of Time-Varying KM Plot

Take Home

Choose the right method and test

Kaplan Meier – simple

Logrank tests – useful, potentially misleading

Cox Proportional Hazards – workhorse

Not everything is proportional – check

Time matters

Changes in protocol matter

Outline

How to Measure Time and Events

Truncation and Censoring

Survival and Hazard Functions

Competing Risks

Models and Hypothesis Testing

Example

Conclusions

Example

Randomized clinical trial at Mayo: survival of patients with liver cirrhosis (NEJM 1982)

Two year survival probability of 0.88, calculated with Kaplan Meier

Compare a new treatment, D-penicillamine with placebo

Trial Information

Data collected at randomization

Presence/absence of ascites

Prothrombin time in seconds -10

Cox model

$$\Lambda(t) = \Lambda_0(t) \exp\{-0.135 X_{\text{TRT}} + 1.737 X_{\text{A}} + 0.346 X_{\text{P}}\}$$

How to say it in English

$$\Lambda(t) = \Lambda_0(t) \exp\{-0.135 X_{\text{TRT}} + 1.737 X_{\text{A}} + 0.346 X_{\text{P}}\}$$

X_{TRT} : 1 = D-penicillamine, 0 = placebo

X_{A} : 1 = ascites, 0 = no ascites

X_{P} : Prothrombin time – 10

Continuous, in seconds

$\Lambda_0(t)$ is the event rate at time t in the placebo arm for subjects without ascites with a prothrombin time of 10 seconds

$$\Lambda(t) = \Lambda_0(t) \exp\{-0.135 X_{\text{TRT}} + 1.737 X_{\text{A}} + 0.346 X_{\text{P}}\}$$

Relative rate of death two years post randomization for a subject on this trial who received the new treatment, had ascites at randomization and a prothrombin time of 10 seconds compared to a similar subject who received placebo?

$$\text{RR} = \exp\{-0.135\} = 0.87$$

Worked Out

Equation: $\exp\{0.135\} = 0.87$ is the relative rate of death for subjects who received treatment compared to those who received placebo

RR at Three Years?

Relative rate does not vary with time according to the proportional hazards model.

At the years the previously described RR is also $\exp\{-0.135\}$

Can work out RR for lots of other subject comparisons

But...

Physicians were initially reluctant to enter patients with ascites on the trial because of potential toxicity concerns

After about a year and a half recruitment became more representative of the clinic population

How does this Effect the Validity of the Kaplan Meier Estimator?

Censoring is not independent

At large t , the risk sets will not include patients with ascites because they were not recruited early enough and therefore are censored early.

The hazard function will be biased too small for larger t and so will be larger than the population survival function at large t .

In Short, What If

From first participant entered until the end of study: 4 years

Enroll for 3 years

Can be on study at least 1 year and up to 4 years

Followed enrollment to end of study

Do not start fully enrolling ascites until year 1.5

Ascites Participants

On study at least 1 yr and up to 2.5 yr

Do not have full population/risk set information at time $t > 2.5$ years

At time points $t > 2.5$ the study does not include a representative population

Ascites -> worse prognosis

KM estimate at $t > 2.5$ too high

Hazard is too small at larger t

Cox Model: Doomed Regression Coefficient Estimates?

No bias because conditional on covariates (including X_A)

Censoring must be independent GIVEN X

Censoring is independent and that is all that is required for consistency of the partial likelihood estimator (i.e. the coefficients)

Outline

How to Measure Time and Events

Truncation and Censoring

Survival and Hazard Functions

Competing Risks

Models and Hypothesis Testing

Example

Conclusions

Survival Picture

**Survival analysis deals with making inference about
EVENT RATES**

Rate at t = Rate among those at risk at t

Look at Median survival (50%) not Mean survival

**If you look at the mean you need everyone to have an
event**

Survival Analysis Can Handle

Right censoring

Left truncation

Recurrent events

Competing risks, etc.

Because we have available representative risk sets at t which allow us to estimate/model event rates.

Kaplan Meier

One way to estimate survival

Nice, simple, can compute by hand

Can add stratification factors

Cannot evaluate covariates like Cox model

No sensible interpretation for competing risks

Inference: Log Rank

Logrank statistics compare event rates and allow the same generality as right censoring, left truncation

For single event data inference about rates -> inference for $S(t)$

**No time dependent covariates, no recurrent events,
no competing risk events**

Cox Model for Event Rates

Provides a framework for making inference about covariate effects

Semi-parametric

$\Lambda_0(t)$ completely unspecified

Multiplicative - $e^{\beta x}$

Effect of covariate is to multiply the rate by a factor

Cox cont.

Requires either that

RR is constant over time (proportional hazards), or

That we model RR over time

Allows time-dependent covariates and stratification factors

Truncation and Censoring

Independence is key

Truncation is about *entering* the study

Right: Event has occurred (e.g. cancer registry)

Left: “staggered entry”

Censoring is about *leaving* the study

Right: Incomplete follow-up (common)

Left: Observed time > survival time

Course in General

Lots of assumptions

What is your n ? Probably small?

Try to have some intuition of data

Exploratory Data Analysis (EDA)

**Mean, median, variance or standard deviation,
quartiles**

Plots: histograms, box and scatter plots

Analyses

Fancy methods

Bread and butter

T-tests, Wilcoxon tests, chi-square

Linear or logistic regression

Basic survival (K-M, Cox PH)

Extensive Exploratory Data Analysis

Plots to match analysis

Your Question Comes First

May need to rewrite

If you change your question later

May not have the power

May not have the data

May have the wrong study population

COME TO THE STATISTICIAN EARLY AND COME OFTEN

Analysis Follows Design

Questions -> Hypotheses ->

Experimental Design -> Samples ->

Data -> Analyses ->Conclusions

**Take all of your design information to a statistician early
and often**

Guidance

Assumptions

Questions?

Each location?

Thanks!

Please fill out the course evaluations

**Please email me with specific examples or suggestions
to further improve the course**

johnslau@mail.nih.gov