JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Public Health Statistics: Dealing with the Element of Time

John McGready, PhD
Johns Hopkins University

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

(Sample) Incidence Rates as Summary Measures for Time-to-Event Analysis and Spatial Data

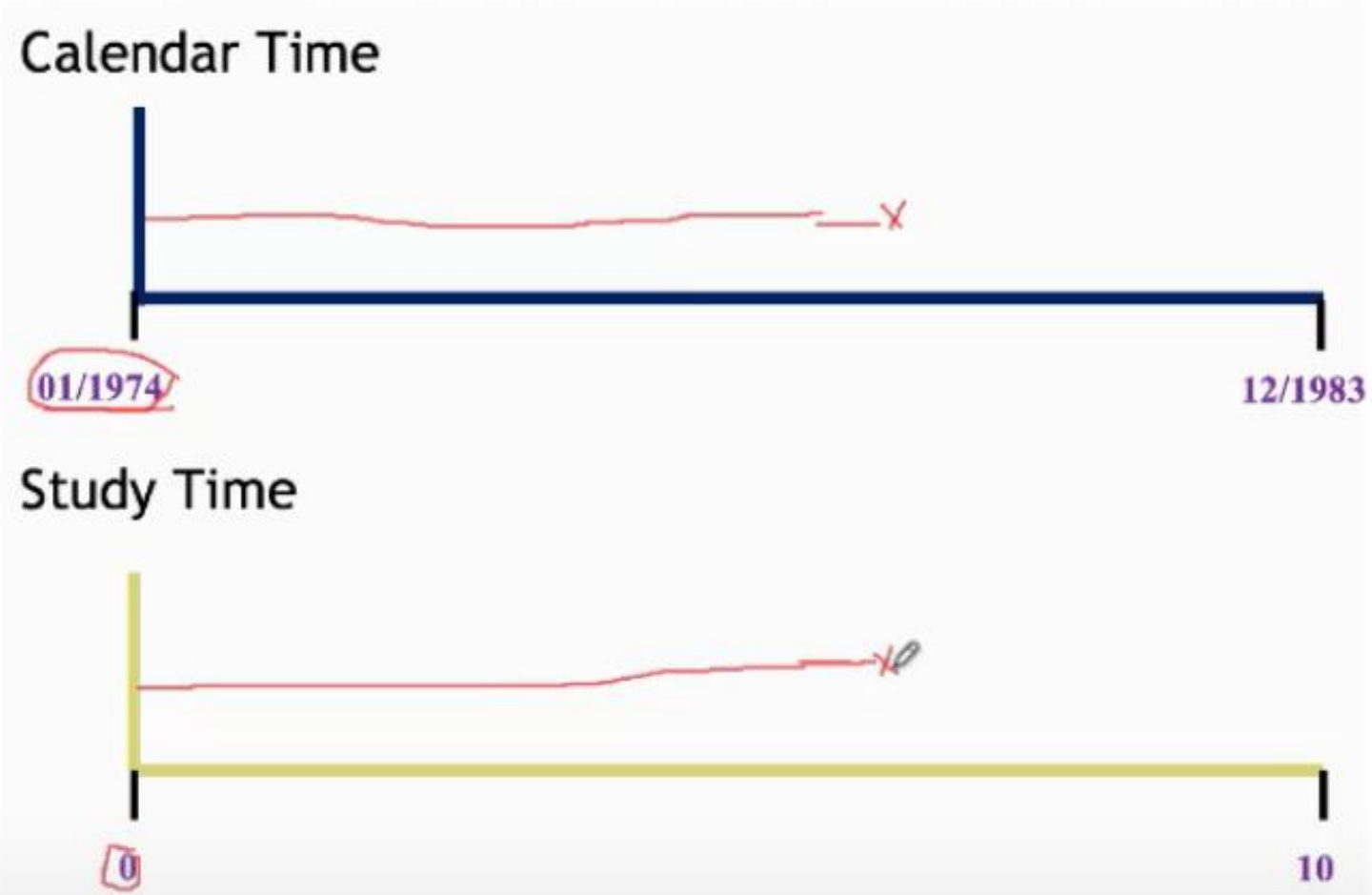
Learning Objectives

- For spatial data, including event counts and person totals in a defined observation period, you will be able to:
  - Summarize the person-count standardized event counts as event rates

- For time-to-event outcomes with individual outcome times known, you will be able to:
  - Distinguish between calendar time and study time scales for time-to-event data
  - Define censoring in the context of time-to-event studies
  - Explain why either ignoring the time component, or averaging subject follow-up times can be problematic for summarizing time-to-event data
  - Compute event incidence rates using event counts and cumulative follow-up times

Example: PBC Trial: Subject 1

- Subject 1: Patient who enters at start of study and dies 7 years later



Event Rates for Data Without Known Event Times

- For some data involving events occurring over time, the exact event times are not recorded but are grouped into time intervals
- This is the case for many death and disease rates grouped by area (country, state, city, region, zip code, etc.) by year (or some other unit of time)

Example: Lung Cancer Cases in Pennsylvania: 2002 Data

- Data from 2002 on incidence of lung cancer diagnoses in the state of Pennsylvania, and on the overall state population
- Data are stratified by county, sex, race, and age groups but can be aggregated
- How to summarize?
  - An incidence rate can be computed by taking the total number of cases divided by the total person-time at risk

Example: Lung Cancer Cases in Pennsylvania: Rescaling Incidence Rate

- The incidence rate can be rescaled for differing time periods: usually done so numerator has an integer component

$$\widehat{IR} = \frac{10{,}279 \text{ cases}}{12{,}281{,}054 \text{ person years}} \approx \frac{0.0008 \text{ cases}}{\text{person year}}$$

- Expressed as events per 10,000 years:

$$\widehat{IR} = \frac{0.0008 \text{ cases}}{\text{person year}} \times 10{,}000 \text{ years} = \frac{8 \text{ cases}}{10{,}000 \text{ person years}}$$

Example: Lung Cancer Cases in Pennsylvania

- Isn't this just a proportion (percent) measure for a binary outcome (lung cancer yes/no?)
  - There is an element of time
  - Nevertheless, you could think of this in percent terms: 0.08% incidence of new cases in the year (i.e., 0.08% of the sample under study developed lung cancer in 2002)
  - However, even if interpreted as a "percent over some unit of time," rates tend to be very small as proportions, and as such, their statistical properties will differ from proportions as we have defined them previously

Time-to-Event Data With Known Event Times

- For some time-to-event data, the individual event times are known, and this information can be incorporated into incidence rate computations
- This is the case for many longitudinal cohort studies where subjects are followedfrom a defined starting point, up to a certain amount of time

Example: Primary Biliary Cirrhosis (PBC) Trial

- Randomized trial conducted at Mayo Clinic

Description of the Study Population

The study began on January 1, 1974, and patient accrual was terminated in December 1983. During that 10-year period, 422 patients with primary biliary cirrhosis satisfied the criteria, and 309 of them entered the trial and were randomized. Of those entering the

Example: PBC Trial

- Mayo Clinic: Primary Biliary Cirrhosis (PBC treatment), randomized clinical trial
- The primary outcome of interest was death in the follow-up period
  - Ultimately, the researchers were interested in evaluating the effect of the drug D-pencillamine (DPCA) on survival
- What can happen when following subjects over time?

Too nerdy?
What can you expect from a data scientist!

The graph on the left credit to Matt Henderson
The graph on the right credit to Pringles

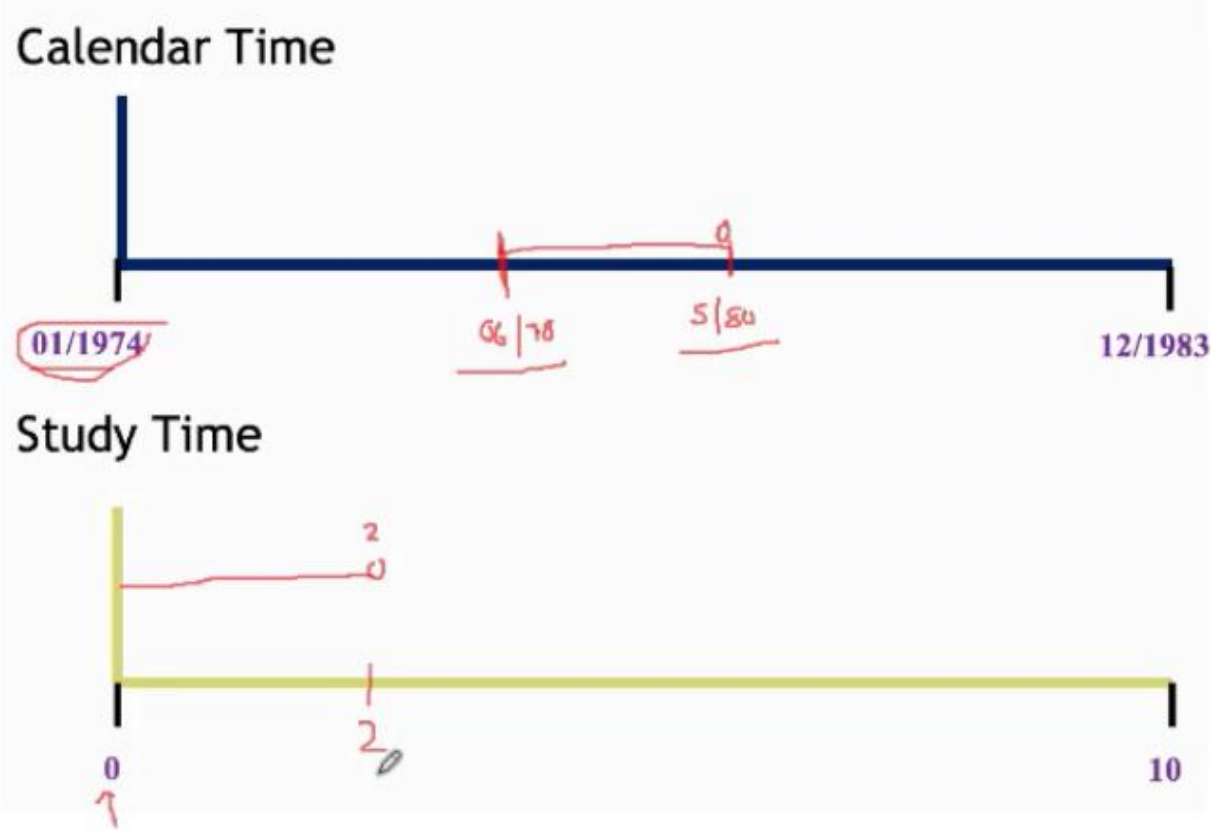
For ML/ AI/ Data Science learning materials, please check my previous posts.

I share my learning journey into Data Science with my amazing LinkedIn friends, please follow me and let's grow together! **Alex (Mengyao) Wang**

#application #technology #mathmatics #innovation

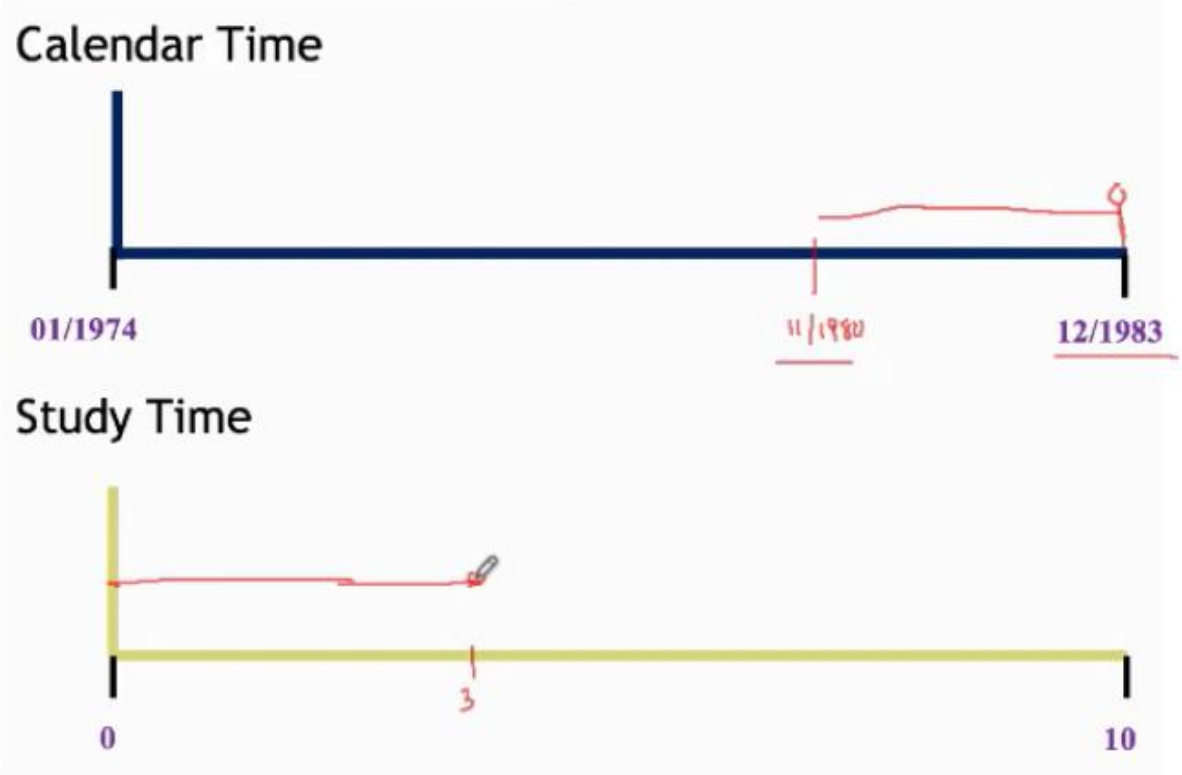
Example: PBC Trial: Subject 2

- Subject 2: Patient who enters in June 1978 and is lost to follow-up in May 1980



Example: PBC Trial: Subject 3

- Subject 3: Patient who enters in November 1980 and stays in study until completion in December 1983 (patient alive at end of study)



Example: PBC Trial: Study Time Graphic

- Putting the 3 subjects together on the Study Time graphic

Complete Versus Censored Observations

- Subject 1 is a complete observation
  - We know that he/she had the outcome under study (death) after seven years in the study

- Subjects 2 and 3 are censored observations
  - We have partial information about the outcome under study (death)
  - While Subject 2 was still alive when he/she was lost to follow-up, we know that he/she survived two years on the study clock
  - Similarly, Subject 3 survived three years on the study clock

Single Numeric Summary, Time-to-Event Data with Known Event Times—1

- How can this be summarized numerically?
- Option A: Treat death as binary and report the proportion who died in the follow-up period. With these 3 subjects, this is:

$$\hat{p} = \frac{1}{3} \approx 0.33\ (33\%)$$

Example: ART and Partner-to-Partner HIV Transmission: Background and Methods

BACKGROUND

Antiretroviral therapy that reduces viral replication could limit the transmission of human immunodeficiency virus type 1 (HIV-1) in serodiscordant couples.

METHODS

In nine countries, we enrolled 1763 couples in which one partner was HIV-1–positive and the other was HIV-1–negative; 54% of the subjects were from Africa, and 50% of infected partners were men. HIV-1–infected subjects with CD4 counts between 350 and 550 cells per cubic millimeter were randomly assigned in a 1:1 ratio to receive antiretroviral therapy either immediately (early therapy) or after a decline in the CD4 count or the onset of HIV-1–related symptoms (delayed therapy). The primary prevention end point was linked HIV-1 transmission in HIV-1–negative partners. The primary clinical end point was the earliest occurrence of pulmonary tuberculosis, severe bacterial infection, a World Health Organization stage 4 event, or death.

ART and Partner-to-Partner HIV Transmission: Results

RESULTS

As of February 21, 2011, a total of 39 HIV-1 transmissions were observed (incidence rate, 1.2 per 100 person-years; 95% confidence interval [CI], 0.9 to 1.7); of these, 28 were virologically linked to the infected partner (incidence rate, 0.9 per 100 person-years, 95% CI, 0.6 to 1.3). Of the 28 linked transmissions, only 1 occurred in the early-therapy group (hazard ratio, 0.04; 95% CI, 0.01 to 0.27; $P<0.001$). Subjects receiving early therapy had fewer treatment end points (hazard ratio, 0.59; 95% CI, 0.40 to 0.88; $P=0.01$).

Example: Infant Mortality: Abstract

ABSTRACT

Background: The effect of vitamin A supplementation on the survival of infants aged <6 mo is unclear. Because most infant deaths occur in the first few month of life, maternal supplementation may improve infant survival.

Objectives: The objective was to assess the effect of maternal vitamin A or β-carotene supplementation on fetal loss and survival of infants <6 mo of age.

Design: Married women of reproductive age in 270 wards of Sarlahi district, Nepal, were eligible to participate. Wards were randomly assigned to have women receive weekly doses of 7000 μg retinol equivalents as retinyl palmitate (vitamin A), 42 mg *all-trans*-β-carotene, or placebo. Pregnancies were followed until miscarriage, stillbirth, maternal death, or live birth of one or more infants, who were followed through 24 wk of age.

Example: Infant Mortality: Results

- From "Results"
  - *"A total of 43,559 women were enrolled; 15,892 contributed 17,373 pregnancies and 15,997 live born infants to the trial"*
- The investigators kindly shared a 2/3 (10,295) random sample of the live births data
- Total follow-up time: 1,627,725 days
- Total deaths in (6-month) follow-up period: 644

Example: Infant Mortality

- Infant mortality rate in 6 months post birth

$$\widehat{IR} = \frac{644 \text{ deaths}}{1{,}627{,}725 \text{ days}} = 0.0004 \text{ deaths/day}$$

- IR estimate per (1 person) year

$$\widehat{IR} = 0.0004 \text{ deaths/day} \times 365 \text{ days/year} \approx 0.146 \text{ deaths/year}$$

- IR estimate per 500 (person) years

$$\widehat{IR} = 0.146 \text{ deaths/year} \times 500 = 73 \text{ deaths} / 500 \text{ years}$$

Note on Terminology

- Analysis techniques for prospective cohort data, where time to an event is of interest, has several synonymous titles:
  - Incidence analysis
  - Survival analysis
  - Time-to-event analysis
  - Failure time data analysis

- "Survival analysis" is the most commonly used term when the individual event times are known
  - The event of interest does not necessarily have to be death

Summary

- Event data collected on a sample of (initially event-free) subjects followed over time is two dimensional
  - For each subject, there is a time measure and also a binary indicator
    - In some cases, the time measure is the same for every individual in the sample (spatial rate data from vital statistics sources)
    - In some cases, the individual event times are known and recorded (randomized clinical trials, prospective observational studies)

- The incidence rate summarizes these two dimensions (of time and whether the event occurred) into a single number

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Comparing Time-to-Event Data Between Two (or More) Samples, Numerically

Learning Objectives

- Upon completion of this lecture section, you will be able to:
  - Estimate a numerical comparison of time-to-event outcomes between two populations, using sample rate estimates
  - Interpret the resulting estimate—the incidence rate ratio—in words and in a public heath/scientific context

Example: Pennsylvania Lung Cancer Data: Incidence Rates

- Primary research question: How do the rates of lung cancer (new diagnoses) compare for men and women in Pennsylvania, 2002?
- Incidence rates for females and males

- Females: 6,351,391 person years (PY), 4,587 lung cancer diagnoses

$$\widehat{IR}_F = \frac{4{,}587 \text{ cases}}{6{,}351{,}391 \text{ PYs}}$$

$$\approx 0.00072 \text{ cases/PY}$$

- Males: 5,929,663 person years (PY), 5,692 lung cancer diagnoses

$$\widehat{IR}_M = \frac{5{,}692 \text{ cases}}{5{,}929{,}663 \text{ PYs}}$$

$$\approx 0.00096 \text{ cases/PY}$$

Example: Pennsylvania Lung Cancer Data: Incidence Rate Ratio

- Incidence rate ratio

$$\widehat{IRR} = \frac{\widehat{IR}_F}{\widehat{IR}_M} = \frac{0.00072 \text{ cases/PY}}{0.00096 \text{ cases/PY}} \approx 0.75$$

- Interpretations:
  - The risk of getting lung cancer (in 2002) for females was 0.75 times the risk for males
  - Females had 25% lower risk of lung cancer in 2002 when compared to males

Example: Clinical Trial, Primary Biliary Cirrhosis (PBC): Incidence Rates

- Primary research question: How does mortality (and hence) survival for PBC patients randomized to receive DPCA (D-penicillamine) compare to survival for PBC patients randomized to received a placebo?
- Incidence rates for DPCA and placebo groups

- DPCA: 872.5 person years (PY) total follow-up, 65 deaths

$$\widehat{IR}_{DPCA} = \frac{65 \text{ deaths}}{872.5 \text{ PYs}} \approx 0.075 \text{ deaths/PY}$$

- Placebo: 842.5 person years (PY) total follow-up, 60 deaths

$$\widehat{IR}_{PLACEBO} = \frac{60 \text{ deaths}}{842.5 \text{ PYs}} \approx 0.071 \text{ deaths/PY}$$

Example: Clinical Trial, PBC: Incidence Rate Ratio

- Incidence rate ratio

$$\widehat{IRR} = \frac{\widehat{IR}_{DPCA}}{\widehat{IR}_{PLACEBO}} = \frac{0.075 \text{ deaths/PY}}{0.071 \text{ deaths/PY}} \approx 1.06$$

- Interpretations
  - The risk of death in the DPCA group (in the study follow-up period) is 1.06 times the risk in the placebo group
  - Subjects in the DPCA group had 6% higher risk of death in the follow-up period when compared to the subjects in the placebo group

ART and Partner-to-Partner HIV Transmission: Results—1

RESULTS

As of February 21, 2011, a total of 39 HIV-1 transmissions were observed (incidence rate, 1.2 per 100 person-years; 95% confidence interval [CI], 0.9 to 1.7); of these, 28 were virologically linked to the infected partner (incidence rate, 0.9 per 100 person-years, 95% CI, 0.6 to 1.3). Of the 28 linked transmissions, only 1 occurred in the early-therapy group (hazard ratio, 0.04; 95% CI, 0.01 to 0.27; $P<0.001$). Subjects receiving early therapy had fewer treatment end points (hazard ratio, 0.59; 95% CI, 0.40 to 0.88; $P=0.01$).

ART and Partner-to-Partner HIV Transmission: Results—2

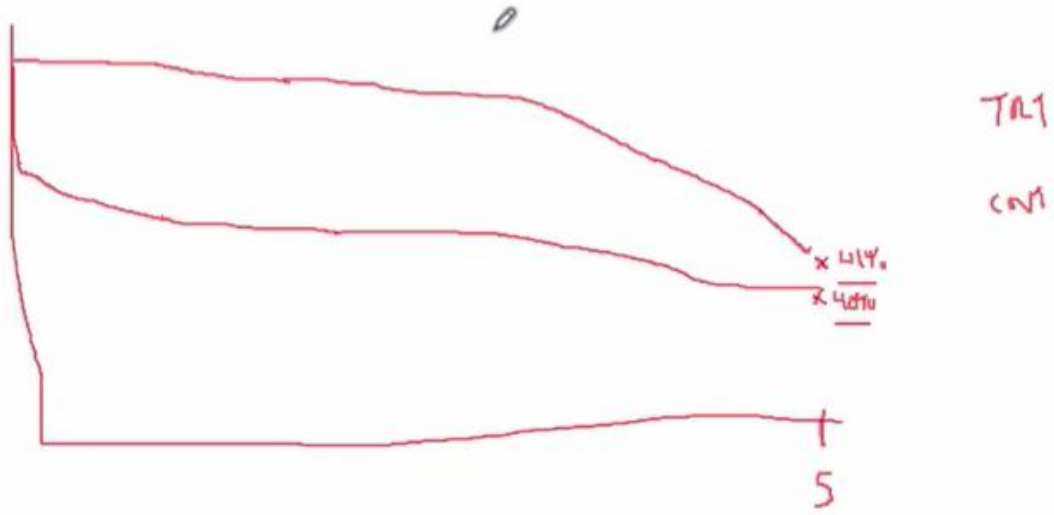
- From "Results"
  - "Of the 28 linked transmissions, only 1 occurred in the early therapy group (hazard ratio 0.04…)"

- *Note: hazard ratio and incidence rate ratio are (nearly) synonymous*

$$\frac{\widehat{IR}_{Early}}{\widehat{IR}_{Standard}} = \frac{\text{1 linked transmission/total PYs, early therapy}}{\text{27 linked transmissions/total PYs, standard}} = 0.04$$

Importance of Incorporating Time into Comparisons, When Follow-Up Times (Person Time) are Available

- What could potentially happen if follow-up time was ignored with the group comparison, and instead a ratio of proportions was computed?



ART and Partner-to-Partner HIV Transmission: Incidence Rate Ratio Interpretations

- Incidence rate ratio interpretation(s)
  - HIV discordant (at baseline) couples in which the HIV+ partner was given early ART therapy had 0.04 times the risk of within-couple transmission, as compared to couples in which the HIV+ partner was given standard therapy
  - HIV discordant (at baseline) couples in which the HIV+ partner was given early ART therapy had 96% lower risk of within-couple transmission, as compared to couples in which the HIV+ partner was given standard therapy

Summary

- The incidence rate ratio (IRR, estimated by $\widehat{IRR}$) is a numerical comparison of event rates between two groups
  - If more than two groups, one can be designated as the reference group, and incidence rate ratios can be computed for each of the other group incidence rates, each compared to this same reference group incidence rate
- The incidence rate ratio can be thought of as a relative risk measure that incorporates differences in subject follow-up times into the comparison when the follow-up times are known

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Summarizing Time-to-Event Data, Graphically

Learning Objectives

- Upon completion of this lecture section, you will be able to:
  - Explain the purpose of a survival curve and its basic properties
  - Interpret the Kaplan-Meier curve estimates of survival curves, with respect to summarizing time-to-event data for samples of data
  - Estimate a Kaplan-Meier curve "by hand" for small samples of data
  - Explain how censored observations are used in the Kaplan-Meier estimation process
  - Give approximate estimates of event time percentiles using a Kaplan-Meier curve
  - Interpret the complimentary presentation of the Kaplan-Meier survival curve

Rationale for Visual Display When Follow-Up Times Are Recorded

- Incidence rates are appropriate numerical summary measures for time-to-event data, in that these incorporate the two dimensions of the data—the time factor and the occurrence (count) of events—into a single statistic

- However, time-to-event is two dimensional
  - To capture the richness in such data visually, a graphic would have to display both dimensions

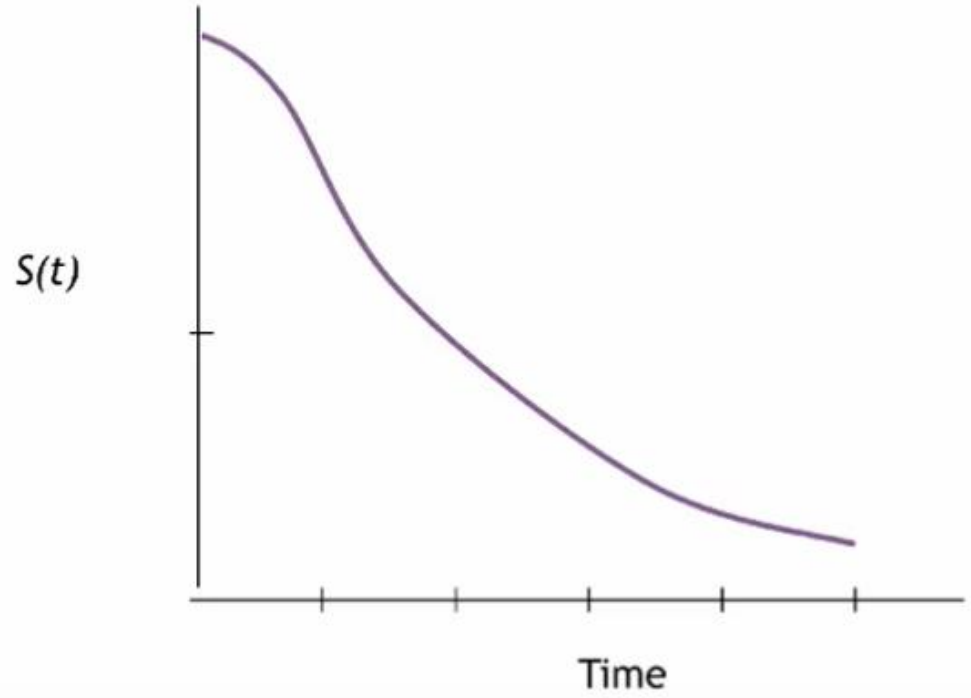
- A common visual display for time-to-event data is a survival curve
  - This can be estimated from sample data using the Kaplan-Meier approach

The Survival Curve, *S(t)*—2

- The survival curve can be estimated (for a population) based on sample data
- The estimated curve, $\hat{S}(t)$, is based on data from all subjects in a sample, both those who have the outcomes of interest and those who are censored
- The estimation procedure will be demonstrated shortly; first, some examples of the estimated curve, estimated by the Kaplan-Meier method (and, hence, called "Kaplan-Meier curves" or "KM Curves") will be shown

The Survival Curve, *S(t)*—1

- *S(t)* = Proportion remaining event free (surviving) at least to time *t* or beyond



Example: Primary Biliary Cirrhosis (PBC) Trial Data

- Mayo Clinic: Primary biliary cirrhosis (PBC treatment), randomized clinical trial
  - Outcome of research interest was death

- Overall incidence rate in follow-up period:

$$\widehat{IR} = \frac{125 \text{ deaths}}{1{,}715 \text{ Person Years (Years)}} \approx .073 \text{ deaths/year}$$

Example: Primary Biliary Cirrhosis (PBC) Trial Data

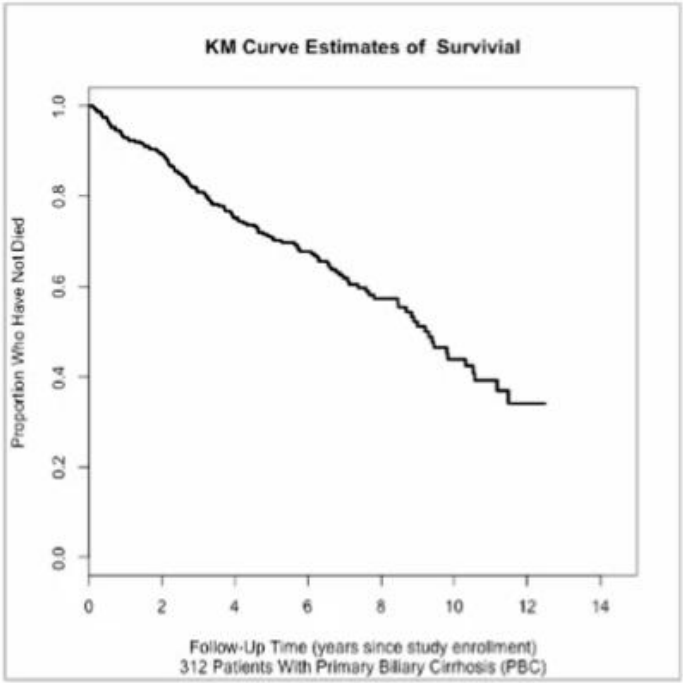
- Mayo Clinic: Primary biliary cirrhosis (PBC treatment), randomized clinical trial
  - Outcome of research interest was death

- Overall incidence rate in follow-up period:

$$\widehat{IR} = \frac{125 \text{ deaths}}{1{,}715 \text{ Person Years (Years)}} \approx .073 \text{ deaths/year}$$

Example: PBC Trial Data: Kaplan-Meier Curve

- The Kaplan-Meier curve shows the estimated proportion of the original sample of 312 patients who survived (did not have the event of death) by the corresponding follow-up time



Example: Infant Mortality, Nepal

- Infant mortality rate in 6 months post birth

$$\widehat{IR} = \frac{144 \text{ deaths}}{1{,}627{,}725 \text{ days}} = 0.0004 \text{ deaths/day}$$

- IR estimate per (1 person) year

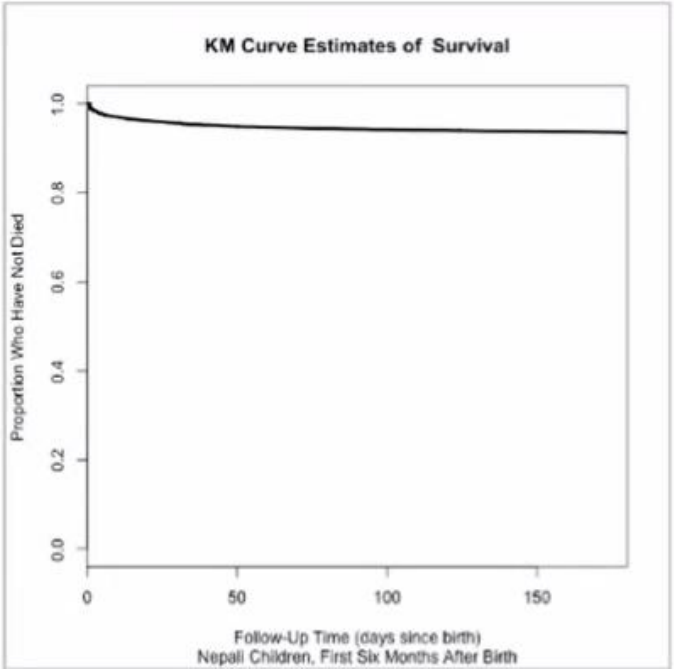
$$\widehat{IR} = 0.0004 \text{ deaths/day} \times 365 \text{ days/year} \approx 0.146 \text{ deaths/year}$$

- IR estimate per 500 (person) years

$$\widehat{IR} = 0.146 \text{ deaths/year} \times 500 = 73 \text{ deaths} / 500 \text{ years}$$

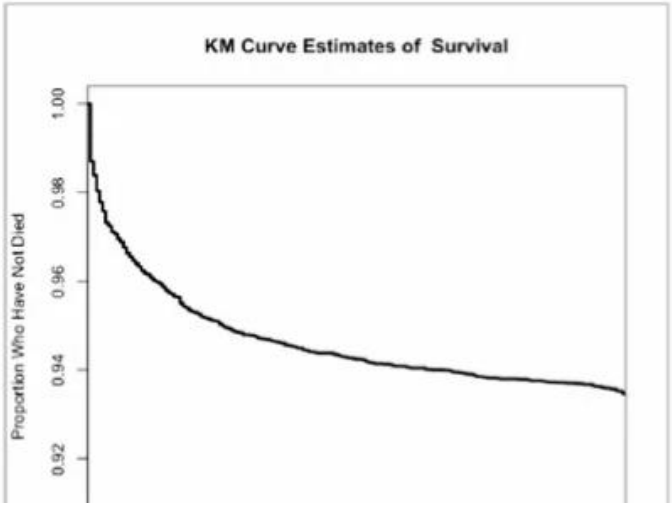
Example: Infant Mortality, Nepal: Kaplan-Meier (KM) Curve

- Infant mortality rate in 6 months post birth: The KM curve tracks the estimated proportion of the original sample of 10,000+ children who survived (did not have the event, death) by the corresponding follow-up time



Example: Infant Mortality, Nepal: KM Curve Zoomed In

- Infant mortality rate in 6 months post birth: The KM curve tracks the estimated proportion of the original sample of 10,000+ children who survived (did not have the event, death) by the corresponding follow-up time



Estimating the Kaplan Meier Curve

- Generally done using a computer
- I will demonstrate the estimation process with a small sample example
- The method uses the complete data in those who actually have the event in the follow-up period
  - The time at which the event occurred
  - The occurrence of the event
- Incomplete data for censored observations gives information about who is “at risk” to have the event at a given time in the follow-up period

Example: Smoking Cessation Data—1

- Twelve subjects who attend a smoking cessation workshop are followed for up to one month after they complete the workshop. They are followed until they quit smoking (the event of interest) or are lost to follow-up (censored)

- The data are as follows (times are in days, censoring is indicated by a "+"):

2 3+ 6 8 9+ 10 15+ 16 18 24+ 27 30+

KM Estimation

- At each event time *t*, the value *S(t)* will be estimated via the following formula:

$$\hat{S}(t) = \frac{N(t) - E(t)}{N(t)} \times \hat{S}\,(previous\ event\ time)$$

  - $N(t)$ is the number still at risk of having the event at time t
  - $E(t)$ is the number who have the event at time t
  - $\hat{S}(t)$ is the estimate proportion of the entire cohort that survives beyond time t (does not have the event by time t)

Example, KM Estimation: Smoking Cessation Data

- The data are as follows (times are in days, censoring is indicated by a "+"):

  2 3+ 6 8 9+ 10 15+ 16 18 24+ 27 30+

- $\hat{S}(0) = 1$, i.e., 100% of the sample have not quit smoking at time 0. The curve will remain at 1 until the first event occurs at 2 days. So $\hat{S}(1) = 1$, $\hat{S}(1.5) = 1$, etc.

Example, KM Estimation: Smoking Cessation Data: Event Time—1

- The data are as follows (times are in days, censoring is indicated by a "+"):

2 3+ 6 8 9+ 10 15+ 16 18 24+ 27 30+

- At t=2, the first event time, all 12 persons in the sample are at risk of quitting smoking, and only 1 person does quit. So $N(2) = 12$, and $E(2) = 1$. There are no previous event times.

$$\hat{S}(t) = \frac{N(t)-E(t)}{N(t)} \times \hat{S}(previous\ event\ time)$$

$$\hat{S}(2) = \frac{N(2)-E(2)}{N(2)} = \frac{12-1}{`12} = \frac{11}{12} \approx 0.92$$

Example, KM Estimation: Smoking Cessation Data: Event Time—2

- The data are as follows (times are in days, censoring is indicated by a "+"):

2 3+ 6 8 9+ 10 15+ 16 18 24+ 27 30+

- At t=6, the next event time, only 10 persons in the sample are at risk of quitting smoking (one quit at t=2 and one was censored at t=3). At t=6, only 1 person does quit. So $N(6)$=10, and $E(6)$=1. The previous event time is t=2.

$$\hat{S}(t) = \frac{N(t)-E(t)}{N(t)} \times \hat{S}(previous\ event\ time)$$

$$\hat{S}(6) = \frac{N(6)-E(6)}{N(6)} \times \hat{S}(6) = \frac{10-1}{10} \times 0.92 = 0.9 \times 0.92 \approx 0.83$$

Example, KM Estimation: Smoking Cessation Data: Event Time—3

- The data are as follows (times are in days, censoring is indicated by a "+"):

  2 3+ 6 8 9+ 10 15+ 16 18 24+ 27 30+

- At t=8, the next event time, only 9 persons in the sample are at risk of quitting smoking (one quit at t=2 ,one was censored at t=3, and one quit at t=6). At t=8, only 1 person does quit. So $N(8)$=9, and $E(8)$=1. The previous event time is t=6.

$$\hat{S}(t) = \frac{N(t)-E(t)}{N(t)} \times \hat{S}(previous\ event\ time)$$

$$\hat{S}(6) = \frac{N(8)-E(8)}{N(8)} \times \hat{S}(6) = \frac{9-1}{9} \times 0.83 = 0.89 \times 0.83 \approx 0.74$$

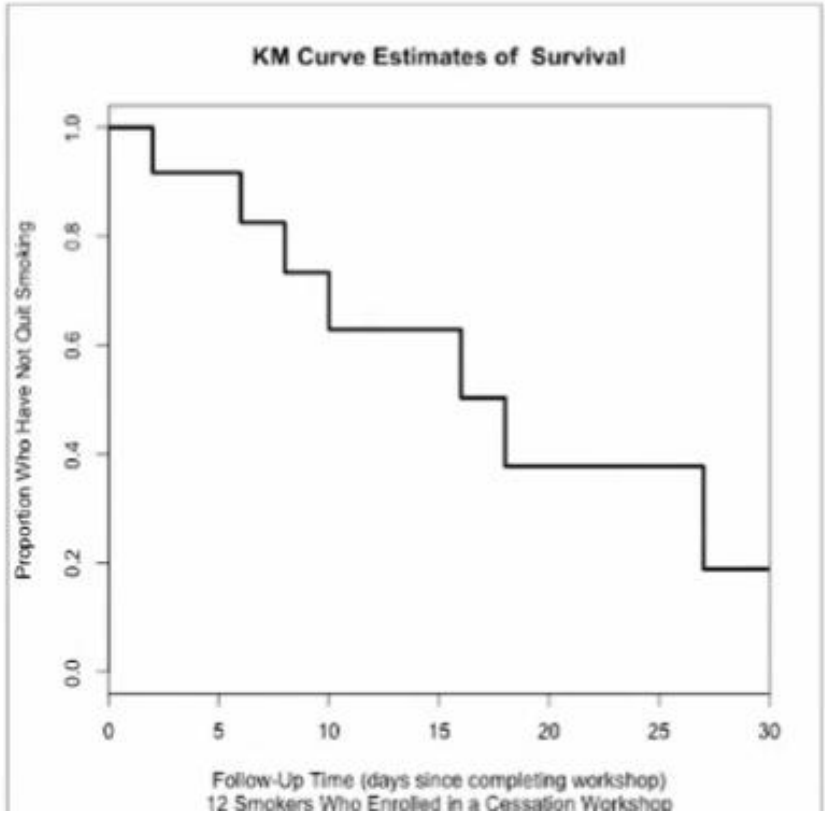
Example, KM Estimation: Smoking Cessation Data: Table

- The complete KM curve estimates are:

t	$\widehat{S}(t)$
2	0.92
6	0.83
8	0.74
10	0.63
16	0.50
18	0.37
27	0.19

Example, KM Estimation: Smoking Cessation Data: Graph

- The graphed KM curve!

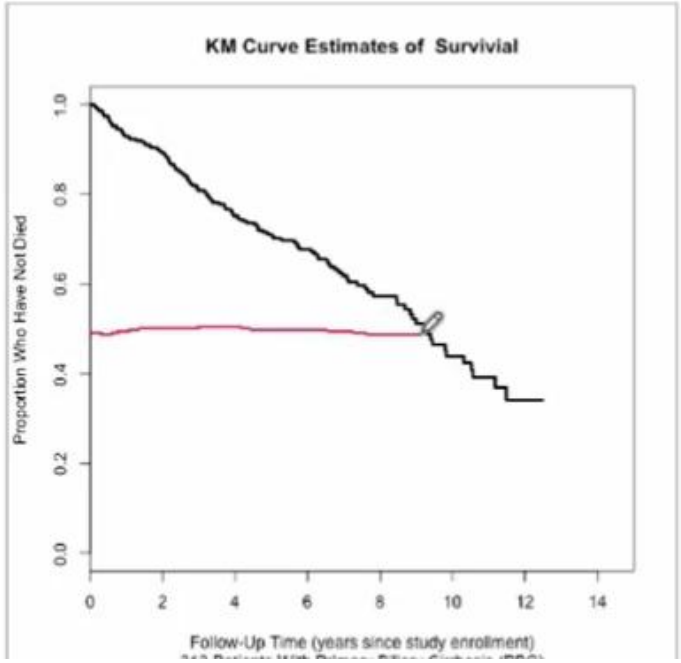


Kaplan-Meier Curve

- Graph is a step function
- "Jumps" at each observed event time
- Nothing is assumed about curved shape between each observed event time
- Can estimate percentiles of event times via this curve

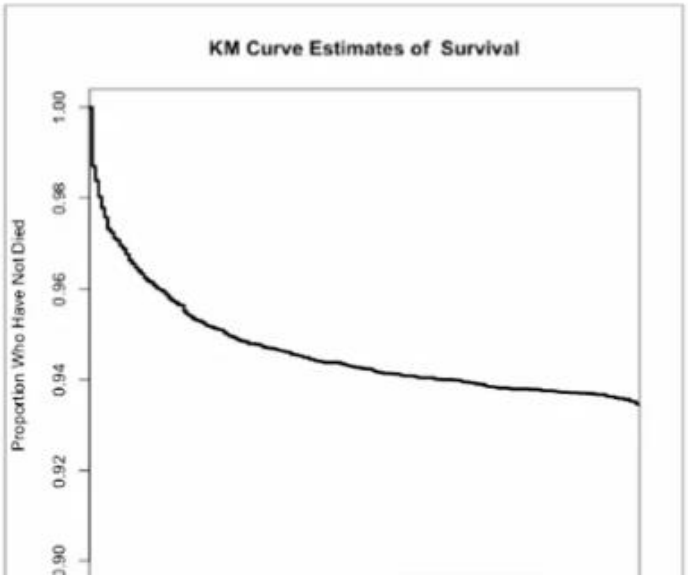
Example: Percentile Estimates of Time to Death, PBC Trial Data

- The Kaplan-Meier curve shows the estimated proportion of the original sample of 312 patients who survived (did not have the event of death) by the corresponding follow-up time



Example: Percentile Estimates of Time to Death, Infants in Nepal

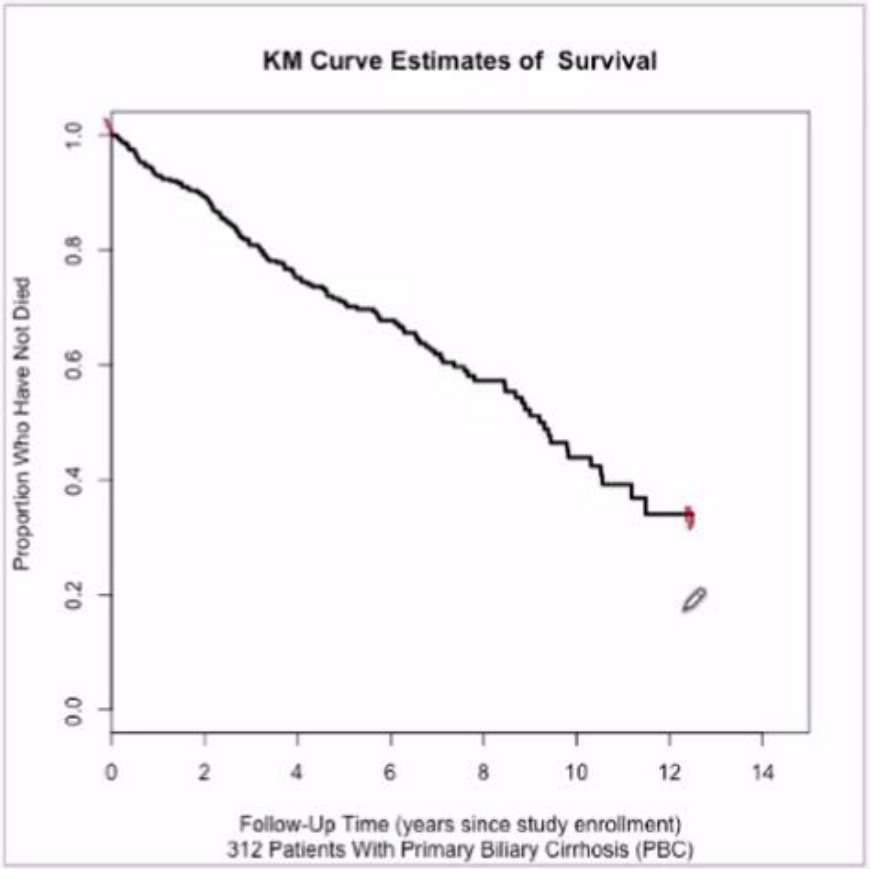
- Infant mortality rate in 6 months post birth: Kaplan-Meier (KM) curve, zoomed in. The curve tracks the estimated proportion of the original sample of 10,000+ children who survived (did not have the event, death) by the corresponding follow-up time

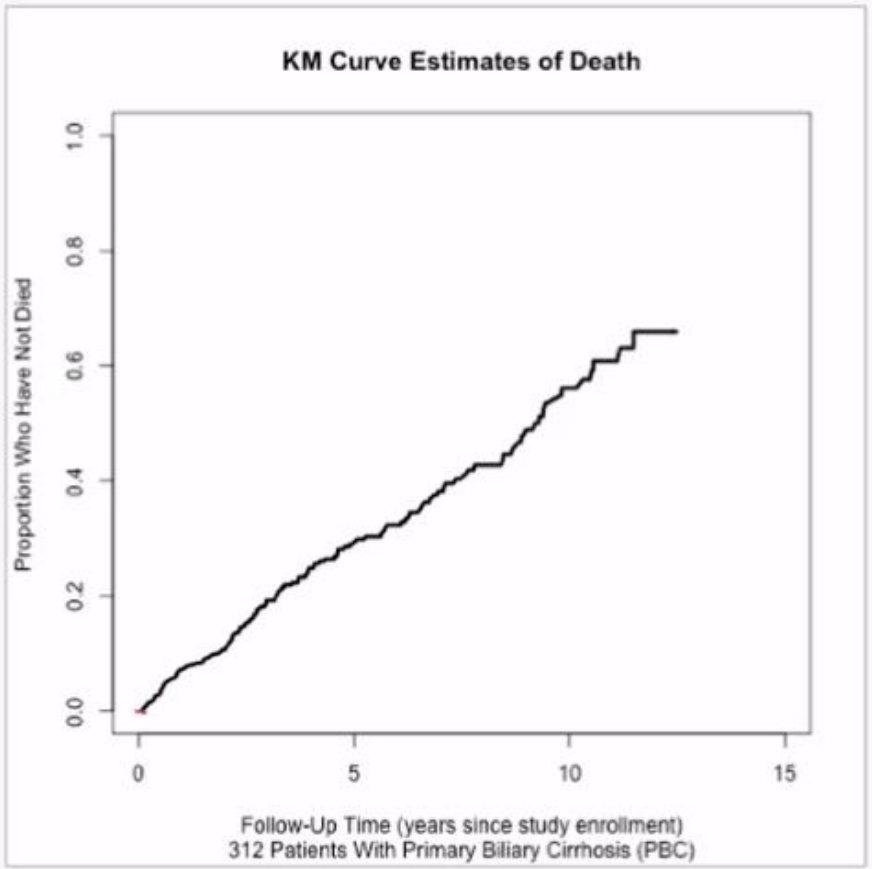


KM Compliment: Cumulative Survival Curve

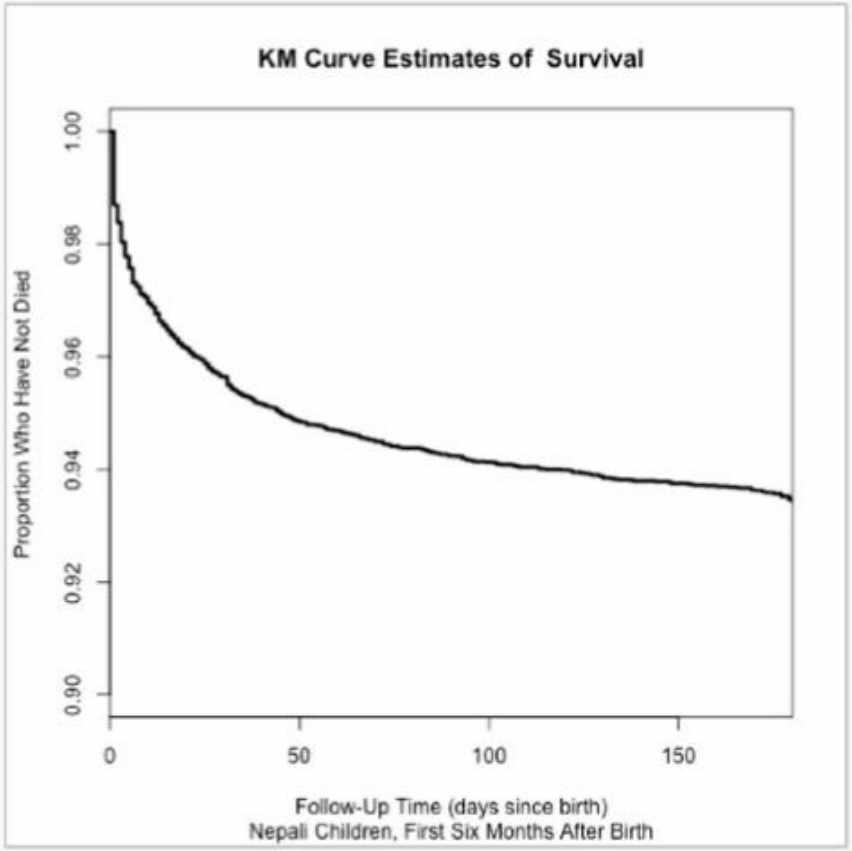
- Frequently, instead of presenting the KM with the examples just shown (more formally the "KM Survival Curve"), $\hat{S}(t)$, researchers will present $1 - \hat{S}(t)$, which shows the cumulative proportion of the original sample that has had the event by a certain time in the follow-up period

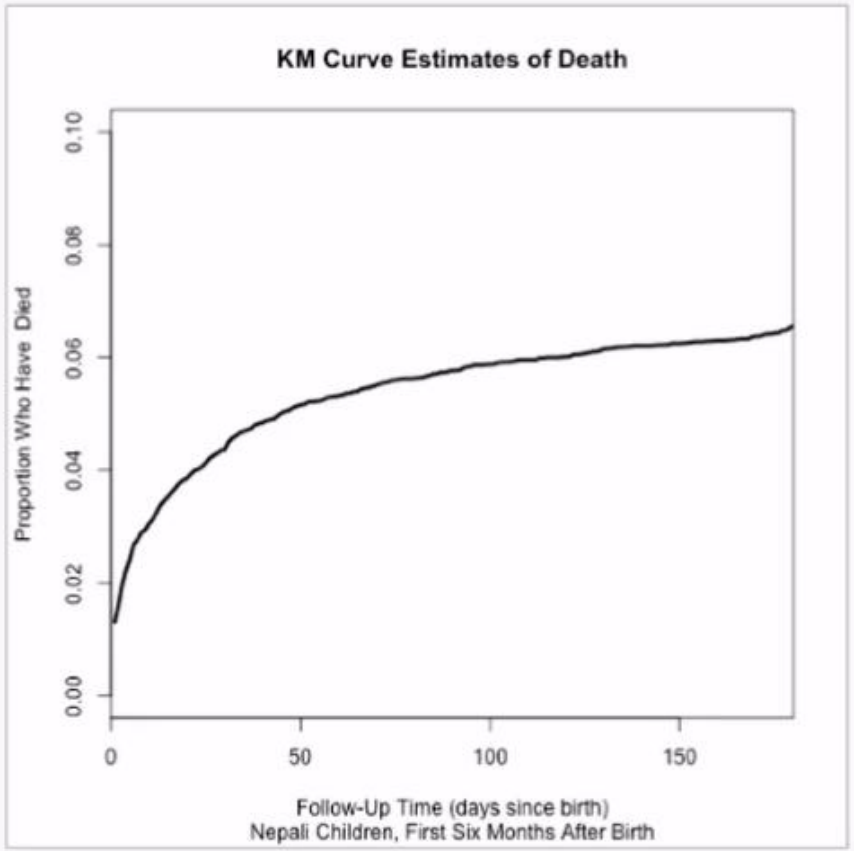
Examples of Both KM Presentations: PBC Trial Patient Mortality





Examples of Both KM Presentations: Infant Mortality in Nepal





Summary—1

- Kaplan-Meier curve estimates add "richness" and understanding to the time-to-event data from a sample by presenting the two dimensions to the data separately
- KM curves use all data in a sample (event and censoring time)
  - The censored observations provide information about who is at risk of having the event of interest at a given time in the follow-up period

Summary—2

- KM curves are summary statistics based on sample data
  - Estimate the underlying, unknown, true population survival curve
- Event time percentiles can be estimated via KM curves
- The are two complimentary ways to present the results of a KM curve estimate

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Comparing Time-to-Event Data Between Two (or More) Samples, Graphically

Learning Objectives

- Visually compare time-to-event data across two or more samples
- Explain how survival proportions across time can remain relatively high (and alternatively, the cumulative probability of having the event, relatively low) even if only a small proportion of the original study is around at the end of the study period

Example: Clinical Trial, Primary Biliary Cirrhosis (PBC): Incidence Rate Ratio

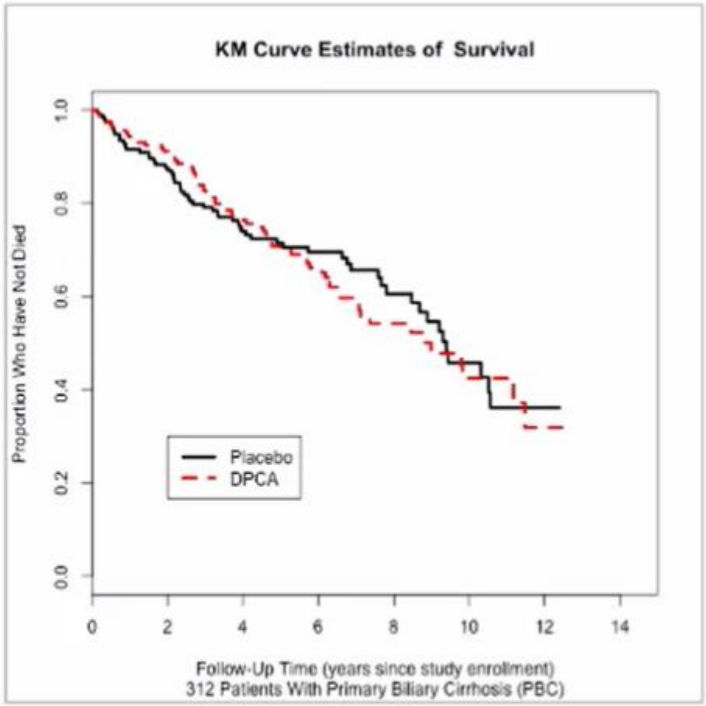
- Incidence rate ratio

$$\widehat{IRR} = \frac{\widehat{IR}_{DPCA}}{\widehat{IR}_{PLACEBO}} = \frac{0.075 \text{ deaths/PY}}{0.071 \text{ deaths/PY}} \approx 1.06$$

- Interpretations
  - The risk of death in the DPCA group (in the study follow-up period) is 1.06 time the risk in the placebo group
  - Subjects in the DPCA groups had 6% higher risk of death in the follow-up period when compared to the subjects in the placebo group

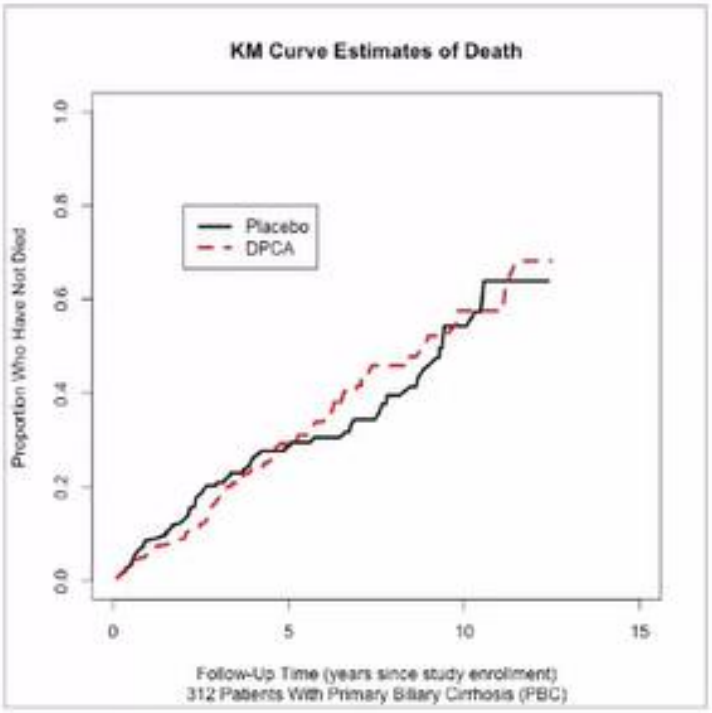
Example: Clinical Trial, PBC: Kaplan-Meier (KM) Curve Estimates of Survival

- KM curves for the DPCA and Placebo groups in the same graphic



Example: Clinical Trial, PBC: KM Curve Estimates of Death

- KM curves for the DPCA and Placebo groups in the same graphic (1-$\hat{S}(t)$ version)

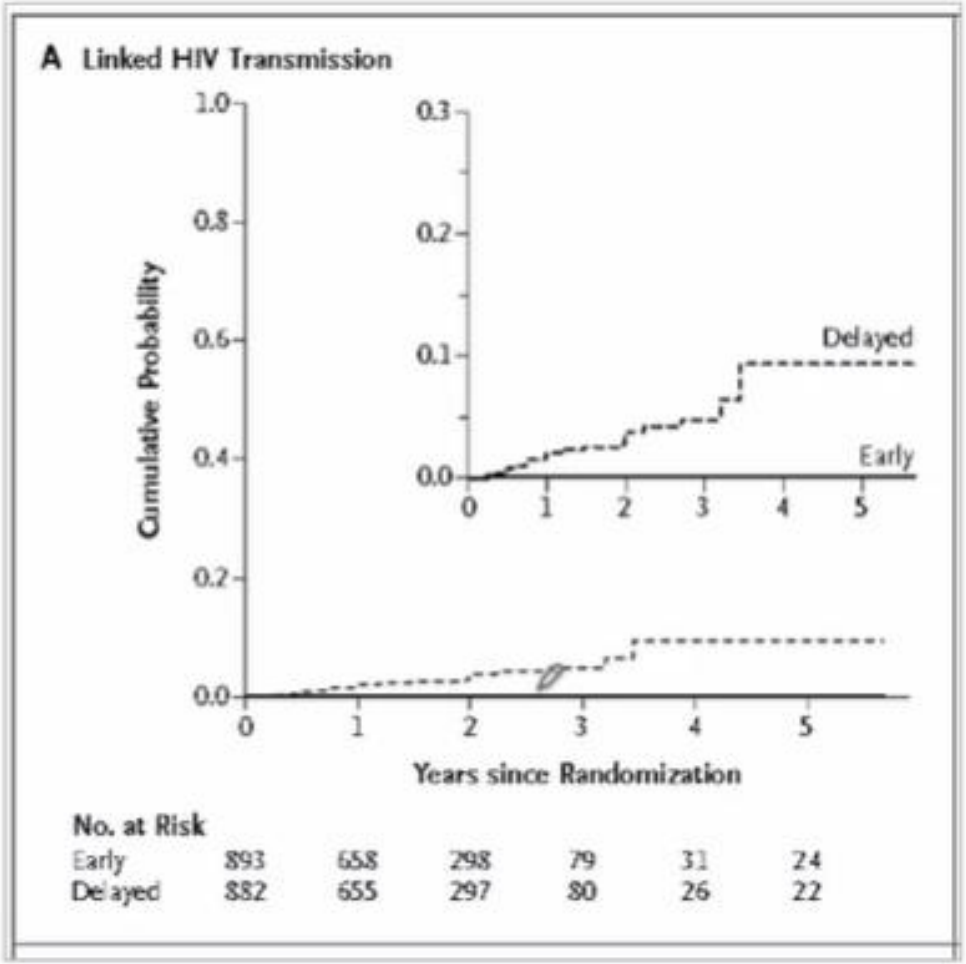


ART and Partner-to-Partner HIV Transmission: Results—1

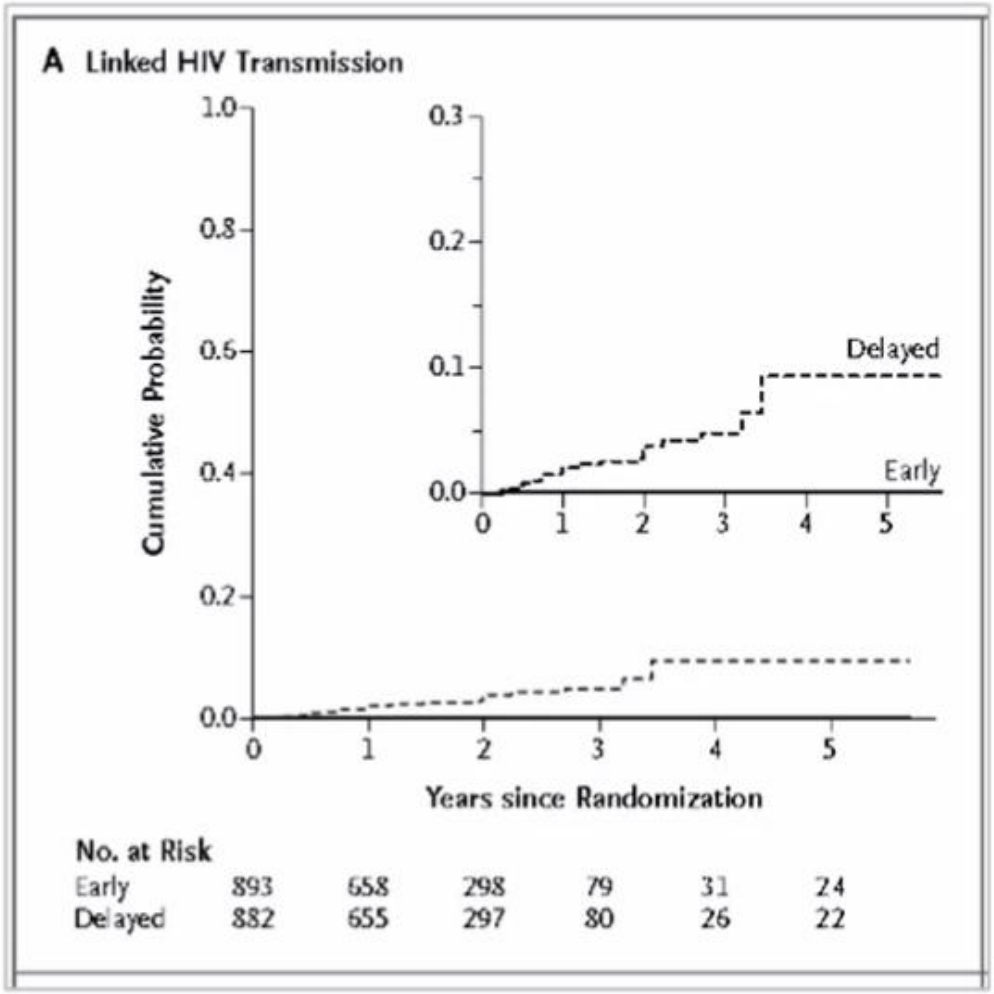
RESULTS

As of February 21, 2011, a total of 39 HIV-1 transmissions were observed (incidence rate, 1.2 per 100 person-years; 95% confidence interval [CI], 0.9 to 1.7); of these, 28 were virologically linked to the infected partner (incidence rate, 0.9 per 100 person-years, 95% CI, 0.6 to 1.3). Of the 28 linked transmissions, only 1 occurred in the early-therapy group (hazard ratio, 0.04; 95% CI, 0.01 to 0.27; $P<0.001$). Subjects receiving early therapy had fewer treatment end points (hazard ratio, 0.59; 95% CI, 0.40 to 0.88; $P=0.01$).

Example: ART and Partner-to-Partner HIV Transmission



Example: ART and Partner-to-Partner HIV Transmission



Summary

- Plotting KM survival curve estimates (or the cumulative event probability curves version) for multiple samples on the same graphic gives a nice overall visual comparison
- KM curves nicely complement incidence rate ratio estimates, and provide more detail
- Kaplan-Meier curve estimates are sample statistics and, hence, estimate the underlying unknown true survival curves in the populations from which the samples are taken