

# Survival Analysis: Analysis of Right Censored Time to Event Data

.....  
Scott S. Emerson, M.D., Ph.D.  
Professor of Biostatistics,  
University of Washington

May 1, 2004

© 2002, 2003, 2004 Scott S. Emerson, M.D., Ph.D.

1

## Course Structure

.....

### Topics:

- First (and a half) Session: General Setting
  - Censored data setting
  - Estimation of survivor functions
  - Survival analysis models
- Second session: Comparison of Two Samples
  - Logrank statistic
  - Nonproportional hazards
  - Weighted logrank statistics
- Third session: Sequential Analysis
  - Stopping rules
  - Weighted logrank statistics in nonproportional hazards
  - Reweighted statistics

2

## Overview

.....

Scientific Studies

3

## Fundamental Philosophy

.....

Statistics is about science.

Science is about proving things to people.

- Other scientists
- Community at large

4

## Scientific Studies

.....

### A well designed study

- Discriminates between the most important, viable hypotheses
  - “Discriminates” defined by what convinces your audience
- Is equally informative for all possible study results
  - Binary search using prior probability of being true
  - Also consider simplicity of experiments, time, cost

(The Scientist Game)

5

## Scientific Questions

.....

Ultimately, scientific questions are most often concerned with investigating cause and effect

- E.g., in biomedical settings:
  - What are the causes of disease?
  - What are the effects of interventions?

6

## Typical Inferential Setting

.....

In the studies considered here, we define

- Some “primary outcome” measurement
  - A “response variable” in regression
- Groups that are homogeneous with respect to the level of some factor(s)
  - Predictor of interest
  - Effect modifiers
  - Confounders
  - Precision variables

7

## Primary Outcome Measurement

.....

The primary outcome can be derived from more than one measured variable

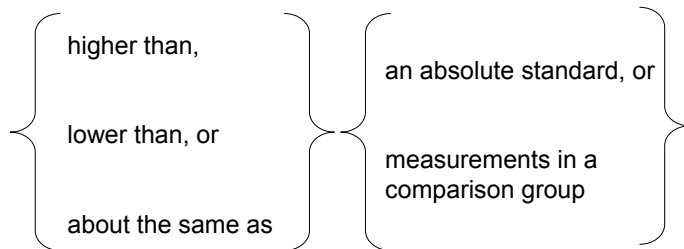
- E.g., for repeated measurements made on the same experimental unit
  - Contrast across repeated measurements
  - Weighted average of repeated measurements
- E.g., for random process defined by longitudinal follow up of experimental units
  - Contrast across time
  - Weighted average over time
  - Time until an event

8

## Typical Scientific Hypotheses

.....

The specified level of some factor will cause outcome measurements that are



9

## Causation vs Association

.....

Truly determining causation requires a suitable interventional study (experiment)

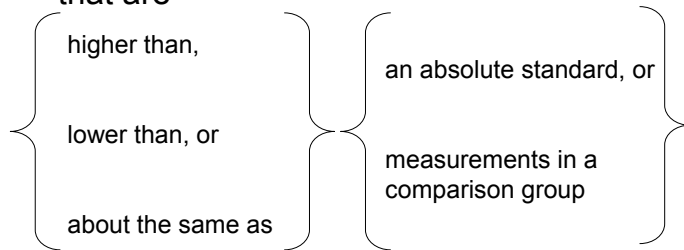
- Statistical analyses tell us about associations
- Associations in the presence of an appropriate experimental design allows us to infer causation
  - But even then, we need to be circumspect in identifying the true mechanistic cause
    - » E.g., a treatment that causes headaches, and therefore aspirin use, may result in lower heart attack rates due entirely to the use of aspirin

10

## First Statistical Refinement

.....

The group with the specified level of some factor will have outcome measurements that are



11

## Deterministic Setting

.....

Conditions of scientific studies might make answering questions difficult even when study results are deterministic

- Difficulties in isolating specific causes
  - E.g., isolating REM sleep from total sleep
  - E.g., interactions between genetics and environment
- Difficulties in measuring potential effects
  - E.g., measuring time to survival
    - » length of study
    - » competing risks

12

## Can Statistics Help?

.....

### Litmus Test # 1:

- If the scientific question cannot be answered by an experiment when outcomes are entirely deterministic, there is NO chance that statistics can be of any help.

13

## Variation in Response

.....

There is, of course, usually variation in outcome measurements across repetitions of an experiment

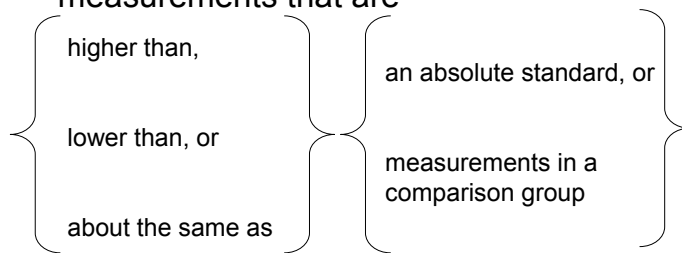
- Variation can be due to
  - Unmeasured (hidden) variables
    - E.g., mix of etiologies, duration of disease, comorbid conditions, genetics when studying new cancer therapies
  - Inherent randomness
    - (as dictated by quantum theory)

14

## Second Statistical Refinement

.....

The group with the specified level of some factor will tend to have outcome measurements that are



15

## Refining Scientific Hypotheses

.....

In order to be able to perform analysis we must define “will tend to have”

- Probability model for response
  - Nonparametric, semiparametric, parametric

(Looking ahead: I am a big proponent of nonparametric interpretations of statistical analyses)

16

## Ordering Probability Distributions

.....

In general, the space of all probability distributions is not totally ordered

- There are an infinite number of ways we can define a tendency toward a “larger” outcome
- This can be difficult to decide even when we have data on the entire population
  - Ex: Is the highest paid occupation in the US the one with
    - » the higher mean?
    - » the higher median?
    - » the higher maximum?
    - » the higher proportion making \$1M per year?

17

## Can Statistics Help?

.....

Litmus Test # 2:

- If the scientific researcher cannot decide on an ordering of probability distributions which would be appropriate when measurements are available on the entire population, there is NO chance that statistics can be of any help.

18

## Summary Measures

.....

Typically we order probability distributions on the basis of some summary measure

- Statistical hypotheses are then stated in terms of the summary measure
  - Primary analysis based on detecting an effect on (most often) one summary measure
    - Avoids pitfalls of multiple comparisons
      - » Especially important in a regulatory environment

19

## Purposeful Vagueness

.....

What I call “summary measures”, others might call “parameters”

- “Parameters” suggests use of parametric and semiparametric statistical models
  - I am generally against such analysis methods

“Functionals” is probably the best word

- “Functional”= anything computed from a cdf
- But too much of a feeling of “statistical jargon”

20

## Marginal Summary Measures

.....

Many times, statistical hypotheses are stated in terms of summary measures for univariate (marginal) distributions

- Means (arithmetic, geometric, harmonic, ...)
- Medians (or other quantiles)
- Proportion exceeding some threshold
- Odds of exceeding some threshold
- Time averaged hazard function (instantaneous risk)
- ...

21

## Comparisons Across Groups

.....

Comparisons across groups then use differences or ratios

- Difference / ratio of means (arithmetic, geometric, ...)
- Difference / ratio of proportion exceeding some threshold
- Difference / ratio of medians (or other quantiles)
- Ratio of odds of exceeding some threshold
- Ratio of hazard (averaged across time?)
- ...

22

## Joint Summary Measures

.....

Other times groups are compared using a summary measure for the joint distribution

- Median difference / ratio of paired observations
- Probability that a randomly chosen measurement from one population might exceed that from the other
- ...

23

## Looking Ahead: Transitivity

.....

The distinction between marginal versus joint summary measures impacts comparisons across studies

- Most often (always?) transitivity is not guaranteed unless comparisons can be defined using marginal distributions
  - Intransitivity: Pairwise comparisons might suggest
    - »  $A > B$ , and
    - »  $B > C$ , but
    - »  $C > A$

24

## Can Statisticians Help?

.....

While I claim that the choice of the definition for “tends to be larger” is primarily a scientific issue, statisticians do usually play an important role

- Statisticians do explain how different summary measures capture key features of a probability distribution

25

## Overview

.....

Choice of Summary Measure  
for Inference

26

## Hypothetical Example: Setting

.....

Consider survival with a particular treatment used in renal dialysis patients

- Extract data from registry of dialysis patients
  - To ensure quality, only use data after 1995
    - Incident cases in 1995: Follow-up 1995 – 2002 (8 years)
    - Prevalent cases in 1995: Data from 1995 - 2002
      - » Incident in 1994: Information about 2<sup>nd</sup> – 9<sup>th</sup> year
      - » Incident in 1993: Information about 3<sup>rd</sup> – 10<sup>th</sup> year
      - » ...
      - » Incident in 1988: Information about 8<sup>th</sup> – 15<sup>th</sup> year

27

## Hypothetical Example: Analysis

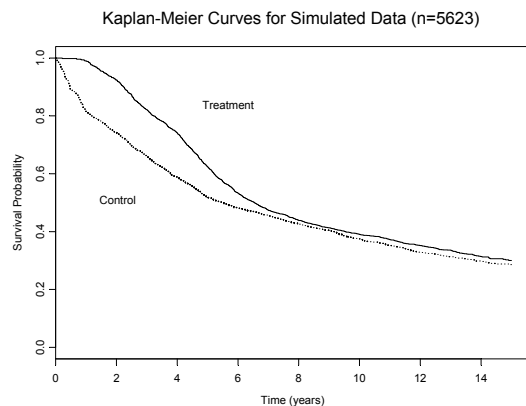
.....

Methods to account for censoring/truncation

- Descriptive statistics using Kaplan-Meier
- Options for inference
  - Parametric models
    - Weibull, lognormal, etc.
  - Semiparametric models
    - Proportional hazards, etc.
  - Nonparametric
    - Weighted rank tests: logrank, Wilcoxon, etc.
    - Comparison of Kaplan-Meier estimates

28

## Hypothetical Example: KM Curves



29

## Who Wants To Be A Millionaire?

Proportional hazards analysis estimates a  
**Treatment : Control** hazard ratio of

- A: 2.07 (logrank P = .0018)
- B: 1.13 (logrank P = .0018)
- C: 0.87 (logrank P = .0018)
- D: 0.48 (logrank P = .0018)

– Lifelines:

- 50-50? Ask the audience? Call a friend?

30

## Criteria for Summary Measure

We choose some summary measure of the probability distribution according to the following criteria (in order of importance)

- Scientifically (clinically) relevant
  - » Also reflects current state of knowledge
- Is likely to vary across levels of the factor of interest
  - » Ability to detect variety of changes
- Statistical precision
  - » Only relevant if all other things are equal

31

## Example of Scientific Issues

- E.g., Is the arithmetic mean's sensitivity to outliers desirable or undesirable?
  - Do we want to detect better infant mortality?
  - Does making one person immortal make up for killing others prematurely?
- E.g., Is the scientific importance of a difference in distribution best measured by the proportion exceeding some threshold?
  - Is an increase in survival time only important if the patient eventually makes it out of intensive care?

32



## Common Practice

.....

The overwhelming majority of statistical inference is based on means

- Means of continuous random variables
  - t test, linear regression
- Proportions (means of binary random variables)
  - chi square test (t test)
- Rates (means) for count data
  - Poisson analyses

33

## Use of the Mean

.....

### Rationale

- Scientific relevance
  - Measure of “central tendency” or “location”
  - Related to totals, e.g. total health care costs
- Plausibility that it would differ across groups
  - Sensitive to many patterns of differences in distributions (especially in tails of distributions)
- Statistical properties
  - Distributional theory known
  - Optimal (most precise) for many distributions
  - (Ease of interpretation?)

34

## Comments on Statistical Criteria

.....

Many of the reasons used to justify other tests are based on misconceptions

- The validity of t tests does NOT depend heavily upon normally distributed data
  - Modern computation allows exact small sample inference for means in same manner as used for other tests
- The statistical theory used to demonstrate inefficiency of the mean is most often based on unreasonable (and sometimes untestable) assumptions

35

## (Right) Censored Data

.....

### The Setting

36

## Censored Variables

.....

### A special type of missing data

- The exact value is not always known
  - Right censoring:
    - For some observations it is only known that the true value exceeds some threshold
  - Left censoring:
    - For some observations it is only known that the true value is below some threshold
  - Interval censoring:
    - For some observations it is only known that the true value is between some thresholds

37

## Example

.....

### A clinical trial is conducted to examine aspirin in prevention of cardiovascular mortality

- 10,000 subjects are randomized equally to receive either aspirin or placebo
- Subjects are randomized over a three year period
- Subjects are followed for fatal events for an additional three year period following accrual of the last subject

38

## The Problem

.....

### At the end of the clinical trial

- Some subjects have been observed to die
  - True time to death is known for these subjects
- Most subjects are likely to be still alive
  - Death times of these subjects are only known to be longer than the observation time
  - “(Right) Censored observations”

39

## What Should We Do?

.....

- Cannot ignore
  - These are our treatment successes
- Cannot just treat as binary (live/die) data:
  - Potential time of follow-up (censoring time) differs across subjects due to time of study entry
    - Confounding vs loss of precision
  - (Censored data may also arise due to loss to follow-up, e.g., moved away)
  - (Could figure out whether alive/dead at earliest censored observation, but this is inefficient and may not answer the question of interest)

40

## Right Censored Data

### Notation:

Unobserved :

True times to event :  $\{T_1^0, T_2^0, \dots, T_n^0\}$

Censoring Times :  $\{C_1, C_2, \dots, C_n\}$

Observed data :

Observation Times :  $T_i = \min(T_i^0, C_i)$

Event indicators :  $D_i = \begin{cases} 1 & \text{if } T_i = T_i^0 \\ 0 & \text{otherwise} \end{cases}$

41

## (Right) Censored Data

### Motivating Example

42

## Motivating Example

### Hypothetical study of subject survival

- Subjects accrued to study and followed until time of analysis
  - Study done at three centers, which started the studies in three successive years
  - Censoring time thus differs across centers

43

## Data (Real Time)

### Staggered study entry by site

		Accrual Group		
Year		A	B	C
1990	On study	100	--	--
	Died	43		
	Surviving	57		
1991	On study	57	100	--
	Died	27	53	
	Surviving	30	47	
1992	On study	30	47	100
	Died	13	22	55
	Surviving	17	25	45

44

## Data (Study Time)

### Realign data according to time on study

Year		Accrual Group		
		A	B	C
1	On study	100	100	100
	Died	43	53	55
	Surviving	57	47	45
2	On study	57	47	--
	Died	27	22	
	Surviving	30	25	
3	On study	30	--	--
	Died	13		
	Surviving	17		

45

## Combined Data

Year		Accrual Group			Combined
		A	B	C	
1	On study	100	100	100	300
	Died	43	53	55	151
	Surviving	57	47	45	149
2	On study	57	47	--	104
	Died	27	22		49
	Surviving	30	25		55
3	On study	30	--	--	30
	Died	13			13
	Surviving	17			17

46

## Problem Posed by Missing Data

### Sampling scheme causes (informative) missing data

- Potentially, we might want to estimate three year survival probabilities
- Different centers contribute information for varying amounts of time
  - One year survival can be estimated at A, B, C
  - Two year survival can be estimated at A, B
  - Three year survival can be estimated at A

47

## Possible Remedies

- **WRONG:** Ignore missing
  - E.g., 17 of 300 subjects alive at three years
- **RIGHT BUT WRONG QUESTION:** Use data only up to earliest censoring time
  - E.g., 149 of 300 subjects alive at one year
- **RIGHT BUT INEFFICIENT:** Use only center A
  - E.g., 17 of 100 subjects alive at three years

48

## Best Remedy

### • RIGHT AND EFFICIENT

- Use all available data to estimate that portion of survival for which it is informative
  - Use Centers A, B, and C to estimate one year survival
  - Use Centers A and B to estimate proportion of one-year survivors who survive to two years
  - Use Center A to estimate proportion of two-year survivors who survive to three years

49

## Theoretical Basis for Approach

### Properties of probabilities

- Probability of event A and B occurring is product of
  - Probability that A occurs when B has occurred
  - Probability that B has occurred

$$\Pr(A \cap B) = \Pr(A | B) \times \Pr(B)$$

50

## Application of Theory to Survival

For times  $T_1 < T_2$ , probability of surviving beyond time  $T_2$  is the product of

- Probability of surviving beyond time  $T_2$  given survival beyond time  $T_1$ , and
- Probability of surviving beyond time  $T_1$

For  $t_0 \leq t_1 \leq t_2 \leq \dots \leq t_k$

$$\begin{aligned}\Pr(T > t_j) &= \Pr(T > t_j \cap T > t_{j-1}) \\ &= \Pr(T > t_j | T > t_{j-1}) \Pr(T > t_{j-1})\end{aligned}$$

51

## Estimation of Conditional Survival

Estimate conditional probability of survival within each time interval

- Condition on surviving up until the start of the time interval
  - Denominator is number of subjects at start of interval
  - Numerator is deaths during the interval

52

## Requirements for Valid Estimates

Consistent estimates of survival probabilities depend on

- The subjects available at the start of each time interval must be a random sample of the population surviving to that time
  - “Noninformative censoring”
    - cf: Nonignorable missing, but noninformative censoring

53

## Estimation of Survival Probability

Estimate probability of survival at the endpoint of each time interval

- Multiply the conditional probabilities for all intervals prior to the time point of interest

54

## Obtaining Estimates

Within interval conditional probabilities

- Use A, B, C to estimate  $\Pr(T > 1)$
- Use A, B to estimate  $\Pr(T > 2 | T > 1)$
- Use A to estimate  $\Pr(T > 3 | T > 2)$

Multiply to obtain unconditional cumulative survival

- $\Pr(T > 1)$
- $\Pr(T > 2) = \Pr(T > 2 | T > 1) \Pr(T > 1)$
- $\Pr(T > 3) = \Pr(T > 3 | T > 2) \Pr(T > 2)$

55

## Combined Data

		Accrual Group			
Year		A	B	C	Combined
1	On study	100	100	100	300
	Died	43	53	55	151
	Surviving	57	47	45	149
2	On study	57	47	--	104
	Died	27	22		49
	Surviving	30	25		55
3	On study	30	--	--	30
	Died	13			13
	Surviving	17			17

56

## Survival Probability Estimates

Survival Probabilities				
Yr	Combined	Each Year		Cumulative
1	On study 300			
	Died 151			
	Surviving 149	149/300 = 49.67%		49.67%
2	On study 104			
	Died 49			
	Surviving 55	55/104 = 52.88%	.4967*.5288 = 26.27%	
3	On study 30			
	Died 13			
	Surviving 17	17/30 = 56.67%	.2627*.5667 = 14.88%	

57

## Improved Precision

Intuitively, these estimates would provide greater precision, because they are based on more data than using Center A alone

- We can show this exactly using confidence intervals

58

## Number At Risk and Number Failed

For notational convenience

For  $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_k$

$n_j$  = Number at risk in interval  $(t_{j-1}, t_j]$   
 $d_j$  = Number failed in interval  $(t_{j-1}, t_j]$

59

## Survival Probability Notation

For notational convenience

For  $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_k$

$$\begin{aligned}
 \pi_j &= \Pr(T > t_j) = \Pr(T > t_j \cap T > t_{j-1}) \\
 &= \Pr(T > t_j \mid T > t_{j-1}) \Pr(T > t_{j-1}) \\
 &= p_{j|j-1} \pi_{j-1} \\
 &= p_{j|j-1} \times p_{j-1|j-2} \times \dots \times p_{1|0} \\
 &= \prod_{i=1}^j p_{i|i-1}
 \end{aligned}$$

60

## Survival Probability Estimates

Maximum likelihood estimates for

- Conditional survival probability within intervals
- Unconditional survival probability

$$\hat{p}_{j|j-1} = 1 - \frac{d_j}{n_j}$$

$$\hat{\pi}_j = \prod_{i=1}^j \hat{p}_{i|i-1}$$

61

## Logarithmic Transformation

Sums are easier to work with than products

- The log transformed unconditional survival probability is the sum of log transformed conditional survival probabilities

$$\log(\pi_j) = \sum_{i=1}^j \log(p_{i|i-1})$$

$$\log(\hat{\pi}_j) = \sum_{i=1}^j \log(\hat{p}_{i|i-1})$$

62

## Basic Approach

We will find the standard error of the log transformed survival probabilities by

- Estimating each conditional survival probability and finding the variance of the log transformed estimates
- Invoking noninformative censoring to argue that the sum of our log transformed estimates must have the same distribution as the sum of log transformed independent estimates

63

## Standard Error of Proportions

From the laws of expectation, for the  $j$ th interval

For  $Y_1, Y_2, \dots, Y_{n_j} \stackrel{iid}{\sim} \text{Bernoulli}(p_{j|j-1})$

$$E[\bar{Y}] = p_{j|j-1}$$

$$\text{Var}(\bar{Y}) = \frac{p_{j|j-1}(1 - p_{j|j-1})}{n_j}$$

$$\text{se}(\bar{Y}) = \sqrt{\frac{p_{j|j-1}(1 - p_{j|j-1})}{n_j}}$$

64



## Large Sample Approximation

From the central limit theorem

For  $Y_1, Y_2, \dots, Y_n \stackrel{iid}{\sim} \text{Bernoulli}(p)$

$$\hat{p}_{j|j-1} = \bar{Y} \sim N\left(p_{j|j-1}, \frac{p_{j|j-1}(1-p_{j|j-1})}{n_j}\right)$$

65

## Logarithmic Transformation

From the delta method

$$\text{For } \bar{Y} \sim N\left(\theta, \frac{V}{n}\right)$$

$$g(\bar{Y}) \sim N\left(g(\theta), [g'(\theta)]^2 \frac{V}{n}\right)$$

So, for  $Y_1, Y_2, \dots, Y_n \stackrel{iid}{\sim} \text{Bernoulli}(p_{j|j-1})$

$$\log(\hat{p}_{j|j-1}) \sim N\left(\log(p_{j|j-1}), \frac{(1-p_{j|j-1})}{n_j p_{j|j-1}}\right)$$

66

## Noninformative Censoring

In the presence of noninformative censoring, the risk set in any interval should look like a random sample of the population at risk

- Estimates of the conditional probability of survival for the intervals should be uncorrelated

$$\log(\hat{\pi}_j) = \sum_{i=1}^j \log(\hat{p}_{i|i-1}) \sim N\left(\sum_{i=1}^j \log(p_{i|i-1}), \sum_{i=1}^j \frac{(1-p_{i|i-1})}{n_i p_{i|i-1}}\right)$$

67

## Confidence Intervals

Using the large sample approximation with plug-in estimates for standard errors

$$100(1-\alpha)\% \text{ CI for } \log(\pi_j)$$

$$\sum_{i=1}^j \log(\hat{p}_{i|i-1}) \pm z_{1-\alpha/2} \sqrt{\sum_{i=1}^j \frac{(1-\hat{p}_{i|i-1})}{n_i \hat{p}_{i|i-1}}}$$

100(1-α)% CI for  $\pi_j$

$$\exp\left\{\sum_{i=1}^j \log(\hat{p}_{i|i-1}) \pm z_{1-\alpha/2} \sqrt{\sum_{i=1}^j \frac{(1-\hat{p}_{i|i-1})}{n_i \hat{p}_{i|i-1}}}\right\}$$

68

## Survival Probability Estimates

Note the improved precision (and accuracy)

- Narrower CI even for the third year estimates

Survival Probabilities (95% CI)

Yr	Site A Only	Combined
1	0.570 (0.473, 0.667)	0.497 (0.443, 0.557)
2	0.300 (0.210, 0.390)	0.263 (0.212, 0.325)
3	0.170 (0.096, 0.244)	0.149 (0.102, 0.217)

69

## Aside: Greenwood's Formula

SE for the survival probabilities by a second application of the delta method

$$\log(\hat{\pi}_j) = \sum_{i=1}^j \log(\hat{p}_{i|i-1}) \sim N\left(\sum_{i=1}^j \log(p_{i|i-1}), \sum_{i=1}^j \frac{(1-p_{i|i-1})}{n_i p_{i|i-1}}\right)$$

$$\hat{\pi}_j = \exp\left(\sum_{i=1}^j \log(\hat{p}_{i|i-1})\right) \sim N\left(\pi_j, \pi_j^2 \sum_{i=1}^j \frac{(1-p_{i|i-1})}{n_i p_{i|i-1}}\right)$$

100(1- $\alpha$ )% CI for  $\pi_j$

$$\hat{\pi}_j \pm z_{1-\alpha/2} \times \hat{\pi}_j \sqrt{\sum_{i=1}^j \frac{(1-p_{i|i-1})}{n_i p_{i|i-1}}}$$

70

## Aside: Alternatives for CI

Three common methods for CI

- Based on log ( S(t) )
- Based on Greenwood's formula
- Based on log ( - log ( S(t) ) )
  - These intervals will always be between 0 and 1

71

## (Right) Censored Data

Product Limit  
(Kaplan-Meier)  
Estimates

72

## Kaplan-Meier Estimates

### Computation of probability of survival

Ordered distinct observation times :  $0 = t_0 \leq t_1 \leq \dots \leq t_k$

Time interval :  $(t_{j-1}, t_j]$

Number at risk at  $t_j$  :  $N_j$

Number of events at  $t_j$  :  $D_j$

Conditional probability of survival in interval :

$$\Pr(T > t_j | T > t_{j-1}) = 1 - \frac{D_j}{N_j}$$

73

## Kaplan-Meier Estimates

### Product Limit Estimate

Cumulative probability of survival :

$$\begin{aligned} \Pr(T > t_j) &= \Pr(T > t_j | T > t_{j-1}) \Pr(T > t_{j-1}) \\ &= \left(1 - \frac{D_j}{N_j}\right) \times \left(1 - \frac{D_{j-1}}{N_{j-1}}\right) \times \dots \times \left(1 - \frac{D_1}{N_1}\right) \\ &= \prod_{i=1}^j \left(1 - \frac{D_i}{N_i}\right) \end{aligned}$$

74

## Kaplan-Meier Estimates

### Note that in the above definition

- An interval which ends in a censored observation with no observed events has conditional probability of surviving within the interval is 1.
- If the largest observation time is censored, the KM (PLE) survivor function never goes to zero
  - We generally regard the KM (PLE) survivor function to be undefined for times beyond the largest observation time in this situation

75

## Kaplan-Meier Estimates

### Properties

- The KM (PLE) survivor functions can be shown to be
  - Consistent: As sample sizes go to infinity, they estimate the true value
  - Nonparametric maximum likelihood estimates
    - (but usual asymptotic theory for regular, parametric MLE's does not necessarily hold)

76

## Alternative Derivations

.....

The KM (PLE) survivor functions can also be derived as the

- Self-consistent estimator
  - (see Miller, Survival Analysis)
- “Redistribute to the right” estimator
  - Provides intuition regarding noninformative censoring

77

## Redistribute to the Right

.....

Basic idea

- Recall the empirical cdf assigns probability  $1/n$  to each observation
  - Each subject in a sample is representative of  $1/n$  of the population
- A censored observation should be equally likely to have event time like any of the remaining uncensored observations
  - Recursively redistribute the mass of each censored observation among the subjects remaining at risk

78

## Redistribute to the Right Example

.....

- Data: 1, 3, 4\*, 5, 7\*, 9, 10 (asterisk means censored)
- Initially: each point has mass  $1/7$
- Determine probability of events at earliest observed (uncensored) event times
  - $\Pr(T = 1) = 1/7$
  - $\Pr(T = 3) = 1/7$

79

## Redistribute to the Right Example

.....

- Censored observation at 4
  - Divide the mass at 4 equally among the remaining subjects at risk
    - Now mass of  $1/7 + 1/28 = 5/28$  for each of 5, 7, 9, 10
- Determine probability of events at next observed (uncensored) event times
  - $\Pr(T = 5) = 5/28$

80

## Redistribute to the Right Example

- Censored observation at 7
  - Divide the mass at 7 equally among the remaining subjects at risk
    - Now mass of  $5/28 + 5/56 = 15/56$  for each of 9, 10
- Determine probability of events at next observed (uncensored) event times
  - $\Pr(T = 9) = 15/56$
  - $\Pr(T = 10) = 15/56$

81

## General Analysis Models

### Risk Sets and Hazard Functions

82

## Hazard Functions

From the approach to nonparametric estimation of survival curves we see the importance of the hazard function

- Hazard = instantaneous risk of failure
  - Conditional upon being still alive, what is the probability (rate) of failing in the next instant

Hazard function :

$$\lambda(t) = \lim_{h \downarrow 0} \frac{\Pr(T \in [t, t+h) | T \geq t)}{h}$$

83

## General Notation

Failure time  $T > 0$  measures time to an event :

Cumulative distribution function :  $F(t) = \Pr(T \leq t)$

Survivor function :  $S(t) = \Pr(T > t) = 1 - F(t)$

Density :  $f(t) = \frac{d}{dt} F(t)$

Hazard function :  $\lambda(t) = \lim_{h \downarrow 0} \frac{\Pr(T \in [t, t+h) | T \geq t)}{h}$

Cumulative hazard function :  $\Lambda(t) = \int_0^t \lambda(u) du$

84

## Relationship to Survivor Function

.....

The survivor function (and, hence, the cdf) is uniquely determined by the hazard and vice versa

$$S(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(u) du\right)$$

$$\lambda(t) = -\frac{d}{dt} \log S(t) = \frac{f(t)}{S(t)}$$

Continuous distributions :

$$\lambda(t) = -\frac{d}{dt} \log S(t) = \frac{f(t)}{S(t)}$$

85

## Estimating the Hazard Function

.....

The intuitive estimator of the hazard function is thus the conditional probability of failure at each point in time

Number at risk at $t$ :	$N_t$
Number of events at $t$ :	$D_t$
Estimated hazard at $t$ :	$\hat{\lambda}(t) = \frac{D_t}{N_t}$

86

## Risk Sets

.....

Survival analysis often focuses on the “risk set” at each time

- “Risk set at time  $t$ ” = the set of subjects in the sample who are at risk for failure at  $t$ 
  - These subjects can be used to compute and compare hazard functions and, hence, survival probabilities

87

## Risk Based Analysis Models

.....

Analyses based on hazard functions afford the opportunity to allow sampling schemes which sample the population at risk at each time

- Advantages:
  - More efficient use of available data
  - Time-varying covariates
- Disadvantages:
  - Less intuitive summary measures
  - Temptation to use time-varying covariates

88

## General Analysis Models

.....

### A Useful Analogy

89

## Urn Model

.....

Balls in an urn of various colors and patterns

- Balls might represent people in a study
  - At any given time, the balls that are in the urn are therefore the risk set
- Colors and patterns represent risk factors

90

## Death Process

.....

Periodically, I come in and choose a ball from the urn and take it

- When a ball is chosen it fails
- My predilection for choosing certain colors or patterns identifies true risk factors
- Characteristics of the balls that I do not notice have no effect on survival probabilities

91

## Evidence for Risk Factors

.....

A certain color/pattern must be my favorite if

- (Time based observations)
  - I come in more often when that color/pattern is in the urn
    - You need not consider what else is in the urn
- (Risk set based observations)
  - I choose that color/pattern with a frequency disproportionate to its frequency in the urn
    - If I am blind to a characteristic, my choices should look like random sampling
    - You need not consider the times that I come in

92

## (Semi)parametric Models

.....

Two general (semi)parametric probability models used in survival analysis

- Accelerated failure time models
  - Consider time of failure
- Proportional hazards models
  - Consider relations among hazards
  - (Additive hazards models also used, but less frequently)

93

## Accelerated Failure Time Models

.....

Two groups that differ in some risk factor have survivor functions related by a parameter measuring acceleration or deceleration of time

$$S_1(t) = S_0(\theta t)$$

- E.g.,
  - A smoker ages twice as fast as a nonsmoker
  - Each human year is seven dog years

94

## Proportional Hazards Models

.....

Two groups that differ in some risk factor have survivor functions related by a parameter measuring increased hazard

$$\lambda_1(t) = \theta \lambda_0(t)$$
$$S_1(t) = [S_0(t)]^\theta$$

- E.g.,
  - At any given time, a smoker is ten times more likely to develop lung cancer as a nonsmoker

95

## Scientific Studies

.....

As a scientist you may

- Observe
  - When I come into the room and take a ball,
  - The colors/patterns on all the balls in the urn, and
  - The color/patterns on the ball that I take
- Experiment
  - Change the composition in the urn and see
    - Whether I come in the room more or less often, and
    - The lengths to which I might go to find balls with certain colors or patterns by restricting my choices

96



## Altering the Risk Set

.....

Censoring and time-varying covariates are analogous to changes in the composition of the urn

- Censoring = removing balls from the urn
- Time-varying covariates = repainting the balls or adding different balls

97

## Caveats: Informative Censoring

.....

Altering the risk set can be problematic

- Recall that in order for survival estimates to be consistent, the risk set in the sample must look like a random sample from the population
  - You should not selectively remove or change balls that were (for their risk factors) particularly more likely or less likely to be chosen
    - If you notice that I search the urn from top to bottom,
      - » Don't just change the balls sitting at the top of the urn
      - » Make sure you stir the urn after each change

98

## Caveats: Time-varying Covariates

.....

Time-varying covariates are far more easily implemented in the hazard based models

- Risk set approach makes this easy

However, scientifically we run the risk of overfitting our data using variables we are less interested in

- A priest delivering last rites is highly predictive of death and that may obscure that it was a gunshot wound that led to the death

99

## General Analysis Models

.....

### Noninformative Censoring

100

## Noninformative Censoring

.....

Censoring must not be informative about subjects who were either more or less likely to have an event in the immediate future

- The censored individuals must look like a random sample of those individuals at risk at the time of censoring
- (Later we shall say that they are a random sample from all subjects at risk having similar modeled covariates)

101

## Examples of Informative Censoring

.....

- Subjects in a clinical trial are withdrawn due to treatment failure (likely they would die sooner than those remaining)
- Subjects in a clinical trial in a fatal condition are lost to follow up when they go on vacation (likely they are healthier than those remaining)

102

## Examples of Informative Censoring

.....

- Leukemia patients in a clinical trial of bone marrow transplantation are censored if they die of infections rather than dying of cancer (the subjects who died of infections might have had a more effective regimen to wipe out existing cancer)

103

## Detecting Informative Censoring

.....

As a general rule it is impossible to use the data to detect informative censoring

- The necessary data is almost certainly missing in the data set
- In some cases, it is impossible to ever observe the missing data
  - Nonfelines can only die once
  - We cannot observe whether subjects dying of one cause are more or less likely to die of another if we cure them of the first cause

104

## Competing Risks

.....

This last situation is often referred to as “Competing Risks”

- Some “nuisance” event sometimes precludes your ability to ever observe the event of interest
- In the presence of competing risks, we must decide how best to address the scientific question of interest

105

## Example: Censoring Mechanisms

.....

Consider a study of smoking as a risk factor for incidence of cancer

- Possible causes of censored observations
  - Subject still alive at time of data analysis
  - Subject lost to follow-up during study
  - Subject died in airline accident
  - Subject died in single car accident
  - Subject died of MI
  - Subject died of emphysema

106

## Censoring Competing Risks

.....

Time to cancer, but competing risk of death

- Suppose we censor deaths
  - If deaths represent noninformative censoring
    - People who died of, say, MI neither more nor less likely to get cancer in the near term
    - Estimates desired hazard rate
  - If deaths represent informative censoring
    - Estimates cause specific hazard in presence of unchanged risk of competing event
    - Results are not generalizable to a population with an altered risk of death

107

## Competing Risks: Alternatives

.....

- Model informative censoring
  - Model must be based on untestable assumptions
- Event free survival
  - Like censoring deaths if competing risk hazard low
  - Like censoring deaths if everyone gets cancer first
  - Loss of power if truly noninformative censoring
- Wilcoxon like statistic
  - Rank first on death times; break ties with cancer dx
  - Like survival only if everyone dies
- Survival only
  - Not really the question, especially if competing risk hazard is high

108

## General Analysis Models

.....  
Time-varying Covariates

109

## Fixed Covariates

.....

In a typical study, we compare the distribution of some outcome across groups defined at the start of the study

- Example: Risk of hang gliding
  - Identify two groups
    - Hang gliders
    - Cowards
  - Follow survival experience over time

110

## Problem

.....

What if a coward obtains courage?

- Misclassification will attenuate the true effect of hang gliding on survival
  - Biased estimates
  - Less precision

111

## A Wrong Approach

.....

We cannot divide the sample into groups according to lifetime habits

- Suppose we consider
  - Ever hang glided (hung glide?) vs Constant coward
- We might detect spurious associations due to “survivorship”
  - If we started study at birth, we might find hang gliding is beneficial
    - Most people don’t start hang gliding until teenaged
    - We would detect the fact that hang gliders survived at least that long

112

## A Correct Approach

Let each subject contribute observation time to the appropriate group according to covariate at the relevant time

- Proportional hazards model
  - Easily done, if noninformative censoring results
- Accelerated failure time model
  - Difficult due to need to integrate hazards over disjoint intervals

113

## Issues

Issues related to the use of time-varying covariates are analogous to those when deciding to adjust for any variable

- Can regard measurements made at different times as different covariates
- Need to consider
  - Causal pathway of interest
  - Confounding (bias)
  - Precision
- Time aspect does increase the dimensionality

114

## Issues: Informative Censoring

Possibility that impending event causes informative censoring (confounding?)

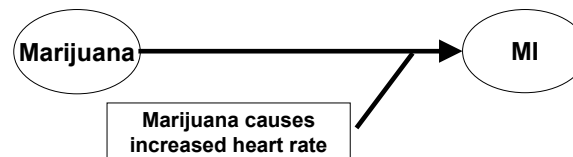
- Types of variables
  - Extrinsic: Unaffected by individual decisions
    - As a rule, time-varying extrinsic variables will not cause informative censoring
    - E.g., Air pollution on a given day in an asthma study
      - » (providing it does not affect relocation)
  - Intrinsic: Potentially affected by impending event
    - E.g., Marijuana use

115

## Causation versus Association

Example: Scientific interest in causal pathways between marijuana use and heart attacks (MI)

- Pictorial representation of hypothetical causal effect of marijuana on MI that might be of scientific interest



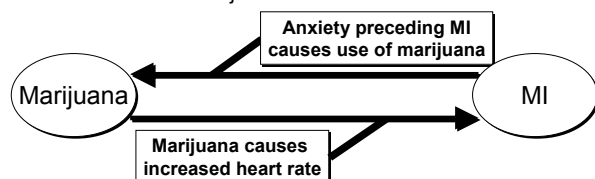
116

## Causation versus Association

.....

In an observational study, we cannot thus be sure which causative mechanism an association might represent

- Either of these mechanisms will result in an association between marijuana use and MI



117

## Issues: Obscuring Effect of Interest

.....

With time-varying covariates, we have increased opportunity to measure short term effects

- This is good if that is our interest
  - Immediate effects of blood pressure on hemorrhagic stroke
- This is bad if we wanted to assess long acting risk factors
  - Chronic effect of asbestos on lung cancer
    - A former asbestos worker is still at high risk

118

## Issues: Causal Pathway of Interest

.....

Capability for modeling time-varying covariates also increases chances for modeling a variable in the causal pathway of interest

119

## Issues: Summary Measure

.....

As illustrated previously, the interpretation of some of the statistics commonly used in survival analysis is heavily dependent upon the censoring distribution

- It is very difficult to explore how the changing size of risk sets might be altering the interpretation of the time-averaged hazard ratio in a proportional hazards model

120

## Issues: Final Comments

.....

Time-varying covariates are definitely of scientific interest

However, they should not be used casually

- Usually, my first choice is to try to address scientific questions with fixed covariates
  - I will put up with some misclassification, to avoid making mistakes that are due to incorrect, untestable assumptions

121

## General Analysis Models

.....

Choice of Summary Measures  
Used for Inference

122

## Summarizing Effect

.....

Based on marginal distributions

- Difference / ratio of means (arithmetic, geometric, ...)
- Difference / ratio of proportion exceeding some threshold
- Difference / ratio of medians (or other quantiles)
- Ratio of odds of exceeding some threshold
- Ratio of hazard (averaged across time?)
- ...

Based on joint distribution

- Median difference / ratio of paired observations
- Probability that a randomly chosen measurement from one population might exceed that from the other
- ...

123

## Statistical Models

.....

Options for inference

- Parametric models
  - Weibull, lognormal, etc.
- Semiparametric models
  - Proportional hazards, etc.
- Nonparametric
  - Weighted rank tests: logrank, Wilcoxon, etc.
  - Comparison of Kaplan-Meier estimates

124

## (Semi)parametric vs Nonparametric

Choice of statistical model can affect

- Computational methods for estimating the summary measure
- Precision of summary measure estimates
- Robustness of inference about the summary measure
- Ability to estimate the summary measure

125

## General Analysis Models

### Probability Models

126

## Right Censored Data

Notation:

Unobserved :

True times to event :  $\{T_1^0, T_2^0, \dots, T_n^0\}$

Censoring Times :  $\{C_1, C_2, \dots, C_n\}$

Observed data :

Observation Times :  $T_i = \min(T_i^0, C_i)$

Event indicators :  $D_i = \begin{cases} 1 & \text{if } T_i = T_i^0 \\ 0 & \text{otherwise} \end{cases}$

127

## Probability Distributions

Failure time  $T > 0$  measures time to an event :

Cumulative distribution function :  $F(t) = \Pr(T \leq t)$

Survivor function :  $S(t) = \Pr(T > t) = 1 - F(t)$

Density :  $f(t) = \frac{d}{dt} F(t)$

Hazard function :  $\lambda(t) = \lim_{h \downarrow 0} \frac{\Pr(T \in [t, t+h) | T \geq t)}{h}$

Cumulative hazard function :  $\Lambda(t) = \int_0^t \lambda(u) du$

Censoring variable distribution :  $C$  has cdf  $G(\cdot)$ ; pdf  $g(\cdot)$

128



## Parametric Models

$F$  is known up to some finite dimensional parameter vectors

$$F(t) = \Psi(t, \vec{\Phi})$$

where :

$\Psi(\cdot, \cdot)$  has known form

$\vec{\Phi}$  is finite dimensional and unknown

129

## Parametric Survival Models

Commonly used parametric survival models are generally accelerated failure time models

- Exponential
- Weibull
- Gamma
- Lognormal
- Log logistic
- Families joining several of the above

130

## Parametric Inference

Parametric inference generally proceeds through likelihood methods

- MLE found by Newton-Raphson iteration
- Asymptotic distributions from theory of regular problems

$$L(\vec{\Phi}; \vec{T}, \vec{D}) = \prod_{i=1}^n \left( f(T_i; \vec{\Phi}) (1 - G(T_i)) \right)^{D_i} \left( S(T_i; \vec{\Phi}) g(T_i) \right)^{1-D_i}$$

131

## Parametric Summary Measures

Mean : 
$$\hat{\theta} = \int_0^{\infty} u f(u; \hat{\Phi}) du$$

Median : 
$$\hat{\theta} = F^{-1}(0.5; \hat{\Phi})$$

Proportion above threshold : 
$$\hat{\theta} = \int_a^{\infty} f(u; \hat{\Phi}) du$$

Weighted average of hazard : 
$$\hat{\theta} = \int_0^{\infty} w(u) \lambda(u; \hat{\Phi}) du$$

132

## Parametric Models: Issues

### Advantages

- Can estimate any of the summary measures
- Can handle sparse data

### Disadvantages

- Not robust to other distributions
  - Parametric estimates with censoring do not generally have easy nonparametric interpretation
    - E.g., lognormal model is not particularly robust
- Little reason to suggest particular distribution
  - But motivation does exist for Weibull and Gamma

133

## Semiparametric Models

Exact form of within group distributions are unknown, but related to each other by some finite dimensional parameter vector

- Full inference only for comparing distributions
- One group's distn can be found from another group's and a finite dimensional parameter
- (Most often: Distributions equal under  $H_0$ )

(My definition of semiparametric models is a little stronger than some statisticians', but agrees with commonly used semiparametric survival models)

134

## Semiparametric Models: Notation

For group  $k$ :  $F_k(t) = \Psi(t, \vec{\Phi}_k)$

where:

- $\Psi(\cdot, \cdot)$  has unknown form (in  $t$ )
- $\vec{\Phi}_0 = \vec{0}$  for identifiability of  $\Psi(\cdot, \cdot)$
- $\vec{\Phi}_k$  is finite dimensional and unknown  
(estimable by comparing two or more groups)

135

## Semiparametric Survival Models

Accel failure:  $F_k(t) = F_0(t\theta_k)$

Prop hzd:  $S_k(t) = [S_0(t)]^{\theta_k}$

where in a regression problem

$$g(\theta_k) = \vec{X}_k^T \vec{\beta}$$

for some link function  $g(\cdot)$

136

## Semiparametric Inference

.....

Semiparametric inference generally proceeds through estimating equations

- Estimates found by iterative search
- Asymptotic distributions from special theory

137

## PH Partial Likelihood

.....

Proportional hazards regression based on hazard of observed failure relative to sum of hazards in the risk set

$$S_i(t) = [S_0(t)]^{\theta_i}; \lambda_i(t) = \theta_i \lambda_0(t) \quad \text{where } \log(\theta_i) = \bar{X}_i^T \bar{\beta}$$

$$\text{Partial likelihood: } L(\bar{\beta}) = \prod_{i=1}^n \left[ \frac{\lambda_i(T_i)}{\sum_{j:T_j \geq T_i} \lambda_j(T_i)} \right]^{D_i} = \prod_{i=1}^n \left[ \frac{\theta_i}{\sum_{j:T_j \geq T_i} \theta_j} \right]^{D_i}$$

138

## Semiparametric Summary Measures

.....

Estimation of summary measures is generally limited to the parameter fundamental to the semiparametric model

- Proportional hazards
  - Only make inference about hazard ratio
- Accelerated failure time
  - Only make inference about ratio of quantiles

139

## Semiparametric Models: Issues

.....

### Advantages

- Can handle sparse data
- More robust than any single parametric model

### Disadvantages

- Not easily interpreted when semiparametric model does not hold
- Little reason to suggest a given risk factor would affect distribution in only one way

140

## A Logical Disconnect

.....

“Because the light is so  
much better  
here under the streetlamp”

- a drunk looking for the keys  
he lost half a block away

141

## Inflammatory Assertion

.....

(Semi)parametric models are not typically in keeping with the state of knowledge as an experiment is being conducted

- The assumptions are more detailed than the hypothesis being tested, e.g.,
  - Question: How does the intervention affect the first moment of the probability distribution?
  - Assumption: We know how the intervention affects the 2nd, 3rd, ...,  $\infty$  central moments of the probability distribution.

142

## The Problem

.....

Incorrect parametric assumptions can lead to incorrect statistical inference

- Precision of estimators can be over- or understated
  - Hypothesis tests do not attain the nominal size
- Hypothesis tests can be inconsistent
  - Even an infinite sample size may not detect the alternative
- Interpretation of estimators can be wrong

143

## (Semi)parametric Example

.....

Survival cure model (Ibrahim, 1999, 2000)

- Probability model
  - Proportion  $\pi_i$  is cured (survival probability 1 at  $\infty$ ) in the  $i$ -th treatment group
  - Noncured group has survival distribution modeled parametrically (e.g., Weibull) or semiparametrically (e.g., proportional hazards)
  - Treatment effect is measured by  $\theta = \pi_1 - \pi_0$
- The problem as I see it: Incorrect assumptions about the nuisance parameter can bias the estimation of the treatment effect

144

## Foundational Issues: Null

.....

Which null hypothesis should we test?

- The intervention has no effect whatsoever

$$H_0 : F(t) = G(t), \forall t$$

- The intervention has no effect on some summary measure of the distribution

$$H_0 : \theta = \theta_0$$

145

## Foundational Issues: Alternative

.....

What should the distribution of the data under the alternative represent?

- Counterfactual
  - An imagined form for  $F(t)$ ,  $G(t)$  if something else were true
- Empirical
  - The most likely distribution of the data if the alternative hypothesis about      were true

146

## My Views

.....

The null hypothesis of greatest interest is rarely that a treatment has no effect

- Bone marrow transplantation
- Women's Health Initiative
- National Lung Screening Trial

The empirical alternative is most in keeping with inference about a summary measure

147

## An Aside

.....

The above views have important ramifications regarding the computation of standard errors for statistics under the null

- Permutation tests (or any test which presumes  $F=G$  under the null) will generally be inconsistent

148

## Problem with (Semi)parametrics

.....

Many mechanisms would seem to make it likely that the problems in which a fully parametric model or even a semiparametric model is correct constitute a set of measure zero

- Treatments are often directed to outliers
- Treatments are often only effective in subsets
- Factors affect rates; outcomes measure cumulative effects

149

## A Non-Solution: Model Checking

.....

Model checking is apparently used by many to allow them to believe that their models are correct.

- From a recent referee's report:
  - “I know of no sensible statistician (frequentist or Bayesian) who does not do model checking.”
- Apparently the referee believes the following unproven proposition:
  - If we cannot tell the model is wrong, then statistical inference under the model will be correct

150

## A Non-Solution: Model Checking

.....

Counter example: Exponential vs Lognormal medians

- Pretest with Kolmogorov-Smirnov test (n=40)
  - Power to detect wrong model
    - 20% (exp); 12% (lnorm)
  - Coverage of 95% CI under wrong model
    - 85% (exp); 88% (lnorm)

151

## A Non-Solution: Model Checking

.....

Model checking particularly makes little sense in a regulatory setting

- Commonly used null hypotheses presume the model fits in the absence of a treatment effect
  - Frequentists would be testing for a treatment effect as they do model checking
- Bayesians should model any uncertainty in the distribution
  - Interestingly, if one does this, the estimate indicating parametric family will in general vary with the estimate of treatment effect

152

## Nonparametric Models

Form of  $F$  is completely arbitrary and unknown within groups

- The summary measure measuring factor effect is just some difference between distributions
- The summary measure is estimated nonparametrically
  - (preferably within groups and then compared across groups)

153

## Comparison of Summary Measures

Typical approaches to compare response across two treatment arms

- Difference / ratio of means (arithmetic, geometric, ...)
- Difference / ratio of medians (or other quantiles)
- Median difference of paired observations
- Difference / ratio of proportion exceeding some threshold
- Ratio of odds of exceeding some threshold
- Ratio of instantaneous risk of some event
  - » (averaged across time?)
- Probability that a randomly chosen measurement from one population might exceed that from the other
- ...

154

## Nonparametric Summary Measures

Nonparametric: Estimate summary measures from nonparametric empirical distribution functions

- E.g., use sample median for inference about population medians
- In the presence of censoring, use estimates based on Kaplan-Meier estimates
- Often the nonparametric estimate agrees with a commonly used (semi)parametric estimate
  - Interpretation may depend on sampling scheme
  - In this case, the difference will come in the computation of the standard errors

155

## Nonparametric Summary Measures

Using Kaplan - Meier survival estimate  $\hat{S}(t)$

Mean : 
$$\hat{\theta} = \int_0^{\infty} \hat{S}(u) du$$

Median : 
$$\hat{\theta} = \hat{S}^{-1}(0.5)$$

Proportion above threshold : 
$$\hat{\theta} = \hat{S}(a)$$

Weighted average of hazard : 
$$\hat{\theta} = \int_0^{\infty} w(u) \hat{\lambda}(u) du$$

156

## Nonparametric Summary Measures

.....

Depending on the censoring scheme, not all summary measures are estimable

- The support of the censoring distribution may preclude estimation of the mean and some quantiles
- Can instead use the mean of the truncated distribution
  - “Average increase in days alive during first 5 years”

Mean of truncated distribution :  $\hat{\theta} = \int_0^a \hat{S}(u) du$

157

## Inference

.....

In most cases, variance estimates can be obtained from the asymptotic theory of the Kaplan-Meier estimates

- There are still some issues to be solved
  - Regression modeling needs to be worked out
  - Software is not readily available (Why not?)

158