SURVIVAL ANALYSIS

A primary objective of a phase II trial is to screen for antitumor activity; agents which are found to have substantial antitumor activity and an appropriate spectrum of toxicity are likely incorporated into combinations to be evaluated for patient benefit in controlled phase III trials.

The secondary aim is often concerned patient survivorship.

Therefore, in addition to estimating the response rate and performing the Chi-square test (one-arm or two-arm trials); other data analyses for phase II trials may involve: (1) Logistic Regression, & (2) Kaplan-Meier Survival Curve & maybe (3) Comparison of 2 survival curves, and (4) Cox's Proportional Hazards Regression.

The reviews of Logistic Regression were in session #2; therefore, this review session #3 is devoted to Survival Analysis. The more advanced part, competing risks, is in the last analysis review #4.

If “survival” is of high priority, “overall survival” (endpoint is death by any cause) is recommended as the endpoint. Use of “time to death due to the disease” may be problematic. Even if cause of death information is reliable, one can never safely assume the independence of causes of death; In addition, cause-specific endpoint does not include all aspects of treatment on survival.

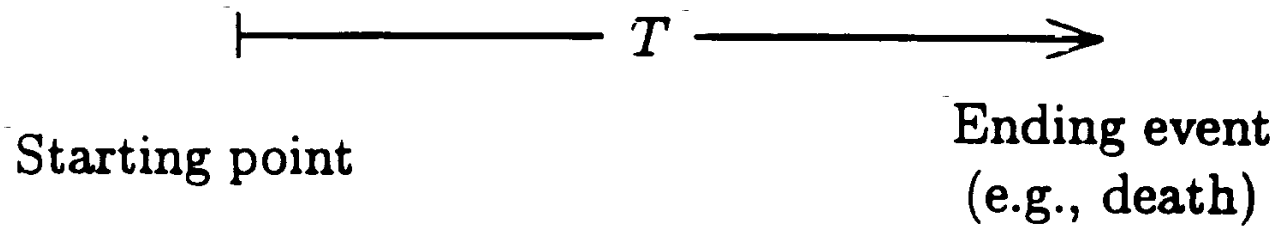
Disease-free survival (endpoint is death or relapse) is often another good choice.

Use of “progression-free survival” is also biased because follow-up on progression is often discontinued after a patient is taken off the study.

To focus on “overall survival” or “disease-free survival”, the most common task is estimating the survival curve, called the “Kaplan-Meier Curve” and some typical “survival rates” – such as “one-year survival rate”.

Basic Outcome: **SURVIVAL TIME**

1. Time origin or starting point
2. Ending event of interest
3. Measurement scale for the passage of time



The “Survival Time” T is from “Enrollment” (Starting Point) to “Death” (if focusing on “Overall Survival” or to “Death or Relapse” (if focusing on “Disease-free Survival” (the Ending Event of Interest) measured in “Years” or “Months” (measurement scale).

SURVIVAL FUNCTION

- Basic Functions to characterize a Survival Time T are: (1) the Survival Function S(t) , and (2) the Hazard Function.
- Survival Function is defined as:

$S(t) = Pr(T>t)$; for $t>0$

$= 1 - \text{cdf } F(t)$

(Emphasize the “good news”).

- It’s relatively easy to estimate – even in the presence of censoring.

HAZARD (OR RISK) FUNCTION

- Hazard, or Risk, Function gives the "Instantaneous Failure Rate" at time "t" assuming survival at "t":

$$\lambda(t) = \lim_{\delta \to 0} \frac{\Pr[t \le T < t + \delta \mid t \le T]}{\delta}$$

- For a small time increment δ, the Probability that an event occurs during the time interval [t,t+δ) is, approximately:

$$\Pr[t \le T < t + \delta \mid t \le T] \cong \{\lambda(t)\}\{\delta\}$$

HAZARD FUNCTION

- The Hazard, Function measures the "proneness" to failure as a function of "age"
- It may increase or decrease with time for long-term or short-term risks, or remains constant- even follows a more complicated process. For example, hazard associated with acute leukemia increases, hazard for organ transplant & AIDS decreases; maybe non-monotonic (e.g. Hodgkin's disease).
- "Bathtub" hazard curve for human life.

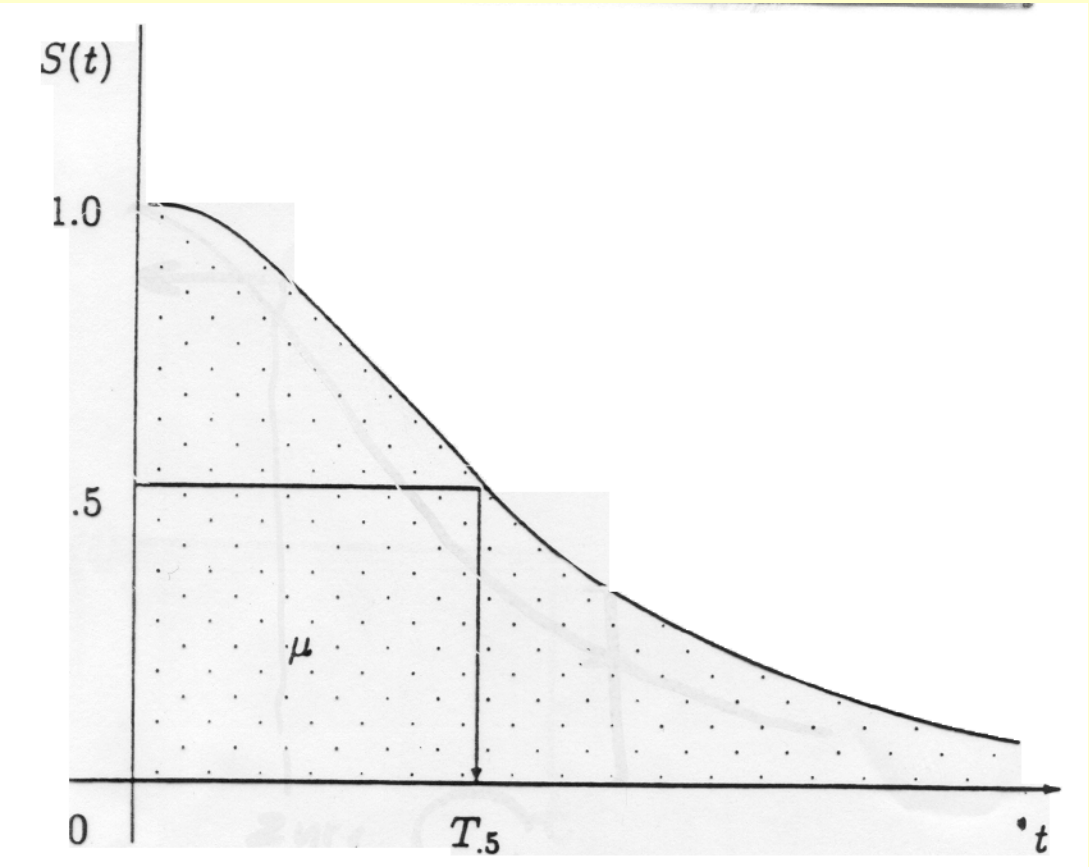
DIFFERENCES

- Survival Function: “How far” to event (for the population)
- Hazard Function: “How fast” (approaching the ending event)- like “velocity”. Also called “Risk Function”.
- Hazard Function is defined similar to the Density Function but “conditional”; the Hazard Function can tell you more, but more difficult to estimate- compared to S(t).

Survival Function

&

Survival Curve



The Survival Function id defined as: S(t) = Pr(T>t); if "t" is in years, S(t) is called "t-year survival rate". For example, S(2) is the 2-year survival rate (%); S(5) or 5-year survival rate is a conventional benchmark in Cancer Survival.

BONE-MARROW TRANSPLANT: ALLOGENIC Versus AUTOLOGOUS

Allogenic:
From a Matched Donor

Autologous:
From Self

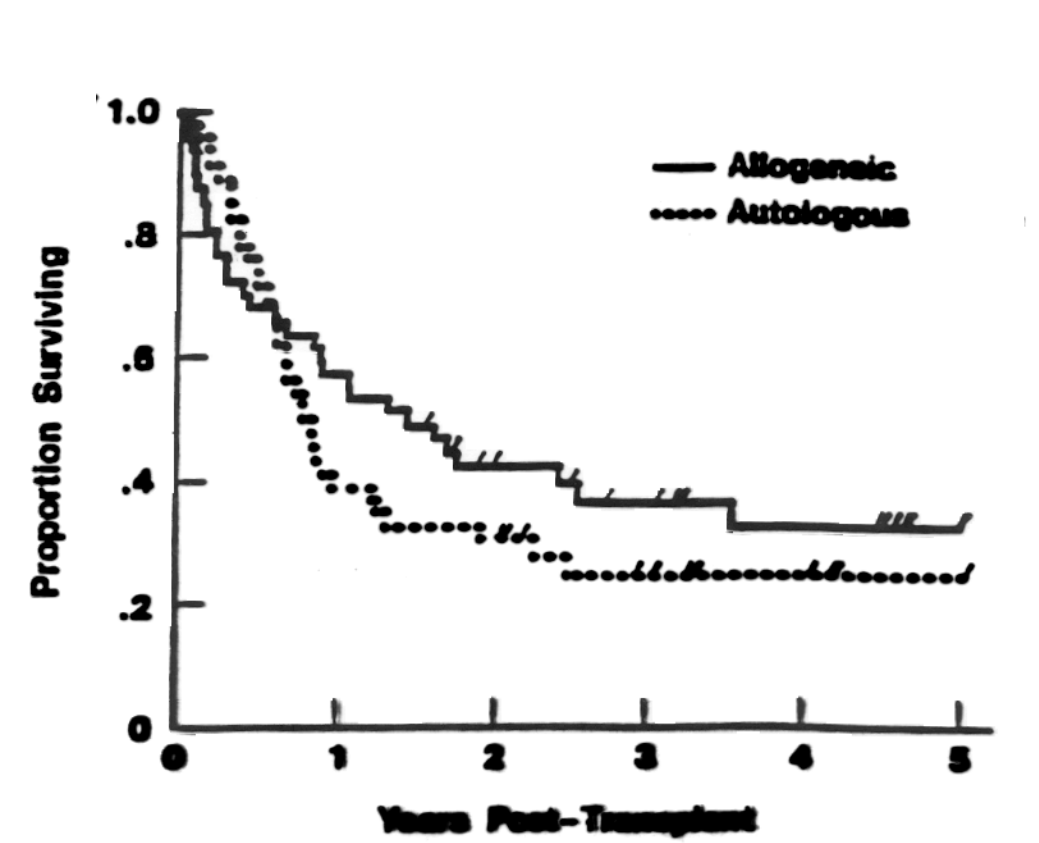


Figure 4. Cumulative Proportions of Patients Alive, According to the Number of Years after Transplantation
(P = 0.32, Mantel-Cox).

For Phase II trials, which are often in shorter term lasting just a year or two, one-year survival rate is often used a bench mark to measure treatment success (e.g. versus historical data).

A DISCRETE MODEL

- A discrete model for survival time T is defined as one taking values $x_1<x_2< \ldots <x_k$ with associated (positive) probability mass: $f(x_i) = Pr(T = x_i) > 0$; and $f(x) = 0$ elsewhere
- The graph is a "Step Curve" with "jumps" at these support points: $x_1<x_2< \ldots <x_k$.
- **Note that**: $S(x_i) = Pr(T>x_i)$ & $S(x_i^-) = S(x_{i-1})$; at each vertical "drop" in the graph, the point is at lower level

SURVIVAL CURVE
FOR
A DISCRETE MODEL

Note: At each vertical "drop" in the graph, the point is at lower level

Value at each drop is equal to the value of the density function at that point.

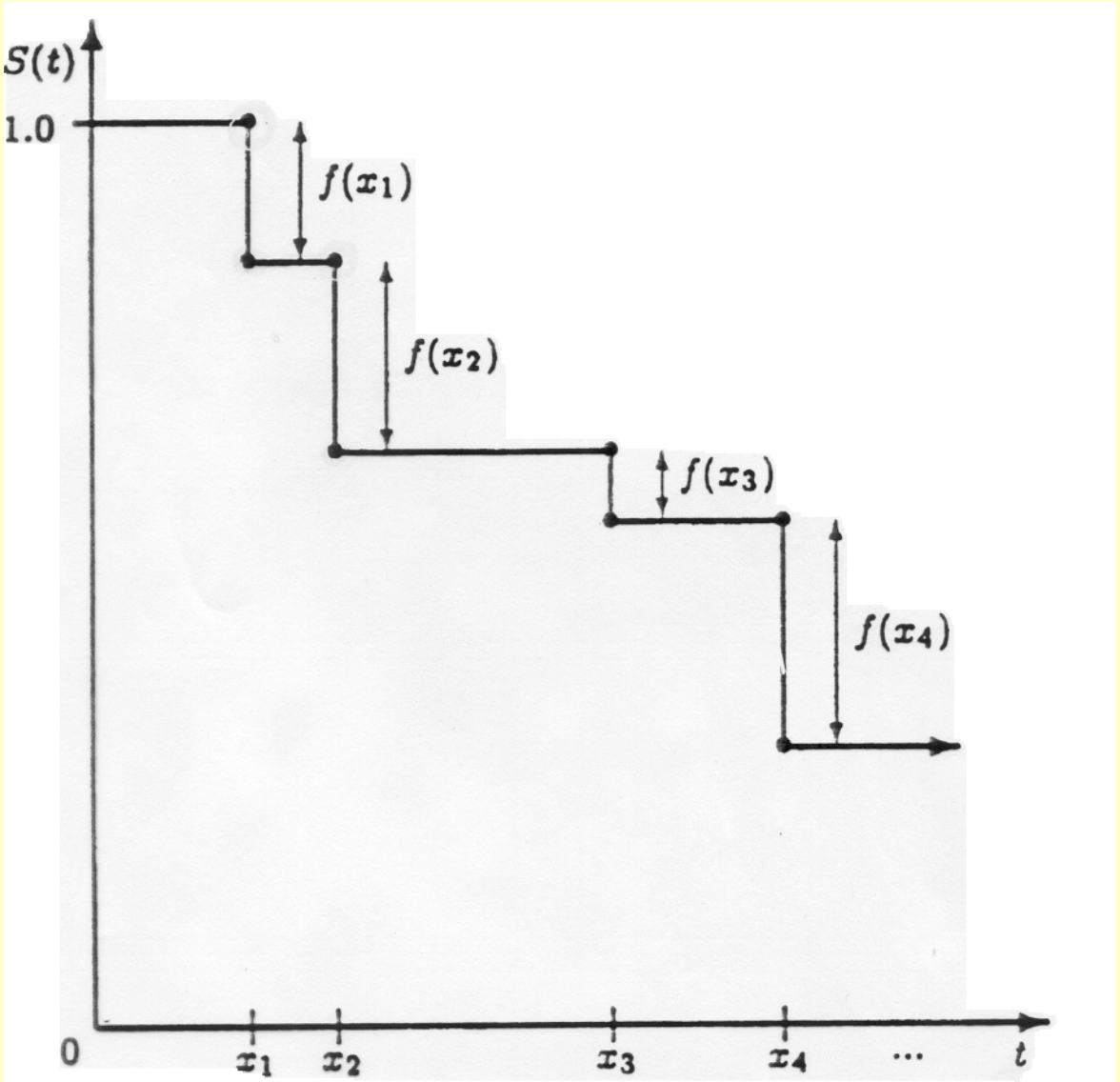


FIGURE 1.7. $S(t)$ is a step function for a discrete T

DISCRETE MODEL

$$\lambda_i = \Pr[T = x_i \mid T \geq x_i] = \frac{f(x_i)}{S(x_i^-)}$$

$$f(x_i) = \lambda_i \prod_{j=1}^{i-1} (1 - \lambda_j)$$

$$S(x_i) = \prod_{j=1}^{i} (1 - \lambda_j)$$

SURVIVAL DATA: An Example

- Forty-two (42) leukemia pts in remission, after chemotherapy, were randomized to receive either Placebo or 6-mercaptopurine (6-MP) in a 52-week trial w/o follow-up. Ending event is “relapse”; times in wks are:
- Placebo:1,2,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12 ,15,17, 22,23 (all 21 patients have relapsed).
- 6-MP: 6,6,6,7,10,13,16,22,23,6+,9+,10+,11+, 17+,19+,20+,25+,32+,32+,34+,35+.

ABOUT DATA IN THE EXAMPLE

- Patients with time “t+” have had no relapse yet! e.g., “10+” was from a patient who enrolled on week 42 and was still in remission.
- This phenomenon is typical of survival data, called “censoring”; the study was terminated before some subjects have their ending event (censored observation). All methods learned, including sample mean & t-tests, do not apply because some subjects only provide “partial information” (e.g. time is more than 10 weeks)

A CLINICAL TRIAL

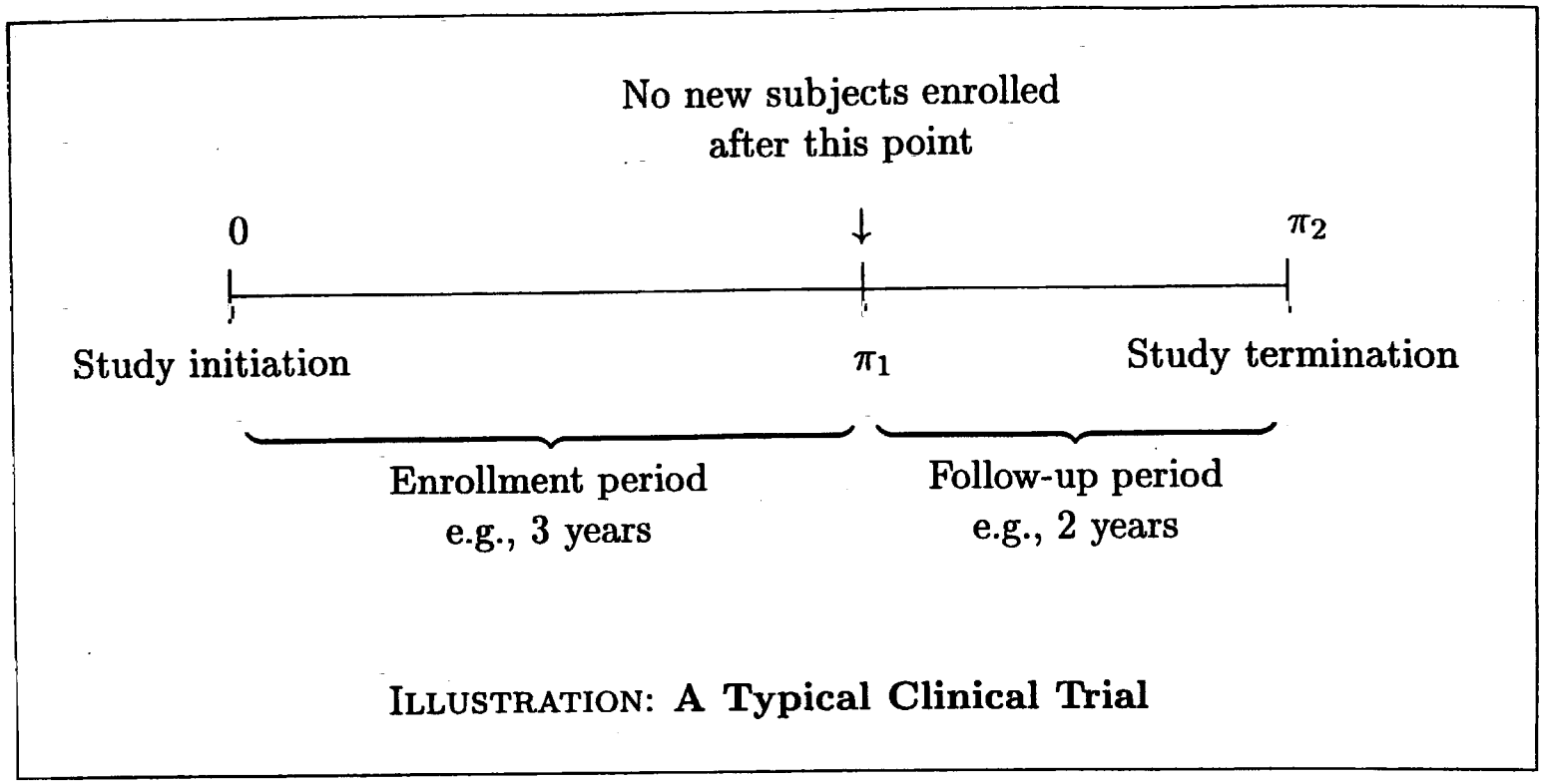


ILLUSTRATION: **A Typical Clinical Trial**

Study subjects without events when trial is terminated were "censored"

THE CONCEPT "At Risk"

- Consider the sample: {7, 9+, 10, 12, 25+}. At t = 10, for example, 3 subjects are "at risk" (those with duration equal 10, 12, & 25; the one with duration 7 "died" before 10 and the one with duration 9 "was censored" before 10 so they are "not at risk" at t=10. In other words, subjects "at risk" at t are those having been followed t units of time or longer.
- Actually, they are at risk at time t^-. In the sample: {7,9+,10,10+,12,25+}; there are four (4) subjects at risk at t=10 because we count "deaths" and censored cases at t among those "at risk" at t^-.

The "risk set" R_i at time t_i consists of subjects "at risk" at t_i^- ; the number of subjects is n_i, these are under observation at time t_i^- . The "death set" or "event set" D_i consists of those with events at time t_i (it is a subset of the risk set); the number of subjects is d_i, these died at t_i.

KAPLAN-MEIER, or PRODUCT-LIMIT METHOD

- The Kaplan-Meier or also called Product-Limit method is the method that generalizes the Empirical Estimate for use with Censored Data.
- Approach: To assume the Discrete Model

$$f(t_i) = \lambda_i \prod_{j=1}^{i-1} (1-\lambda_j) \quad \& \quad S(t_i) = \prod_{j=1}^{i} (1-\lambda_j)$$

- Then use "MLE" method to estimate the λ_i's where λ_i's are the hazards at times with event(s)

The Likelihood Function is :

$$\prod_{i=1}^{n}[f(t_i)]^{\delta_i}[S(t_i)]^{1-\delta_i} = \prod_{i=1}^{m}\lambda_i^{d_i}(1-\lambda_i)^{n_i-d_i}$$

where $\delta_i = 1$ for event and $\delta_i = 0$ if censored

KAPLAN-MEIER ESTIMATE

Results:

$$\hat{\lambda}_i = \frac{d_i}{n_i} \qquad \hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right)$$

Notes:

- The Product is over all $t_i \le t$; "Step-function"
- At each vertical "drop" on the graph, the point is at lower level; "right-continuous"

EXAMPLE #1

- Consider a small sample of times (in years): {3,4,4,9,9}; n=5 and no censoring.

t_i	n_i	d_i	$1-(d_i/n_i)$	$\hat{S}(t_i)$
3	5	1	.800	.8000
4	4	2	.500	.4000=(.8000)(.500)
9	2	2	0	0=(.4000)(0)

- Kaplan-Meier Method applies to data without censoring as well (why not?);
- Results are identical to empirical estimates.

EXAMPL #2: **Leukemia Data**

- 6MP:6,6,6,7,10,13,16,22,23,6+,9+,10+,11+, 17+,19+,20+,25+,32+,32+,34+,35+(n=21,d=9)

t_i	n_i	d_i	$1-(d_i/n_i)$	$\widehat{S}(t_i)$
6	21	3	.8571	.8571
7	17	1	.9412	.8067
10	15	1	.9333	.7529
13	12	1	.9167	.6902=(.7529)(.9167)
16	11	1	.9091	.6275
22	7	1	.8671	.5378
23	6	1	.8333	.4482

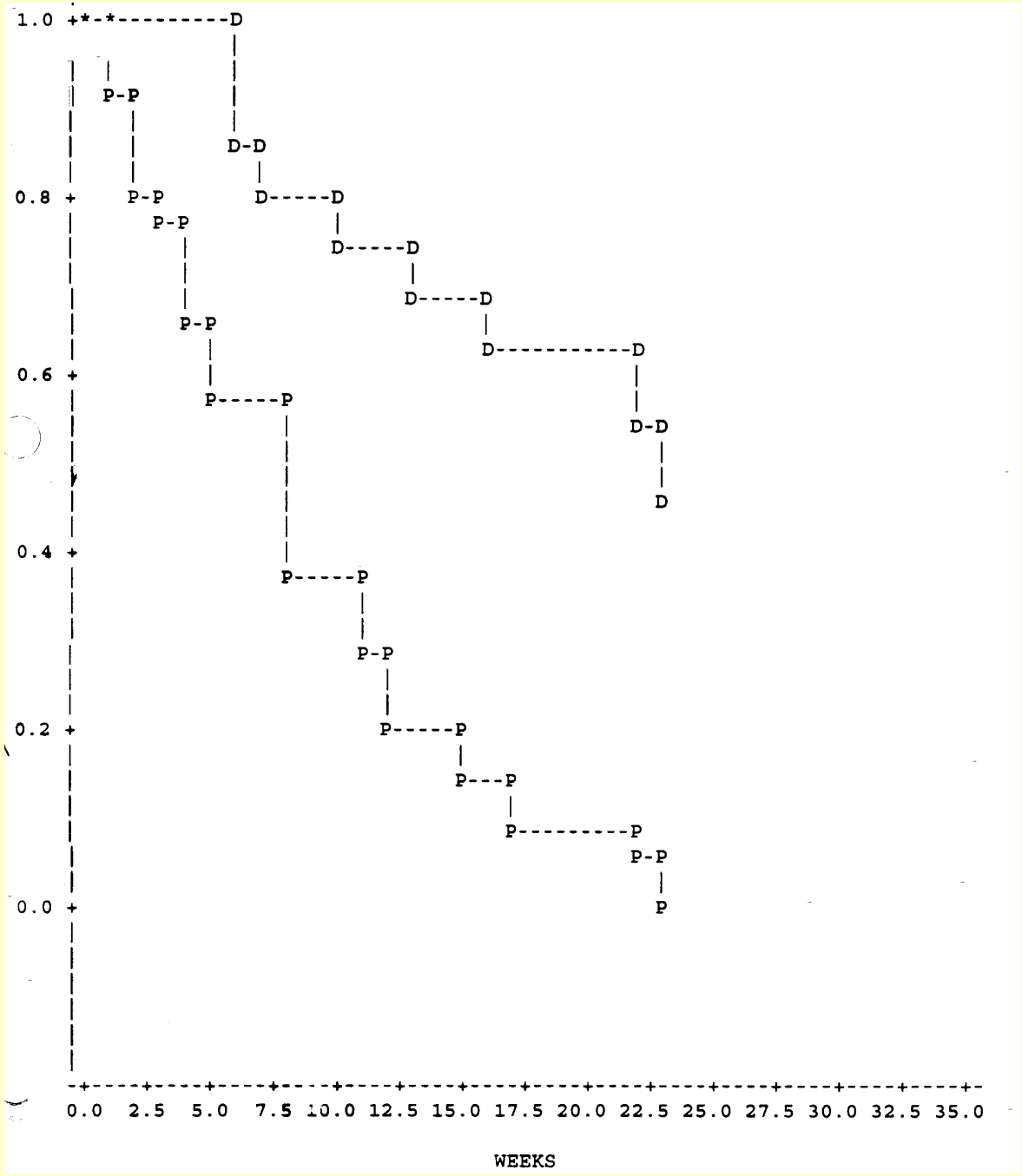
- (Example: Median is between 22 & 23)

D: Drug 6-MP

P: Placebo

RESULTS:

Medians are 23 and 8 for 6-MP and Placebo groups respectively.



BONE-MARROW TRANSPLANT: **ALLOGENIC Versus AUTOLOGOUS**

Allogenic:
From a Matched Donor

Autologous:
From Self

Event: Death; Same Pattern of Curves for Leukemia Relapse: Evidence of "Cure"!

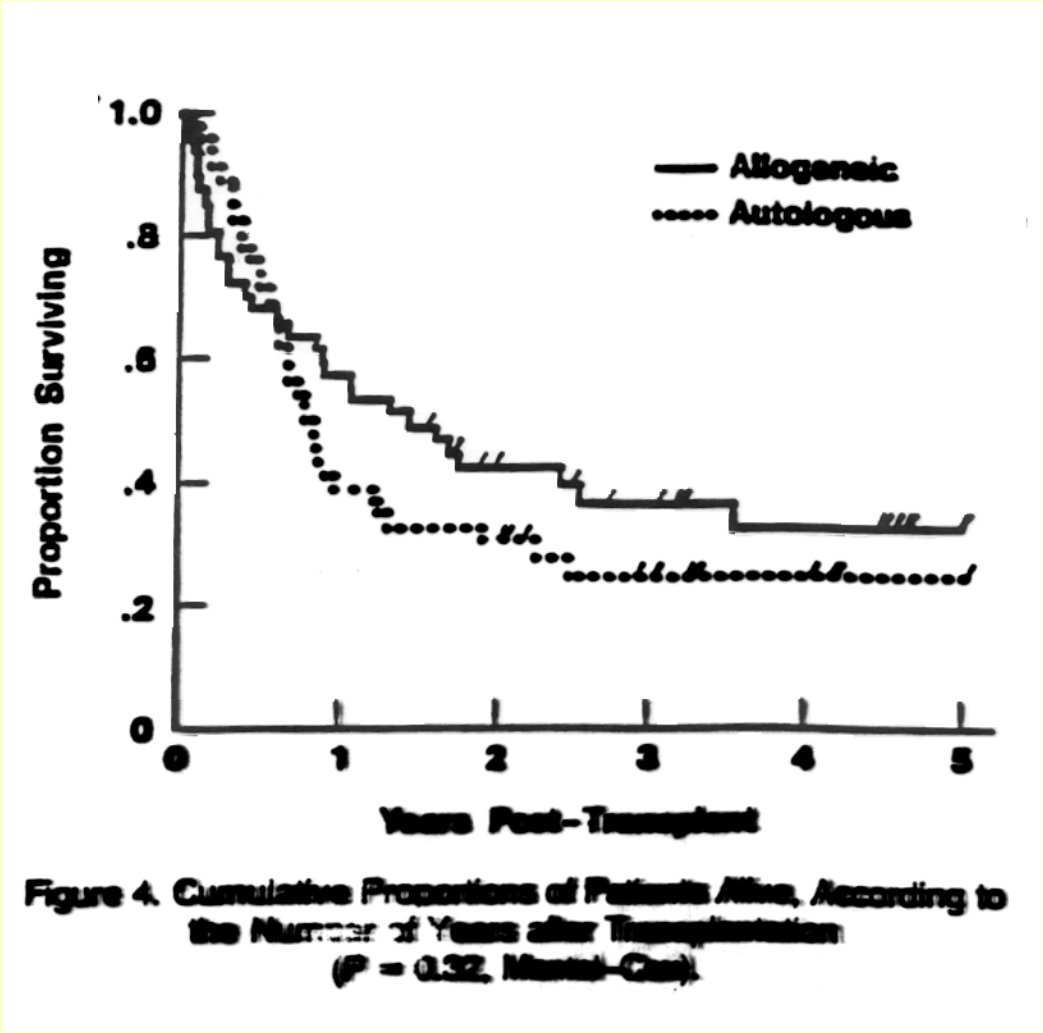


Figure 4. Cumulative Proportions of Patients Alive, According to the Number of Years after Transplantation
(P = 0.32, Mantel-Cox)

COMPARISON OF TWO SURVIVAL DISTRIBUTIONS

- Suppose we have n_1 and n_2 individuals receiving treatments 1 and 2, respectively, in a clinical trial. The study provides 2 samples of survival data: $\{t_{1i}, \delta_{1i}\}$, i from 1 to n_1 and $\{t_{2j}, \delta_{2j}\}$, j from 1 to n_2.
- The sample sizes n_1 & n_2 may or may not be equal
- The Null Hypothesis to be tested is:
  H_0: $S_1(t) = S_2(t)$
  Goal: to generalize the Wilcoxon test for use with censored data (Surprise: We'll get more!)

STARTING STEP

- At ordered time t_i, $1 \le i \le m$, the data may be summarized into a 2-by-2 table as follows:

	Status		
Sample:	Dead	Alive	Total
1	d_{1i}	a_{1i}	n_{1i}
2	d_{2i}	a_{2i}	n_{2i}
Total	d_i	a_i	n_i

n_{1i} , for example, is the number of subjects in group 1 who were "at risk" at time t_i (i.e. alive at t_i^-); subjects who died or were censored at t_i are included.

EXAMPLE #1A

- Sample #1:{3,5+,10} & Sample #2:{3,6,9+,15+}.
- We have: N=7, d=# of events=4, and m=# of time points = 3 at t = 3 (2 events) and 6, 10 (1 event each)
- At t=6, for example, we form a 2x2 table:

	Status		
Sample:	Dead	Alive	Total
1	0	1	1 ← {10}
2	1	2	3 ← {6,9+,15+}
Total	1	3	4

d_{1i}

BASIC APPROACH

- In this arrangement,

	Status		
Sample:	Dead	Alive	Total
1	d_{1i}	a_{1i}	n_{1i}
2	d_{2i}	a_{2i}	n_{2i}
Total	d_i	a_i	n_i

The Null Hypothesis of equal survival distributions implies the independence of "Status" and "Sample" in each cross-classified 2-by-2 table. In other words, the first row (sample 1) can be considered as "a random sample" from the last row (**Total**, which serves as a finite population).

BASIC APPROACH

	Status		
Sample:	Dead	Alive	Total
1	d_{1i}	a_{1i}	n_{1i}
2	d_{2i}	a_{2i}	n_{2i}
Total	d_i	a_i	n_i

When the first row (sample 1) can be considered as "a random sample" from the last row (Total, acts as finite population), we have from the "Hypergeometric Distribution":

$$E_0(d_{1i}) = \frac{n_{1i} d_i}{n_i}; Var_0(d_{1i}) = \frac{n_{1i} n_{2i} a_i d_i}{n_i^2 (n_i - 1)}$$

TARONE-WARE FAMILY

- The Test Statistic:

$$\boldsymbol{\theta} = \sum_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{w_i}[\mathbf{d_{1i}} - \frac{\mathbf{n_{1i}}\mathbf{d_i}}{\mathbf{n_i}}]; E_0(\theta) = 0$$

$$Var_0(\theta) = \sum_{i=1}^{m} \frac{w_i^2 n_{1i} n_{2i} a_i d_i}{n_i^2 (n_i - 1)}; z = \frac{\theta}{\sqrt{Var_0(\theta)}}$$

represents not "a Test" but "a Family of Tests" each indexed by a "weight" representing the relative importance of time point (relative to other time points). This family is called the "Tarone-Ware" family of tests.

CHOICES OF WEIGHTS FOR THE TARONE-WARE FAMILY

- How to choose the weight?

(1) The Generalized Wilcoxon, with $w_i = n_i$, puts more weight on the early observations and because of that its use is more powerful in detecting the effects of "short-term risk".

(2) The Log-Rank test, with $w_i = 1$, puts equal weight on each observation and, by default, is more sensitive to later differences (as compared to the Generalized Wilcoxon); in fact, it is most powerful under PHM (constant, long-term risks).

EXAMPLE #1B

- Sample #1:{3,5+,10} & Sample #2:{3,6,9+,15+}.
- We have: m=# of time points = 3 at t = 3, 6, 10.
-

t_i	n_i	d_{1i}	$E_0(d_{1i})$	$Var_0(d_{1i})$
3	7	1	.857	.408
6	4	0	.250	.188
10	2	1	.500	.250

- Generalized Wilcoxon:

$$z_W = \frac{(7)(1-.857)+(4)(0-.250)+(2)(1-.500)}{\sqrt{(49)(.408)+(16)(.188)+(4)(.250)}} = .204$$

- Log-Rank:

$$z_L = \frac{(1-.857)+(0-.250)+(1-.500)}{\sqrt{(.408)+(.188)+(.250)}} = .438$$

CHOICES OF WEIGHTS FOR THE TARONE-WARE FAMILY

- Which test should we use without apriori emphasis?
- May be it's not a bad idea to try both; the results would tell whether the two groups are different and, in addition, where the differences are: (i) if the generalized Wilcoxon is more significant, there are more differences at early times, (ii) if the Log-Rank test is more significant, there are more differences at later times.

TARONE-WARE FAMILY

- Because of the way the test statistic is formulated (terms in the sum are "not squared"):

$$\theta = \sum_{i=1}^{m} w_i [d_{1i} - \frac{n_{1i} d_i}{n_i}]$$

the tests in this family, regardless of the choices of weights, are only powerful when one group is "better" than the other all the times; otherwise, some terms in the sum are positive and others are negative, and some of them are cancelled out . For example, the tests are virtually powerless against "crossing curves" alternative.

EXAMPLE #2

- Forty-two (42) leukemia pts in remission, after chemotherapy, were randomized to receive either Placebo or 6-mercaptopurine (6-MP) in a 52-week trial w/o follow-up. Ending event is "relapse"; times in wks are:
- Placebo:1,2,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17, 22,23.
- 6MP: 6,6,6,7,10,13,16,22,23,6+,9+,10+,11+,17+,19+,20+,25+,32+,32+,34+,35+.

D: Drug 6-MP

P: Placebo

RESULTS:

(1)Log-Rank

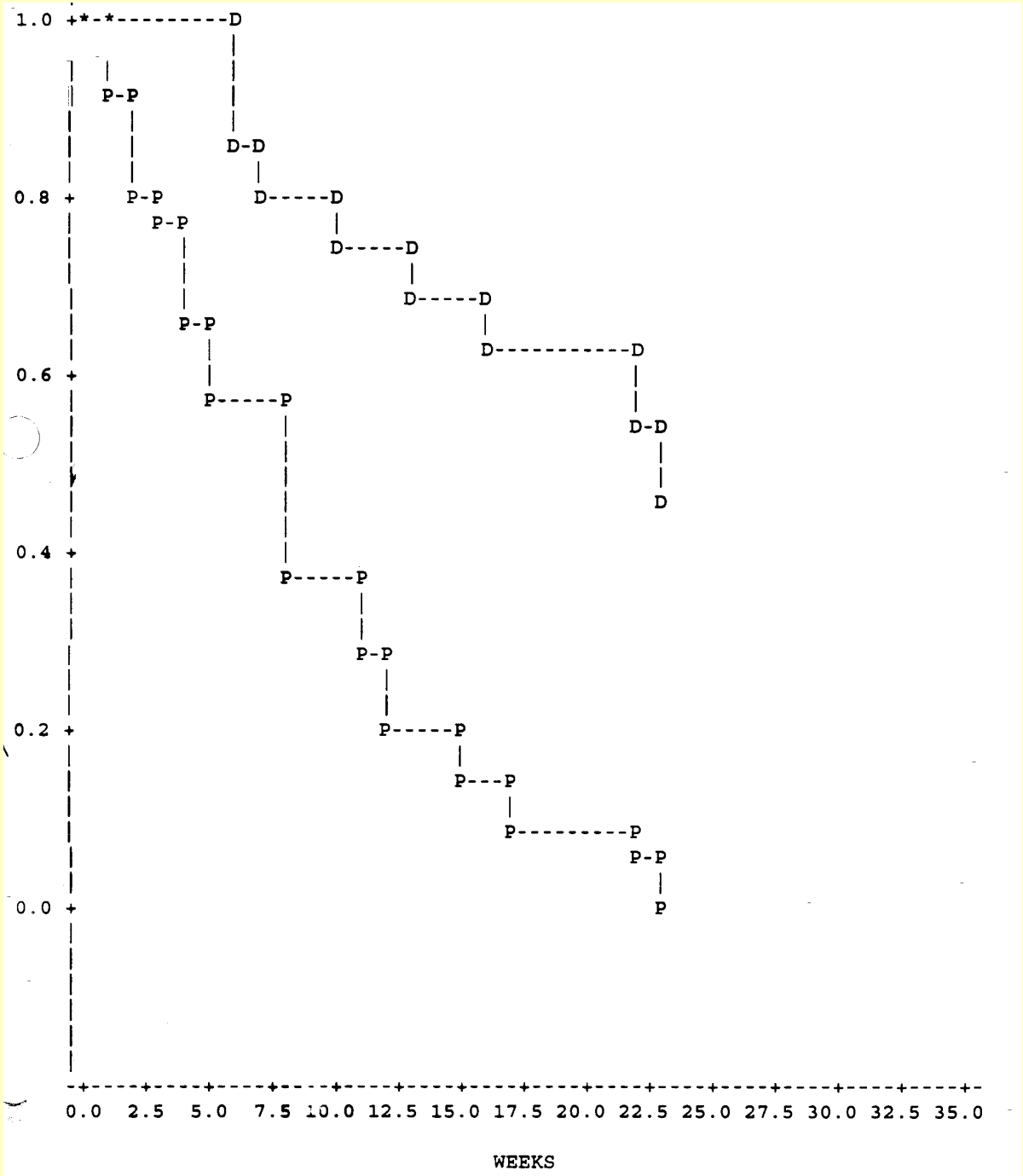
X^2=16.793,

p-value=.0001

(2)Wilcoxon

X^2=13.458,

p-value=.0002



BONE-MARROW TRANSPLANT: ALLOGENIC Versus AUTOLOGOUS

Allogenic:
From a Matched Donor

Autologous:
From Self

Event: Death; Same Pattern of Curves for Leukemia Relapse: Evidence of "Cure"!
Note: The Cox-Mantel test (i.e. Log-Rank test) seen under the graph.

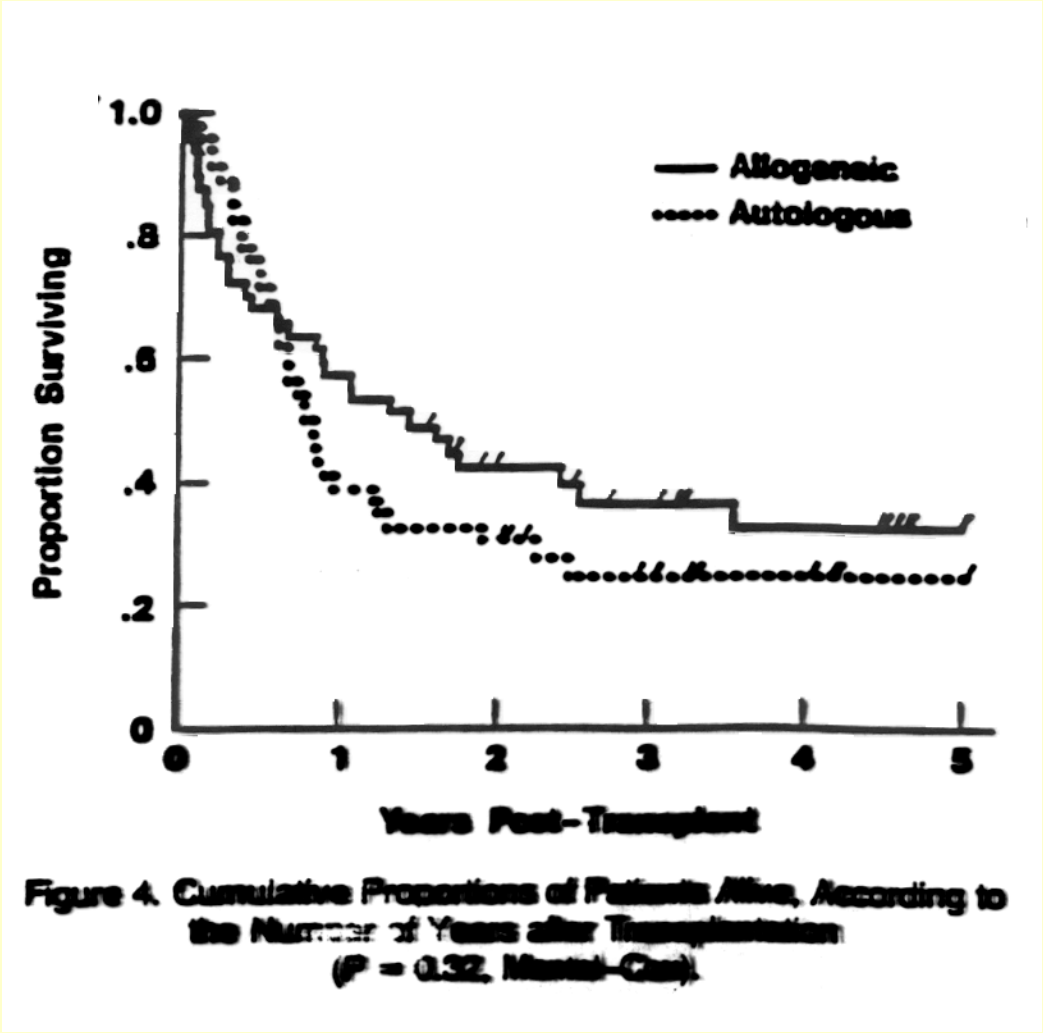


Figure 4. Cumulative Proportions of Patients Alive, According to the Number of Years after Transplantation (P = 0.32, Mantel-Cox)

Nonparametric methods are popular. However, there are instances where we "know" the distribution; in some special applications – e.g. in the sample size determination - we may assume it.

If we know the distribution of time to event, or we may reasonably assume it, say exponential, it would more efficient or powerful to use this distribution to form (parametric) tests of significance than to apply non-parametric tests.

REGRESSION ANALYSIS

- Primary focus is on estimating the response rate but sometimes we also want to see if some factor or factors might affect the "survival time"; such as age or gender of the patient, or other factors from the patient selection process dictated by "inclusion criteria" (and "exclusion criteria").
- It is important to identify and understand the influence of such factors in order to optimize conditions for using the treatment.

Response rates generally decrease as the extent of prior therapy increases. Patients who have failed several prior regimens are more likely to have tumors composed large numbers of resistant cells, and such patients are also less likely to be able to tolerate full doses of the investigational drug. The presence or absence of prior therapy is an important factor.

OTHER POSSIBLE GAINS

- It is important to identify both optimal and sub-optimal conditions.
- Insights can lead to modifications of inclusion/exclusion criteria.
- Even if we find, in research, that a factor does not influence the treatment's result, we may be able to relax the conditions under which the test is applied. (so, i.e. "negative findings" are good/beneficial as well)

REGRESSION MODEL

- Traditionally, in Regression, we "model'" the mean - In our survival data: area under the Survival Curve - as a function of "covariate(s)".
- The usual "Regression Model" takes the form: Observation = a function of covariate(s) + Error. The problem is that some of the observations were censored! This approach does not work unless we assume a Model for the "Error Term" - such as Exponential or Weibull.

NEWER APPROACH

- We know that the Survival Function is related to the Hazard Function.
- If an Exposure has some effect on the Mean of Survival Time (which is related to the Survival Function), it would have some effects on the Hazard Function as well.
- That means we can "model" the Hazard instead of modeling the Mean.

WEIBULL MODELS

- Weibull Model (with 2 parameters:p>0 & λ>0; λ is the "location" and p the "shape"):

$$\lambda(t) = \lambda p(\lambda t)^{p-1} \text{ for all } t > 0$$

$$S(t) = \exp[-(\lambda t)^{p}]$$

- Weibull Regression:

$$p_i = p \text{ (a constant)},$$

$$\lambda_i = \alpha e^{\beta x_i}$$

TERMINOLOGY

- “Relative Risk” is also called “Risks Ratio” (similar to “Odds Ratio”) or, simply, Hazards Ratio.
- Used to measure the effect of exposure, “exposed” vs. “non-exposed”; the “strength of the relationship”, very much like the coefficient of correlation.

AN EXAMPLE OF PROPORTIONAL HAZARDS

- For two Weibull's:

$$\frac{\lambda_1(t)}{\lambda_0(t)} = \frac{\lambda_1 p_1 (\lambda_1 t)^{p_1 - 1}}{\lambda_0 p_0 (\lambda_0 t)^{p_0 - 1}}$$

- So that the Hazards are proportional if and only if $p_1 = p_0 = p$ (same "shape parameter"); Under PHM:

$$\rho = \frac{\lambda_1(t)}{\lambda_0(t)} = \{\frac{\lambda_1}{\lambda_0}\}^p$$

- So, Proportional Hazards exist! (this is similar to the "equal variance" model for 2-sample t-test)

PHM:POSSIBLE GENERALIZATION

- The PHM for a binary risk: $RR(t) = \rho$ can be written as follows: $\lambda(t) = \lambda_0(t)e^{\beta x}$
- where $\lambda_0(t) = \lambda(t; E')$ the "baseline hazard" and the "covariate" X is defined as (=1 if exposed) and (=0 if not exposed); $\rho = e^{\beta}$
- First, we can extend the "model", the Proportional Hazard Model (PHM) from Binary Risk to cover any risk factor, discrete or continuous.
- Then we'll extend to make it multivariate:

$$\lambda(t) = \lambda_0(t)e^{\sum_{i=1}^{k}\beta_i x_i}$$

REGRESSION COEFFICIENT

- Consider a binary risk factor: (X=1 if exposed, X=0 if not exposed), and consider 2 subjects:

(1) Subject A: X=0, his/her hazard is $\lambda_A(t) = \lambda_0(t)$.

(2) Subject B: X=1, his/her hazard is $\lambda_B(t) = \lambda_0(t)e^{\beta}$.

- This ratio represents the "Relative Risk" of B versus A, or the "Relative Risk due to exposure":

$$RR(t) = \frac{\lambda_B(t)}{\lambda_A(t)} = \frac{\lambda_0(t)e^{\beta}}{\lambda_0(t)} = e^{\beta} = \text{constant}$$

- It explains the term "Proportional Hazards" and the regression coefficient β represents the Relative Risk on the log scale.

REGRESSION COEFFICIENT

- Consider a continuous risk factor X & 2 subjects:

(1) Subject A: X=x, his/her hazard is $\lambda_A(t)=\lambda_0(t)\, e^{\beta x}$.

(2) Subject B: X=x+1, his/her hazard is: $\lambda_B(t) = \lambda_0(t)e^{\beta(x+1)}$.

- This ratio represents the "Relative Risk" of B versus A, the "Relative Risk due 1 unit increase in X":

$$RR(t) = \frac{\lambda_B(t)}{\lambda_A(t)} = \frac{\lambda_0(t)e^{\beta(x+1)}}{\lambda_0(t)e^{\beta x}} = e^{\beta} = \text{constant}$$

- It explains the term "Proportional Hazards" and coefficient β represents the Relative Risk, on the log scale, due to "one unit increase in X".

ASSUMPTIONS OF PHM

- There actually are two (2) assumptions in PHM:

(1) Proportional Hazards: Ratio of hazards, or Relative Risk, is independent of time

(2) Linearity: For a continuous factor, the coefficient β represents the Relative Risk, on the log scale, due to "one unit increase in X":

$$RR(t) = \frac{\lambda_B(t)}{\lambda_A(t)} = \frac{\lambda_0(t)e^{\beta(x+1)}}{\lambda_0(t)e^{\beta x}} = e^{\beta} = \text{constant}$$

Note that this RR is independent of X=x, a result of the "linearity assumption" which may not be true.

CONDITIONAL APPROACH

- Denote the ordered distinct death times by t_i's: and let R_i be the risk set just before time t_i, n_i the number of subjects in R_i, D_i the dead set at t_i, d_i the number of subjects (i.e. deaths) in D_i. Let C_i be the collection of all possible combinations of subjects from R_i, each combination - a subset of R_i - has d_i members; D_i is itself one of these combinations.
- Example: $R_i = \{A,B,C\}$ and $D_i = \{A,B\}$; Then $n_i = 3$, $d_i = 2$, and $C_i = \{\{A,B\} = D_i, \{A,C\}, \{B,C\}\}$. The collection C_i has 3 members; in general:

$$N_i = \binom{n_i}{d_i}$$

APPROACH: COX'S ARGUMENT

- Cox writes (1972): "Suppose then that $\lambda_0(t)$ is arbitrary. No information can be contributed about β by time intervals in which no failures occur... We therefore argue conditionally on the set of instants at which failures occur…. Once we require a method of analysis holding for all $\lambda_0(t)$, consideration of this conditional distribution seems inevitable."

LIKELIHOOD FUNCTION

- With his "conditional argument", Cox proposed the following "Partial Likelihood Function":

$$L = \prod_{i=1}^{m} \Pr(D_i \mid R_i, d_i)$$

- We'll consider separately the cases without and with "ties" and illustrate using the following small example with 4 subjects: {A: (t = 4,x = 2, δ = 1), B: (t = 8,x = 5, δ = 1), C: (t = 8,x = 7, δ = 1), and D: (t=11,x=18, δ =0).

LIKELIHOOD FUNCTION

- Without ties, the component of the Partial Likelihood at t_i is:

$$L_i = \Pr(D_i \mid R_i, d_i) = \frac{\lambda_0(t_i)e^{\beta x_i}}{\sum_{j=1}^{n_i} \lambda_0(t_i)e^{\beta x_j}} = \frac{e^{\beta x_i}}{\sum_{j=1}^{n_i} e^{\beta x_j}}$$

- For our example {A:(t=4,x=2,δ=1), B:(t=8,x=5,δ=1), C:(t=8,x=7,δ=1), and D:(t=11,x=18,δ=0), at t=4, the component of the Partial Likelihood is:

$$L_{t=4} = \frac{e^{2\beta}}{e^{2\beta} + e^{5\beta} + e^{7\beta} + e^{18\beta}}$$

LIKELIHOOD FUNCTION

- With ties, component of the Partial Likelihood at t_i is:

$$L_i = \frac{\prod_{D_i} \lambda_0(t_i) e^{\beta x_i}}{\sum_{C_i} \prod_{D_u} \lambda_0(t_i) e^{\beta x_j}} = \frac{e^{\sum_{D_i} \beta x_i}}{\sum_{C_i} e^{\sum_{D_u} \beta x_j}} = \frac{e^{\beta s_i}}{\sum_{C_i} e^{\beta s_j}}$$

- For our example {A:(t=4,x=2,δ=1), B:(t=8,x=5,δ=1), C:(t=8,x=7,δ=1), and D:(t=11,x=18,δ=0), at t=8, the component of the Partial Likelihood is:

$$L_{t=8} = \frac{e^{12\beta}}{e^{12\beta} + e^{23\beta} + e^{25\beta}}$$

EXAMPLE

- For the small example with 4 subjects:
  {A: (t = 4,x = 2, δ = 1), B: (t = 8,x = 5, δ = 1),
  C: (t = 8,x = 7, δ = 1), and D: (t=11,x=18, δ =0),
- The Partial Likelihood Function is:

$$L(\beta) = \{\frac{e^{2\beta}}{e^{2\beta} + e^{5\beta} + e^{7\beta} + e^{18\beta}}\}\{\frac{e^{12\beta}}{e^{12\beta} + e^{23\beta} + e^{25\beta}}\}$$

- We can : (i) maximize it to obtain MLE for β, Tsiatis (1981) gives a proof of the asymptotic normality of $\hat{\beta}$;and (ii) use the score test, for example, to test Null Hypothesis: H_0: β=0

ESTIMATION

- From the Partial Likelihood:

$$L = \prod_{i=1}^{m} \frac{e^{\beta s_i}}{\sum_{C_i} e^{\beta s_j}}, \text{or}, \ln L = \sum_{i=1}^{m} \{\beta s_i - \ln \sum_{C_i} e^{\beta s_j}\}$$

- Estimation, by iteration, is carried out using:

$$\frac{d}{d\beta} \ln L = \sum_{i=1}^{m} \{s_i - \frac{\sum s_j e^{\beta s_j}}{\sum e^{\beta s_j}}\}, \text{and}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\beta^2} \ln L = -\sum_{i=1}^{m} \frac{\sum e^{\beta s_j} \sum s_j^2 e^{\beta s_j} - [\sum s_j e^{\beta s_j}]^2}{\{\sum e^{\beta s_j}\}^2}$$

EXERCISES

T10.1 Under the logistic Regression model with a continuous covariate X, find the Odds Ratio associated with m units increase in X

T10.2 Use this simple data set (of Example #1), say, time is in months: {2,3+,5,5,6+,8,11,11+,12,15+}. Run SAS to form the Kaplan Meier curve and calculate the 95% confidence interval for "one-year survival rate".