CYP-C Data Analysis Using SAS II

CYP-C Research Champion Webinar

November 24, 2017

Jason D. Pole, PhD

PEDIATRIC ONCOLOGY GROUP OF ONTARIO

Overview

- Data Analysis
  - Introduction to time-to-event analysis
  - Kaplan-Meier Curves
  - Cumulative Incidence Curves
  - Introduction to Cox Proportional Hazards Modeling

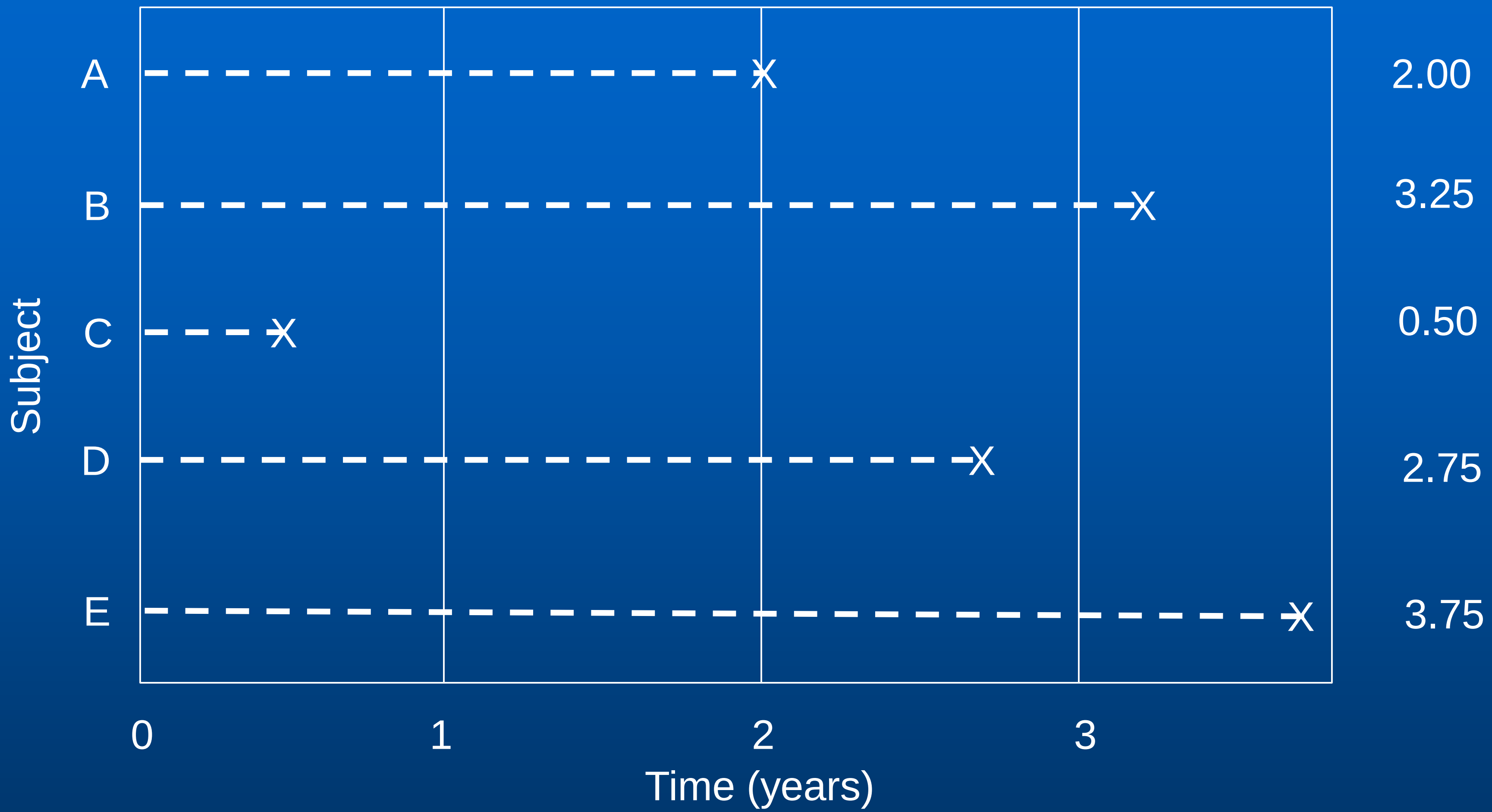
Introduction to Time-To-Event Analysis

Time-To-Event (TTE) Analysis

- Often referred to as survival analysis
- Modelling technique allows you to examine the occurrence and timing of any event
- Time has two components
  - Scale
    - Years, days, hours, minutes, seconds, microseconds
    - Selection of scale has little impact on analysis
      - Only effects the intercept
  - Origin
    - Often implicit but can have large effect on estimates
    - We use time of diagnosis as Time = 0 but really we really want time of disease onset
    - Time of diagnosis is affected by so many things
      - Age, sex, access to care, symptoms etc.
    - In RCTs it is time of randomization

Time-To-Event (TTE) Analysis II

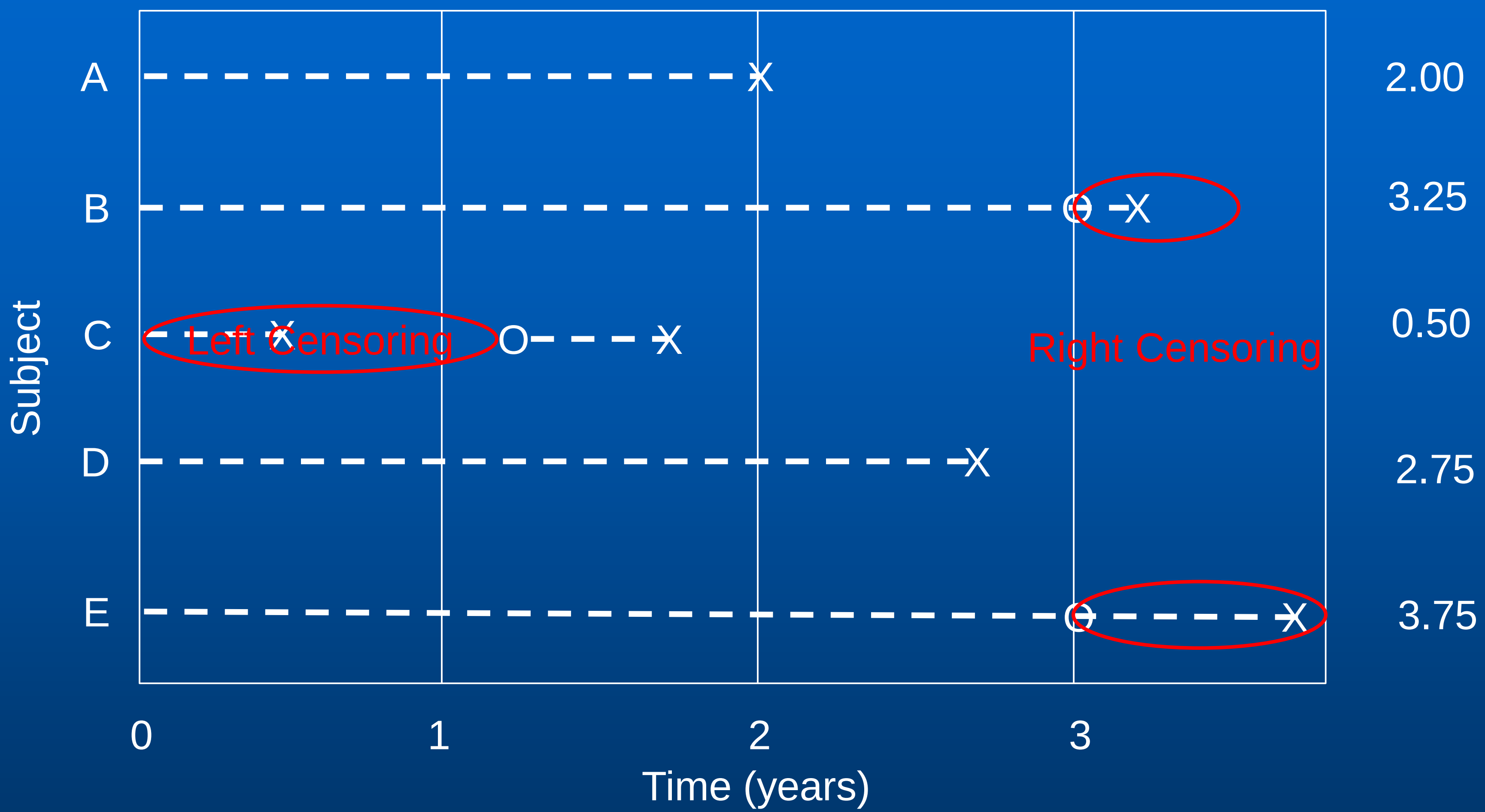
- Interested in the frequency of events happening over a period of observation
- By counting frequency over time we can think of this as the density of events



A
B
C
D
E
Subject
2.00
3.25
0.50
2.75
3.75
0
1
2
3
Time (years)

Censoring

- Describes periods of no observation
  - Many different kinds of censoring
    - Left – some period before you start observing where events could occur
    - Right – some period after you stopped observing where events could occur
    - Interval – combines both left and right censoring



A
B
C
D
E
Subject
2.00
3.25
0.50
2.75
3.75
Left Censoring
Right Censoring
0
1
2
3
Time (years)

Describing TTE Distributions

- **Cumulative Distribution Function**
  - **Tells us the probability that the variable *T* will be less than or equal to any value of time (*t*) we choose F(*t*)**
- **Survival Function**
  - **Probability of surviving beyond *t***
  - **S*(t)* = 1 - F(*t*)**
  - **S*(t)* is a probability: bounded by 0 and 1**

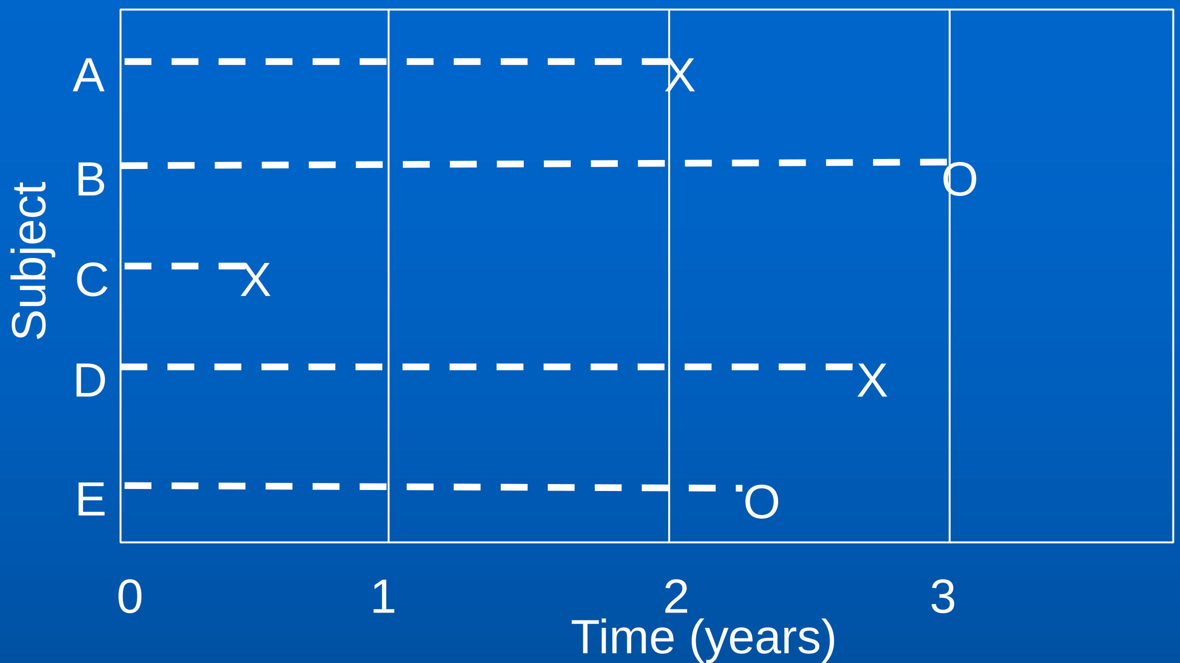
Describing TTE Distributions 2

- Hazard Function
  - Quantifies the instantaneous risk that an event will occur at time *t*
  - We condition this on having survived to time *t*
  - Describes the number of events per interval of time

- The survival function and hazard function are all equivalent ways describing a continuous probability function

Data Structure in TTE

- For basic TTE analysis (no left censoring)
- For each unit of analysis
  - time from start of observation (origin) to event or censoring (measure in any scale you choose)
  - Status at end of time (often called censor)
    - Status = 0 = person had event (observed event)
    - Status = 1 = person was censored (observation ended)



Subject	Time	Censor	Tx_Arm	Age
A	2.00	0	2	1
B	3.00	1	2	2
C	0.50	0	1	2
D	2.75	0	1	3
E	2.25	1	2	1

Kaplan-Meier Estimator

Kaplan-Meier Estimator

- Most widely used method to estimate the survival function
- Also known as the product-limit estimator
- In 1958, Kaplan and Meier demonstrated that this method was the nonparametric maximum likelihood estimator (although the method had been used for years earlier)

Overall Survival

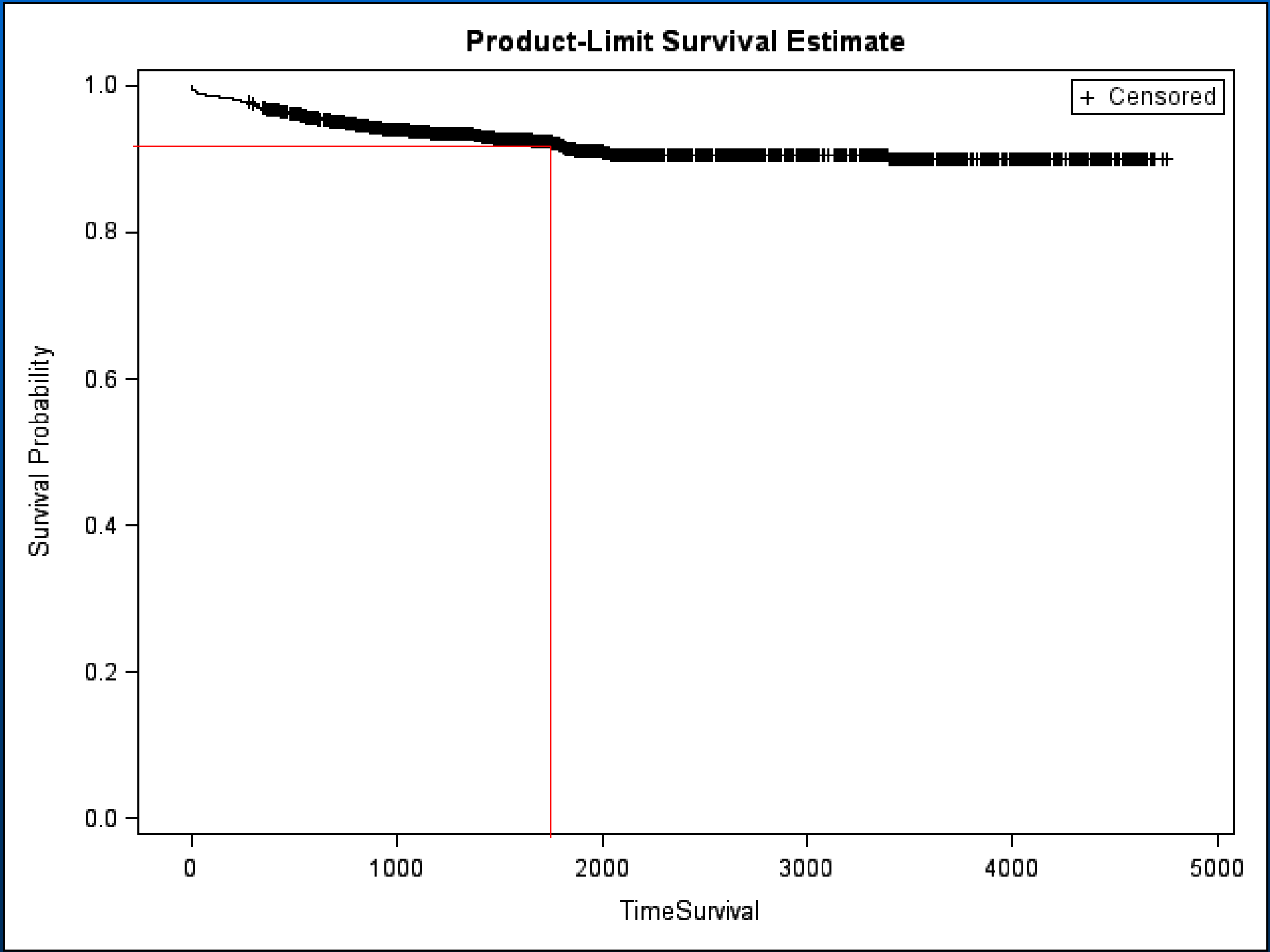
```
DATA T7; SET T6;

IF DUMALL = 1;

TimeLastFU = LAST_CONTACT_DATE - DX_DATE;
LABEL TIMELASTFU = 'NO. OF DAYS BETWEEN DIAGNOSIS AND LAST FU';

TimeDeath = DateDeath - DX_DATE;
LABEL TIMEDEATH = 'NO. OF DAYS BETWEEN DIAGNOSIS AND DEATH';

/* SETS ALL POST-MORTEM DEATHS TO DAY ZERO */
/* CensOS = 1 = PATIENT IS ALIVE */
If TimeDeath < 0 then TimeDeath = 0;
If DateDeath = . then TimeDeath = .;
If TimeDeath = . then CensOS = 1; else CensOS = 0;
TimeSurvival = Min (TimeLastFU, TimeDeath);

RUN;

PROC LIFETEST DATA = T7;
TIME TIMESURVIVAL*CENSOS(1);
RUN;
```

The LIFETEST Procedure

Product-Limit Survival Estimates

Survival Time	Survival	Failure	Survival Standard Error	Number Failed	Number Left
1825.00	0.9127	0.0873	0.00600	209	927

```
PROC LIFETEST DATA = T7 PLOT = (S);
TIME TIMESURVIVAL*CENSOS(1);
RUN;
```



```
DATA T7; SET T7;
IF 0 <= DX_AGE <= 0 THEN EARLY_AGE = 'INFANT';
IF 1 <= DX_AGE <=5 THEN EARLY_AGE = 'YOUNG';
IF 6 <
RUN;
```

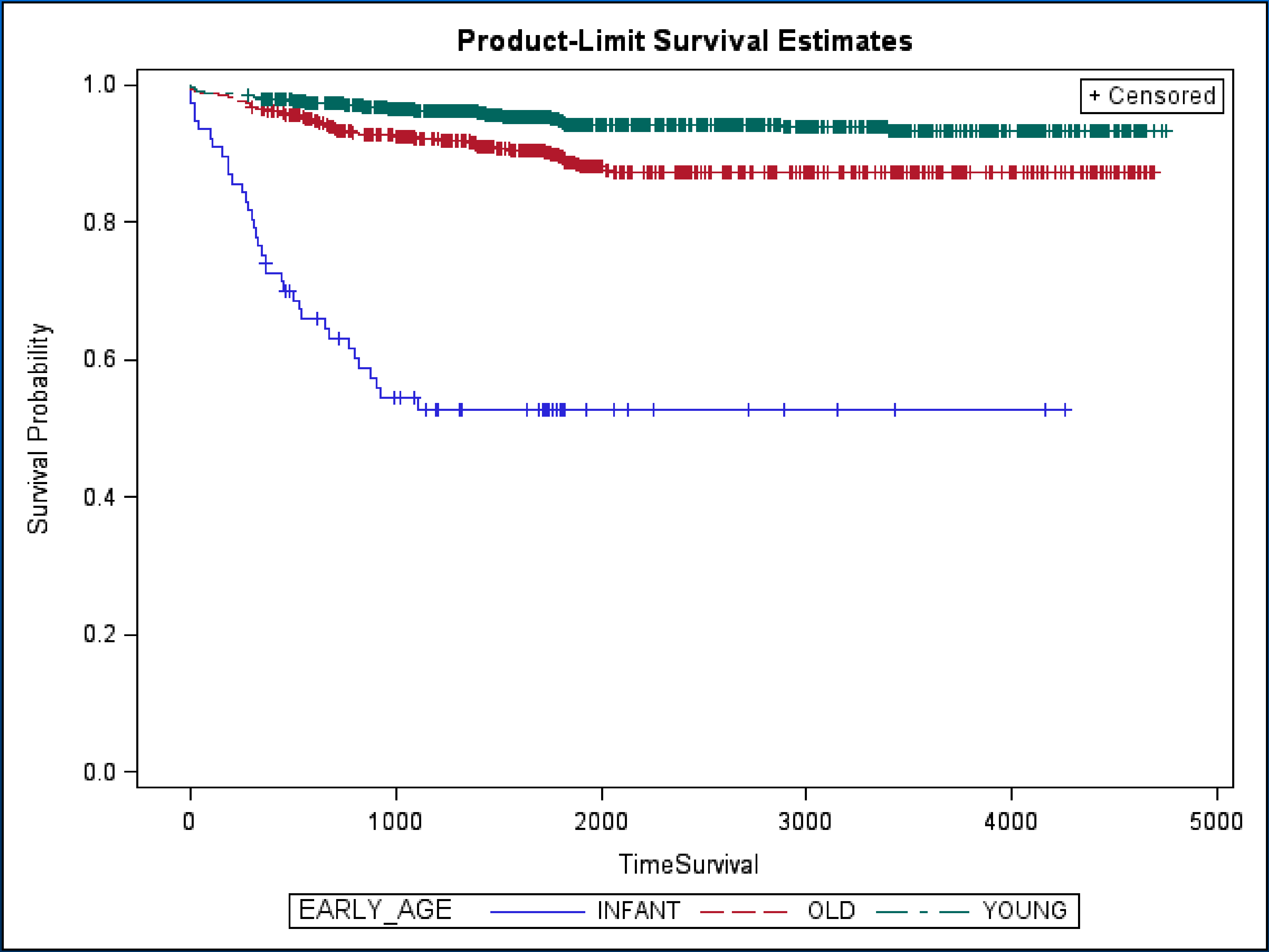
The LIFETEST Procedure

Summary of the Number of Censored and Uncensored Values

Stratum	EARLY_AGE	Total	Failed	Censored	Percent Censored
1	INFANT	77	35	42	54.55
2	OLD	939	96	843	89.78
3	YOUNG	1707	85	1622	95.02
Total		2723	216	2507	92.07

NOTE: 9 observations with invalid time, censoring, or strata values were deleted.

3.00	0.9977	0.00234	0.00117	4	1703
5.00	0.9971	0.00293	0.00131	5	1702



Product-Limit Survival Estimates
+ Censored
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
TimeSurvival
EARLY_AGE
INFANT
OLD
YOUNG

Test of Equality over Strata

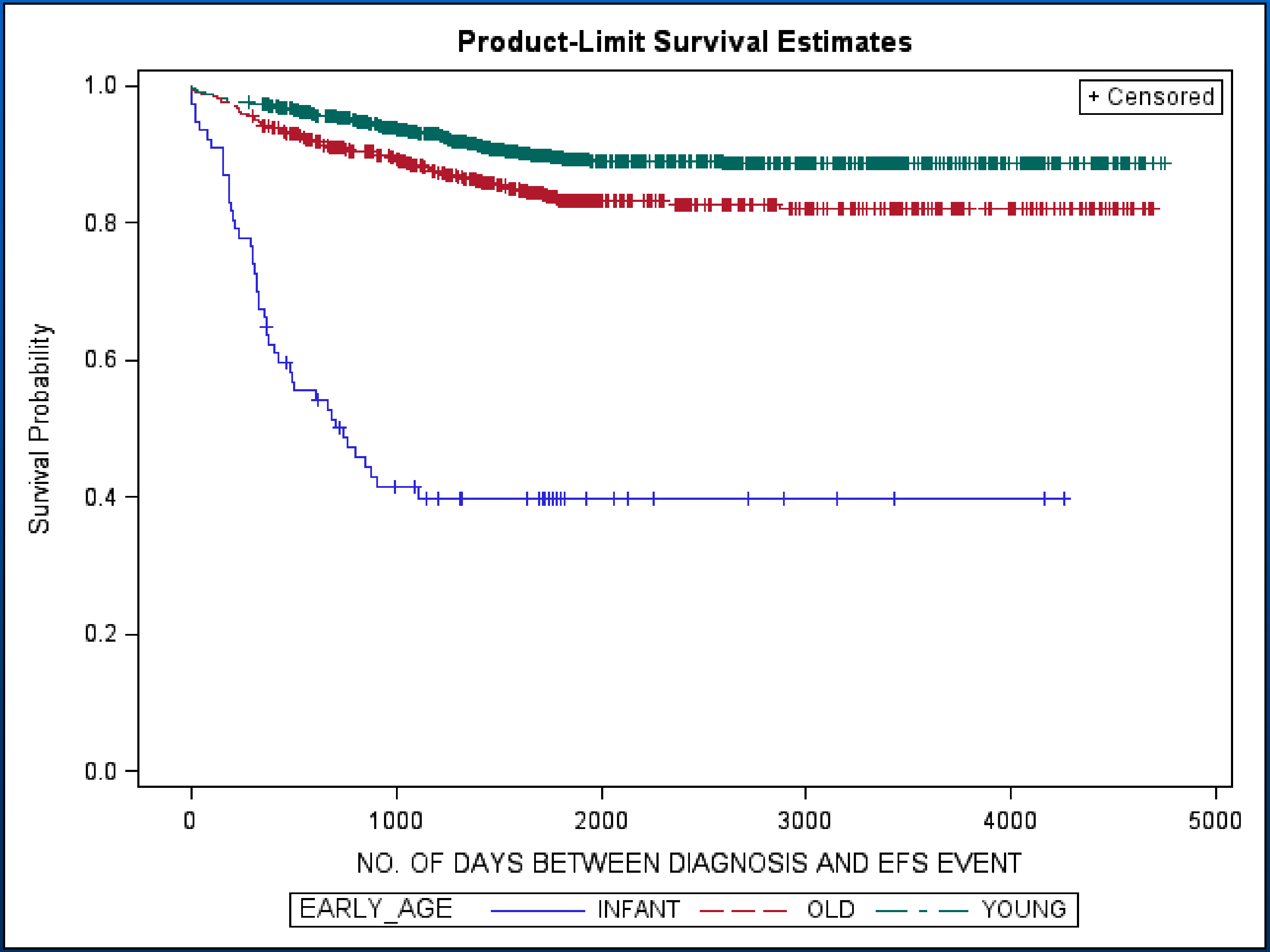
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	242.9273	2	<.0001
Wilcoxon	263.9643	2	<.0001
-2Log(LR)	124.7142	2	<.0001

- **Each test has different properties**
  - **Wilcoxon is more sensitive to early times (is a weighted sum of deviations and by definition there are more observations in the early period)**

Event-Free Survival

```
DATA T7; SET T6;

IF DUMALL = 1;

TimeRelapse = RX_DATE1 - DX_DATE;
LABEL TIMERELAPSE = 'NO. OF DAYS BETWEEN DIAGNOSIS AND FIRST RELAPSE';

TimeLastFU = LAST_CONTACT_DATE - DX_DATE;
LABEL TIMELASTFU = 'NO. OF DAYS BETWEEN DIAGNOSIS AND LAST FU';

TimeDeath = DateDeath - DX_DATE;
LABEL TIMEDEATH = 'NO. OF DAYS BETWEEN DIAGNOSIS AND DEATH';
/* SETS ALL POST-MORTEM DEATHS TO DAY ZERO */
IF TimeDeath < 0 then TimeDeath = 0; if DateDeath = . then TimeDeath = .;

/* DEFINES EFS USING RELASPE, DEATH AND LAST FOLLOW-UP */
TimeEvent = Min(TimeRelapse, TimeLastFU, TimeDeath);
LABEL TIMEEVENT = 'NO. OF DAYS BETWEEN DIAGNOSIS AND EFS EVENT';

/* IF PATIENT DID NOT RELASPE AND DID NOT DIE THEN CENSORED */
IF (TIMERELAPSE = . AND TIMEDEATH = .) THEN CensEFS = 1; else CensEFS = 0;

IF 0 <= DX_AGE <= 0 THEN EARLY_AGE = 'INFANT';
IF 1 <= DX_AGE <=5 THEN EARLY_AGE = 'YOUNG';
IF 6 <= DX_AGE THEN EARLY_AGE = 'OLD';

RUN;
```

```
PROC LIFETEST DATA = T7 PLOT = (S);
TIME TIMEEVENT*CENSEFS(1);
STRATA EARLY_AGE;
RUN;
```



Cumulative Incidence

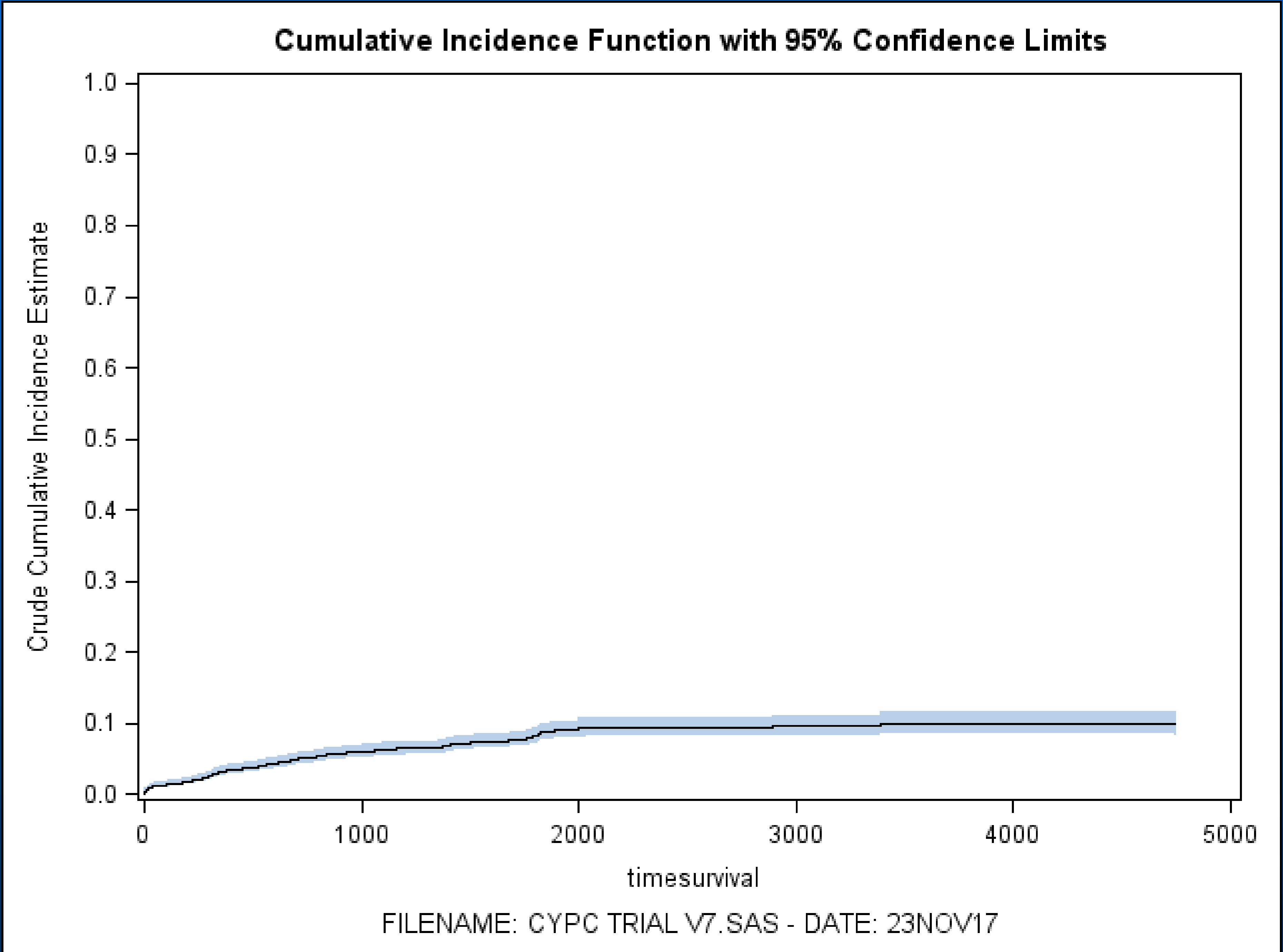
Cumulative Incidence

- probability that a particular event, such as occurrence of a particular disease, has occurred before a given time

- In situation with only right censoring equivalent to 1-survival

- In SAS 9.4 can be estimated using PHREG procedure, prior need to use macro

```
DATA T8; SET T7;
IF TIMESURVIVAL = 0 THEN TIMESURVIVAL = 0.005;
RUN;

%CumIncid(data=t8,
          out=CumInc,
          time=timesurvival,
          status=censos,
          event=0,
          compete=2,
          censored=1,
          strata=,
          alpha=.05,
          options=noprint plotcl);
RUN;
```


Cumulative Incidence Function with 95% Confidence Limits
Crude Cumulative Incidence Estimate
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1000
2000
3000
4000
5000
timesurvival
FILENAME: CYPC TRIAL V7.SAS - DATE: 23NOV17

Cox Proportional Hazards Regression

Cox Proportional Hazards

- K-M Curves are limited by not being able to control or adjust survival for other co-variates (only stratified analysis)

- Cox Regression is semi-parametric (you do not need to specify a probability distribution for survival times)

- Can easily incorporate time-dependent co-variates

- Can use discrete and continuous time measures (you may only measure an outcome every year)

Cox Proportional Hazards II

- Reminder
  - Hazard Function quantifies the instantaneous risk that an event will occur at time *t*

- Key Assumption is proportional hazards
  - survival curves for two strata (defined by any covariate you put in the model) must have hazard functions that are proportional over time (i.e. constant relative hazard)
  - Test this by introducing an interaction with time for each covariate and testing if the interaction term is statistically significant

```
PROC PHREG DATA = T7;
CLASS EARLY_AGE (REF="YOUNG");
MODEL TIMESURVIVAL*CENSOS(1) = EARLY_AGE / RL;
RUN;

/* note I have recoded early_age to be numeric */
```

The PHREG Procedure

Model Information

Data Set	WORK.T7
Dependent Variable	TimeSurvival
Censoring Variable	CensOS
Censoring Value(s)	1
Ties Handling	BRESLOW

Number of Observations Read	2732
Number of Observations Used	2723

Class Level Information

Class	Value	Design Variables	
EARLY_AGE	INFANT	1	0
	OLD	0	1
	YOUNG	0	0

The PHREG Procedure

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
EARLY_AGE	2	163.2025	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
EARLY_AGE INFANT	1	2.57624	0.20166	163.2015	<.0001	13.148	8.855	19.521	EARLY_AGE INFANT
EARLY_AGE OLD	1	0.74155	0.14893	24.7912	<.0001	2.099	1.568	2.811	EARLY_AGE OLD

```
PROC PHREG DATA = T7;
CLASS EARLY_AGE (REF="YOUNG") MALE (REF="0");
MODEL TIMESURVIVAL*CENSOS(1) = EARLY_AGE MALE / RL;
RUN;
```

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
EARLY_AGE	2	166.9010	<.0001
MALE	1	6.9780	0.0083

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
EARLY_AGE	INFANT	1	2.61208	0.20220	166.8854	<.0001	13.627	9.169	20.255	EARLY_AGE INFANT
EARLY_AGE	OLD	1	0.73262	0.14897	24.1858	<.0001	2.081	1.554	2.786	EARLY_AGE OLD
MALE	1	1	0.37553	0.14216	6.9780	0.0083	1.456	1.102	1.924	MALE 1

Testing Proportionality Assumption

```
PROC PHREG DATA = T7;
CLASS EARLY_AGE (REF="YOUNG") MALE (REF="0");
MODEL TIMESURVIVAL*CENSOS(1) = EARLY_AGE MALE AGE_T MALE_T/RL;
AGE_T = EARLY_AGE*LOG(TIMESURVIVAL);
MALE_T = MALE*LOG(TIMESURVIVAL);
PROPORTIONALITY_TEST: TEST AGE_T, MALE_T;
RUN;
```

The PHREG Procedure

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
EARLY_AGE	2	138.6045	<.0001
MALE	1	0.0982	0.7541
AGE_T	1	1.5175	0.2180
MALE_T	1	0.0887	0.7658

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
EARLY_AGE	INFANT	1	3.05714	0.40235	57.7340	<.0001	21.267	9.665	46.793	EARLY_AGE INFANT
EARLY_AGE	OLD	1	0.22946	0.42496	0.2915	0.5892	1.258	0.547	2.893	EARLY_AGE OLD
MALE	1	1	0.18871	0.60235	0.0982	0.7541	1.208	0.371	3.933	MALE 1
AGE_T		1	0.07987	0.06484	1.5175	0.2180	1.083	0.954	1.230	
MALE_T		1	0.02923	0.09814	0.0887	0.7658	1.030	0.849	1.248	

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
PROPORTIONALITY_TEST	1.6654	2	0.4349

Topics Covered

- Time-To-Event Data Analysis
  - Introduction to time-to-event analysis
  - Kaplan-Meier Curves
    - Testing difference over strata
  - Cumulative Incidence Curves
    - Use of macro
  - Introduction to Cox Proportional Hazards Modeling
    - Testing proportional hazards assumption