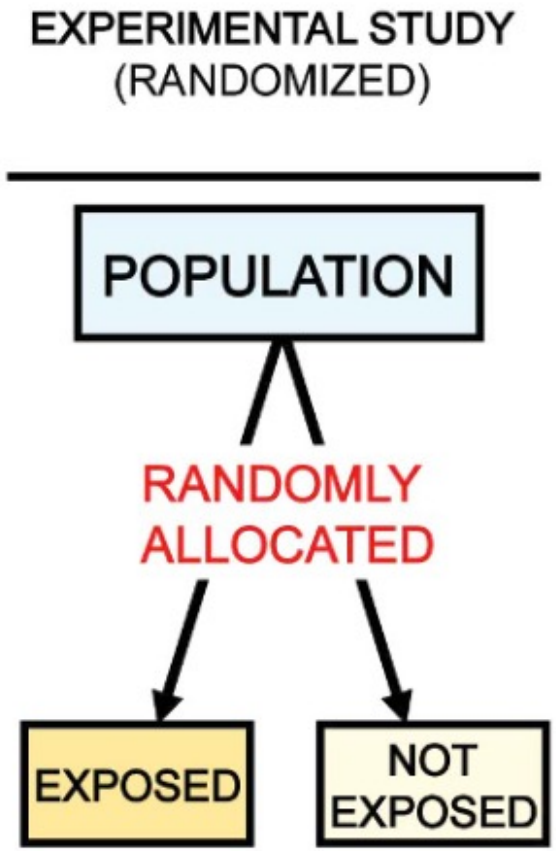
Lecture 9: Cohort Studies

Lecture prepared by Dr. Hailey Banack, PhD

Lecture Outline

1. RCTs vs. observational studies
2. Characteristics of cohort studies
3. Endpoints and exposures in cohort studies
4. Effect measures in cohort studies

Review



In an experimental study, exposure status is randomly assigned

However, you can't randomize all exposures.

Examples

- HPV and cervical cancer
- Poverty and oral health

Consider

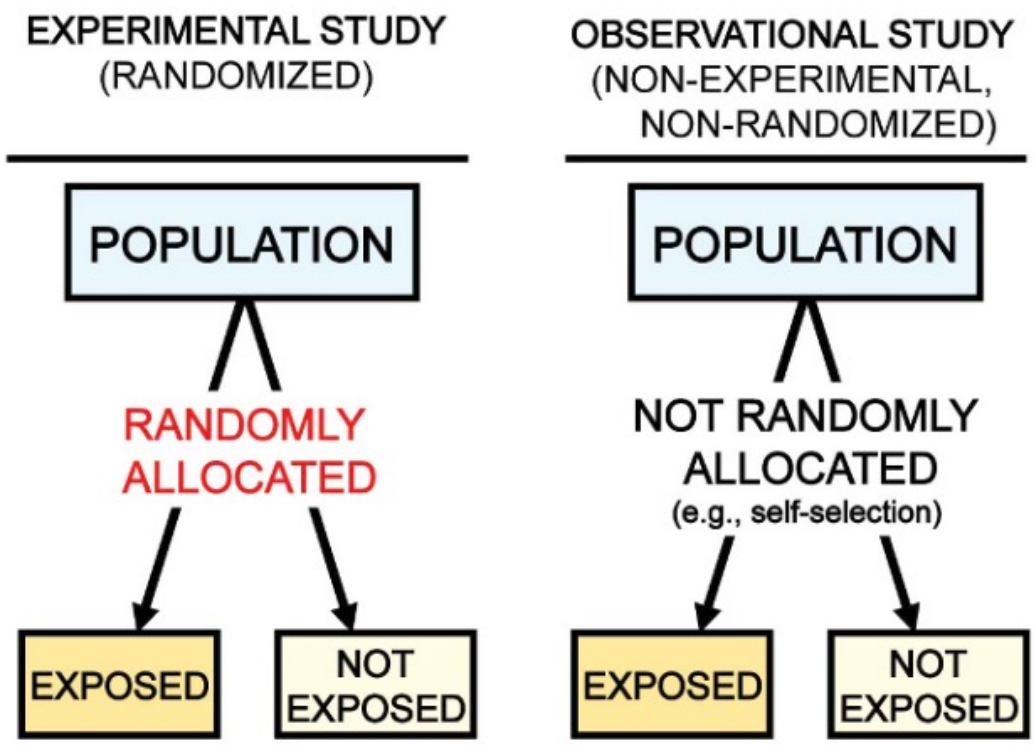
- What is the effect of needle stick injuries on the risk of HIV infection among health care workers?
- What is the effect of maternal smoking during pregnancy on risk of birth defects?
- What is the effect of race on cardiovascular disease risk?

For ethical and practical reasons, we cannot always study the effect of an exposure using an experimental design

RCTs versus observational studies

Key difference:

Use of randomization to separate individuals into exposure groups



Non-randomized studies

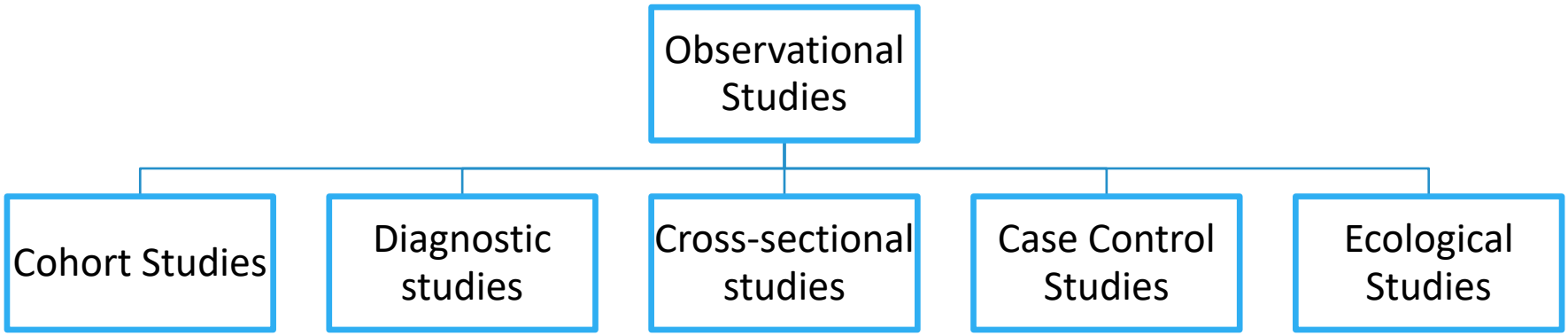
Pros:

- Simpler to implement
- Can study a broader range of exposures

Cons:

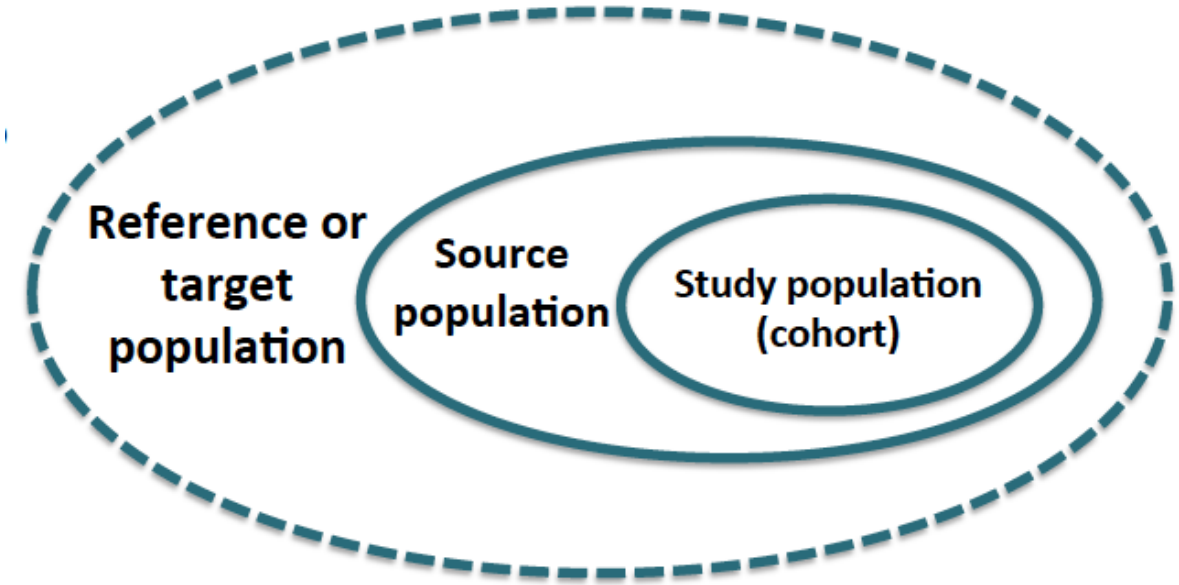
- Cannot assume the groups are exchangeable with respect to other risk factors for disease
  - Introduces uncertainty as to whether an observed association between the exposure and disease is *causal*

Types of Observational Studies



What is a cohort?

"Any designated group of individuals who are followed or traced over a period of time" (Rothman, 2012)



Shared characteristics

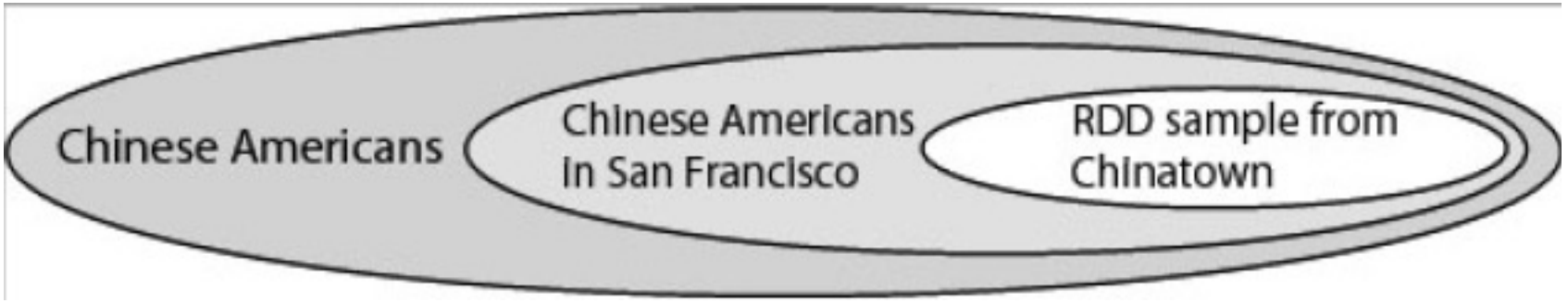
- Cohorts of individuals share common experience or condition:
  - Birth cohort
  - Demographic characteristics
  - Occupation
  - Shared exposure (e.g., smokers, dietary/vegan)

Definitions

- Target population is the broadest category for which you wish to extend your conclusions
- For practical reasons (efficiency/timeliness), your study will be conducted in a more restricted population – the source population
- From your study population, a sample will be drawn, and those will be your study participants
  - Occasionally, you can include the entire study population in your study though it's usually not feasible or necessary

Example

Does daily consumption of soy sauce by Chinese Americans cause GI cancer?



Your target population is Chinese Americans. Suppose that you are based in San Francisco, CA and it is not feasible for you to conduct a national study.

A practical source population would be Chinese Americans living in San Francisco.

To obtain subjects for your study, you might choose to random-digit dial (RDD) using the telephone prefixes used in San Francisco's Chinatown.

Important question

Who are my results generalizable to?

This question has very important implications for how the study results are used

Definition: Generalizability

- Also known as external validity
- Who do my study results apply to?
- Can the study results be generalized to different persons, settings, and times?

(Steckler & McLeroy, 2008)

Ask yourself these questions

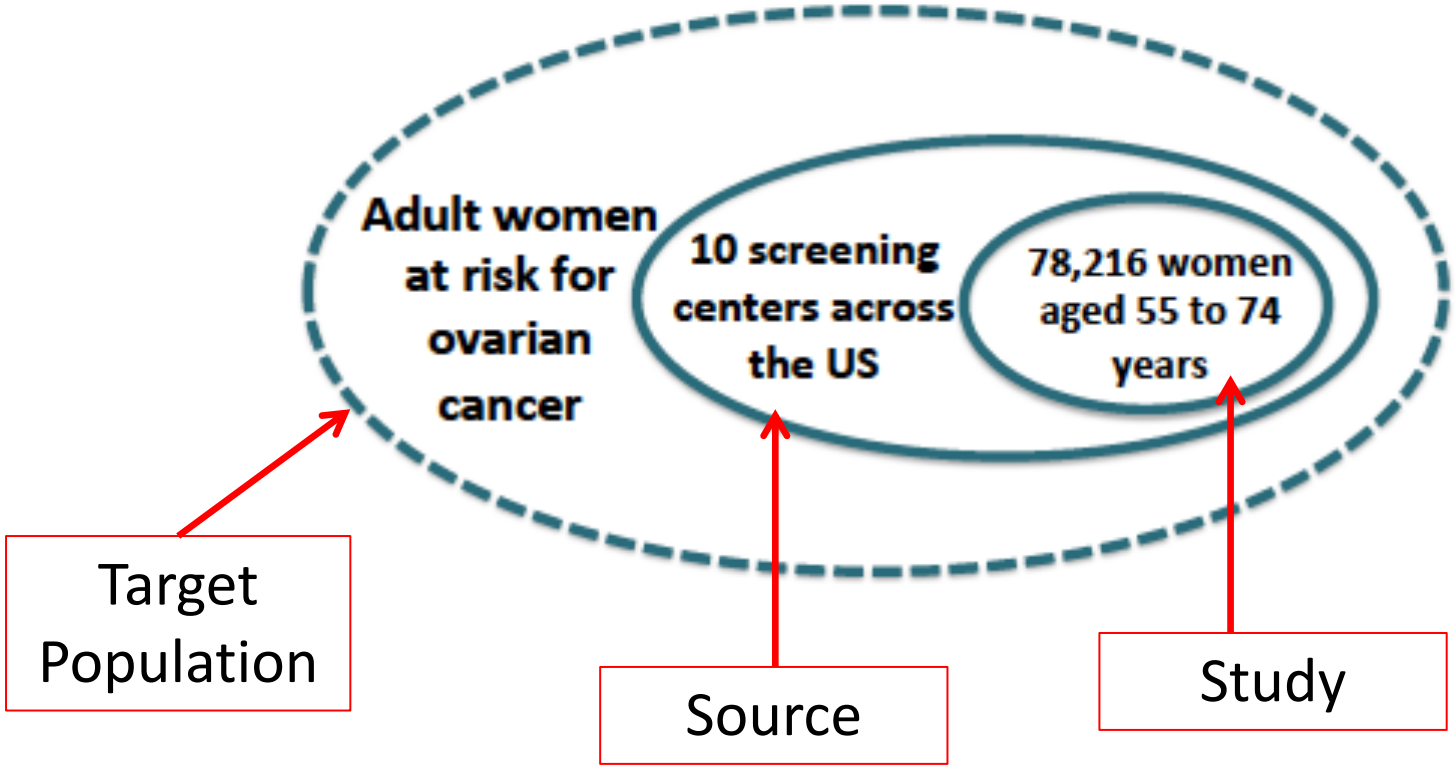
1. Who do you want to generalize to?

2. What population do you have access to?

3. Who is in your study?

Example

Study of the effect of screening on ovarian cancer mortality



Famous Cohorts

- Early era of cohort studies 1940s-1950s
  - Framingham Heart Study
  - Japanese atomic bomb survivors
  - British doctors cohort study

- Key features: cohort size, richness of data, sustained follow-up

Framingham Heart Study

- Begun in 1948 to address rising incidence of CVD
- Key features in its success:
  - Selection of a small and cooperative community
  - Sustained financial support
  - Rigorous and standardized protocols for data collection
- Third generation of family members now enrolled (grandchildren of the original cohort!)
- >1200 publications over 50 years!

Original Cohort

The Original Cohort of the Framingham Heart Study consisted of 5,209 respondents of a random sample of 2/3 of the adult population of Framingham, Massachusetts, 30 to 62 years of age by household, in 1948. Exam 28 for the Original Cohort ended in December of 2005. Exam 29 for the Original Cohort began in April of 2006.

AGE-SEX DISTRIBUTION AT ENTRY (1948)				
Age	**29-39**	**40-49**	**50-62**	**Totals**
Men	835	779	722	2,336
Women	1,042	962	869	2,873
Totals	1,877	1,741	1,591	5,209

Offspring Cohort

The Offspring Study was initiated in 1971 when the need for establishing a prospective epidemiologic study of young adults was recognized. A sample of 5,124 men and women, consisting of the offspring of the Original Cohort and their spouses was recruited. Offspring Exam 8 began in March 2005.

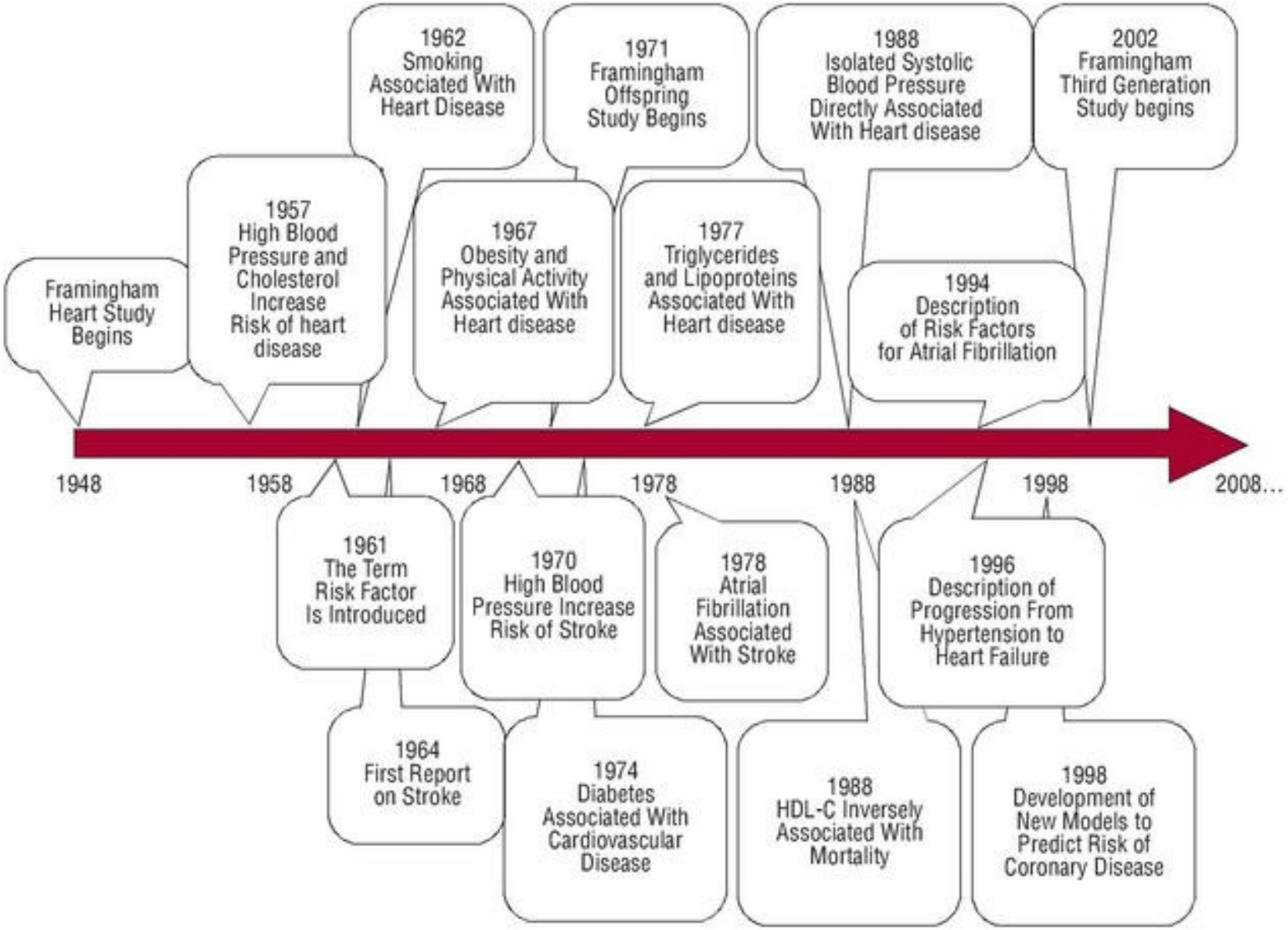
AGE-SEX DISTRIBUTION AT ENTRY (1971)								
Age	**<10**	**10-19**	**20-29**	**30-39**	**40-49**	**50-59**	**60-70**	**Totals**
Men	0	126	543	789	694	293	38	2,483
Women	6	113	692	835	740	242	13	2,641
Totals	6	239	1,235	1,624	1,434	535	51	5,124

Third Generation Cohort (Gen III)

A recent major component of the Framingham Heart Study protocol has been the enrollment and examination of a third generation of participants which will provide greater resources of phenotypic and genotypic information. During Offspring Exam Cycles 6 and 7, the Offspring participants were asked to update information about their children. To assess interest in participation prior to the start of clinic exams, 5,500 letters and response cards were sent in November 2001 to prospective third generation participants who had at least one parent in the Offspring Study and would be at least 20 years old by the close of the first exam cycle. Later an additional 1,241 invitation letters were sent. A prioritization of the recruitment list was prepared. Considerations were given to family size, completeness of data, stored DNA and responsiveness of the Gen III members of the families.

By consenting to and completing Exam 1 of Gen III, the participant was considered enrolled in the Framingham Heart Study Gen III. Special efforts were being made to complete sibships and families in the course of enrollment. A recruitment target of 4,095 Gen III participants was achieved by July of 2005. The details of clinic attendance for Gen III are described in the table to follow.

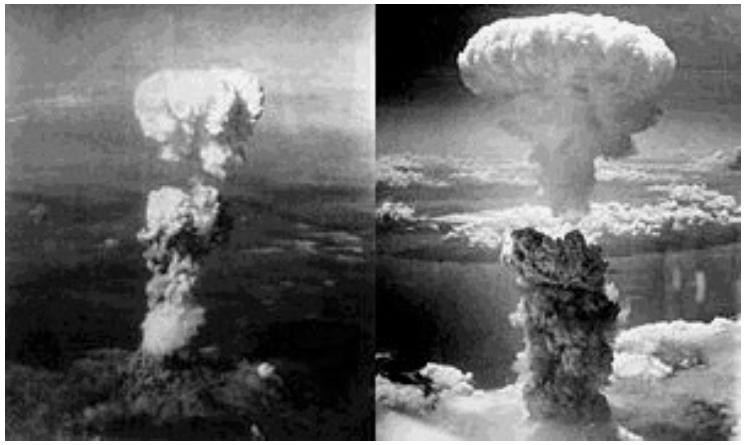
GEN3: AGE-SEX DISTRIBUTION AT ENTRY (THROUGH 2005)								
Age	**19**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-79**	**Totals**
Men	4	220	656	737	276	19	1	1,913
Women	3	262	759	848	293	16	1	2,182
Totals	7	482	1,415	1,585	569	35	2	4,095



(O'Donnell & Elosua, 2008)

Japanese Atomic Bomb Survivor Study

- Addressed consequences of ionizing radiation exposure
- Unlike Framingham study (which was designed to test multiple hypotheses) this study had only one goal: to address the consequences of ionizing radiation exposure
- Radiation doses for sampled survivors were reconstructed and they were entered into a cohort study with regular medical exams
- This study provides the underpinnings of radiation standards worldwide



Atomic bomb mushroom clouds Hiroshima (L) and Nagasaki (R)

Atomic Bomb Survivors

- An in-depth follow-up study of mortality in the study population of 120,000 persons who survived the bombings of Hiroshima and Nagasaki has continued since 1950
- Follow-up studies also have been conducted on in-utero-exposed persons and first-generation offspring of the survivors.
- Radiation effects: leukemia, tumors, cataracts, thyroid disease, growth delay
- No evidence of genetic effects in children of A-bomb survivors

British Doctor's Study

- Follow-up to initial observations and circumstantial evidence that smoking *may* be bad for health
- In October 1951, the researchers wrote to all registered physicians in the United Kingdom and recruited, 40,701 participants.
- No further cohorts were recruited. Because of the limited sample size females were excluded from most analyses.
- 1956: tobacco linked to lung cancer
  - Study demonstrated relevance of epidemiology and statistics in questions of public health

50 year follow-up

Doll et al., 2004

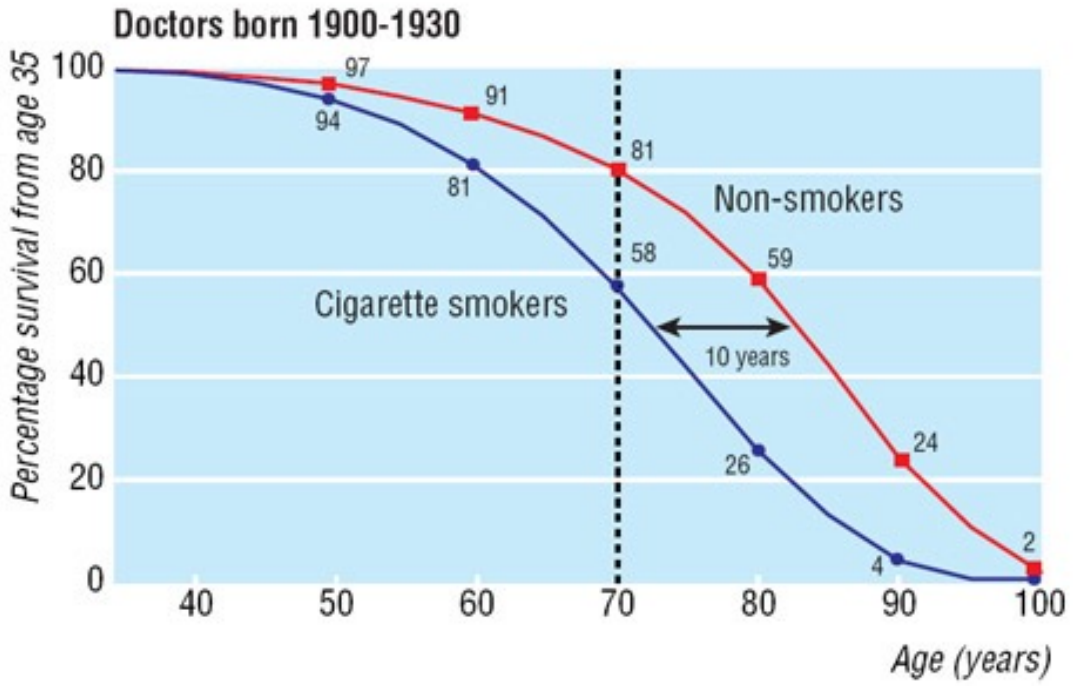


Fig 3 Survival from age 35 for continuing cigarette smokers and lifelong non-smokers among UK male doctors born 1900-1930, with percentages alive at each decade of age

Generalizability

- Do the results from the Framingham Heart study apply to the entire US population? The population of California? A population of adults living in South America?
- Do the results from the Japanese-atomic bomb survivor study apply to X-ray technicians working in hospitals today? To Japanese-Americans living in Honolulu?

Cohort Studies

Study Design

- A cohort study tracks two or more groups forward from exposure to outcome
- Selects a group of exposed and unexposed individuals and follows them over time for development of the outcome of interest

Note:
Participants must be disease-free at the start of a cohort study (for your outcome of interest)

Marching Toward Outcomes

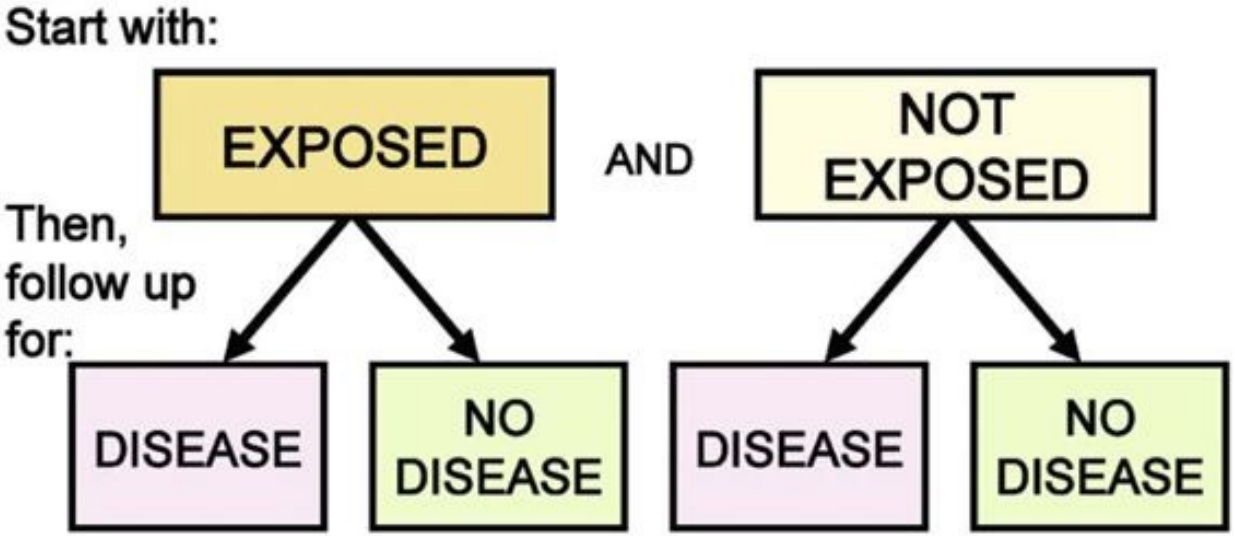
The term cohort has military, not medical, roots. A cohort was a 300–600-man unit in the Roman army. A cohort study consists of bands or groups of persons marching forward in time from an exposure to one or more outcomes.

(Grimes et al., 2003)

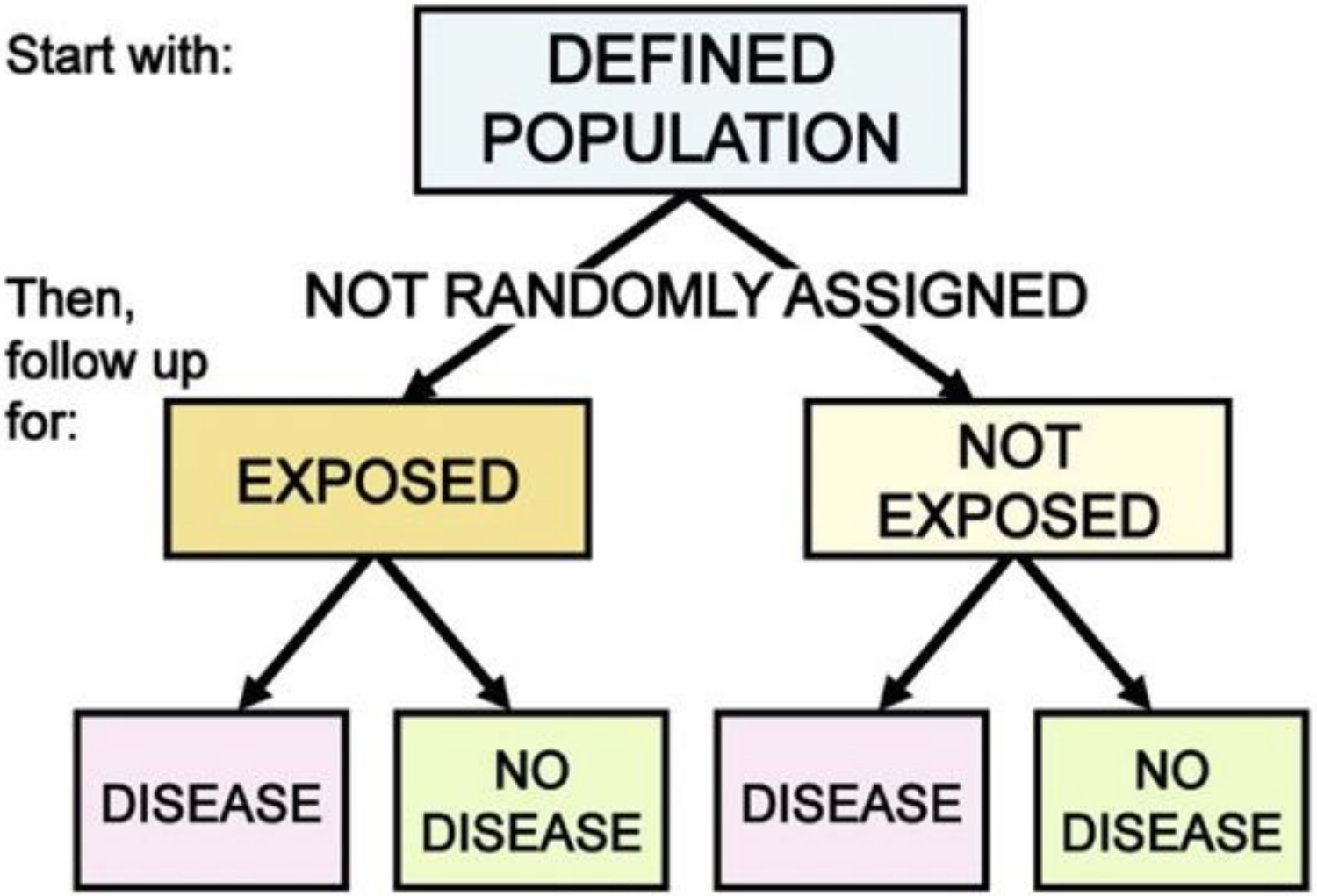
Selecting a study population

Study population = exposed group + unexposed group

1. Based on exposure status

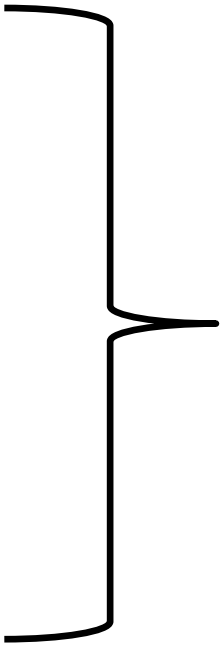


2. Based on a defined population (eg., geography, occupation)



Types of Cohort Studies

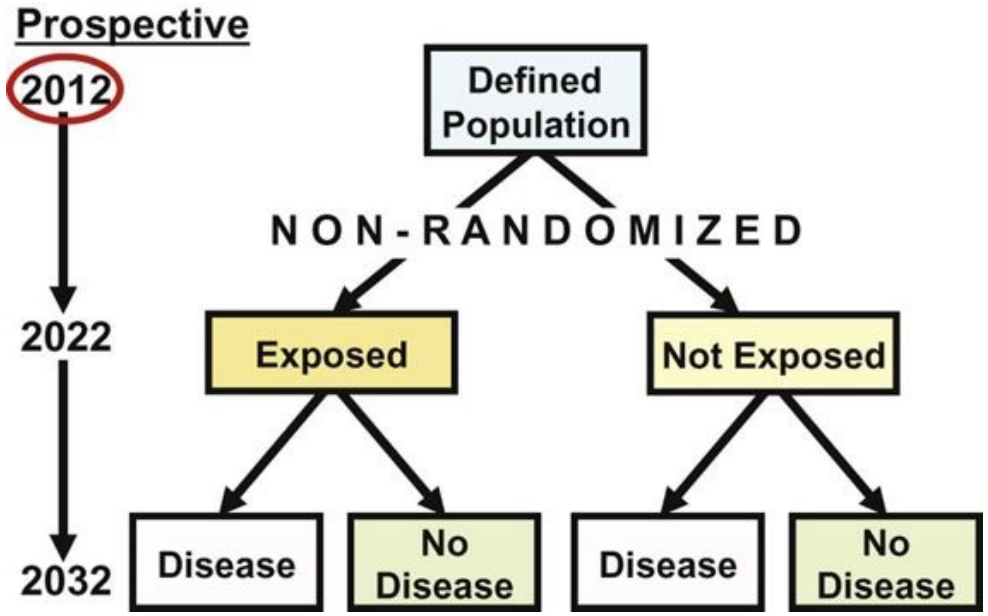
1. Prospective
2. Retrospective
3. Ambispective



Differ with respect to the timing of data collection but the study design is still the same

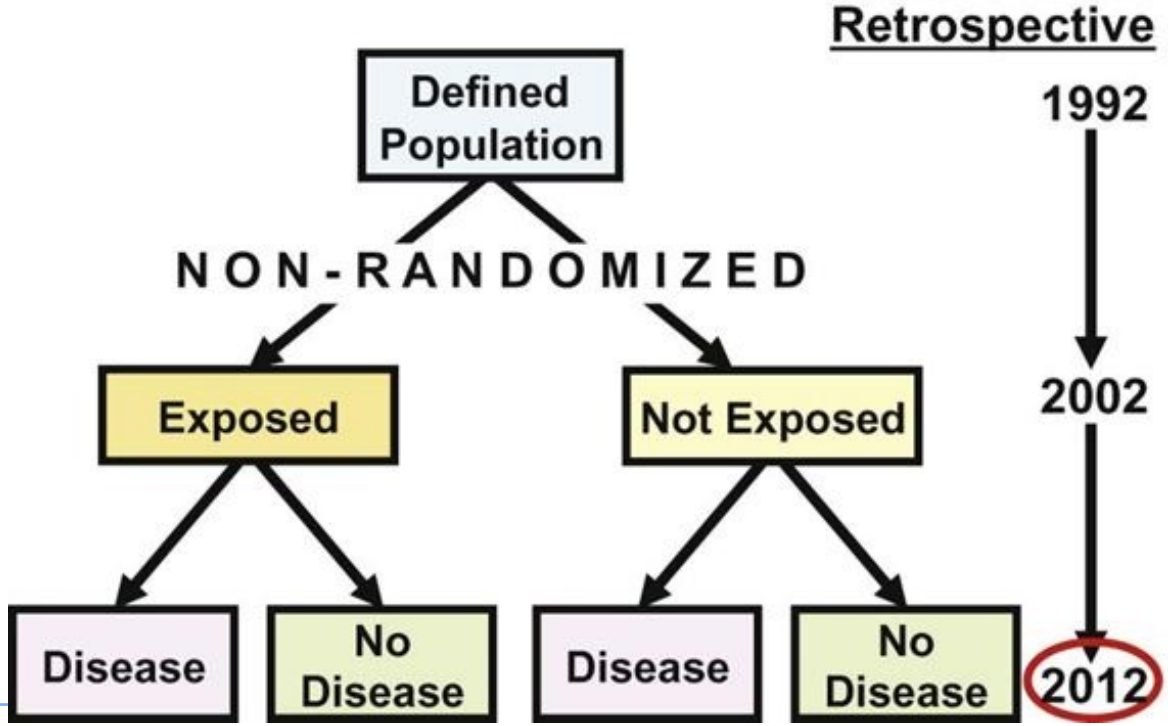
Prospective Cohort Study

- Investigator identifies a study cohort
- Data on individual exposure status collected at start (baseline) and updated during the study
- Exposure groups followed over time for development of disease



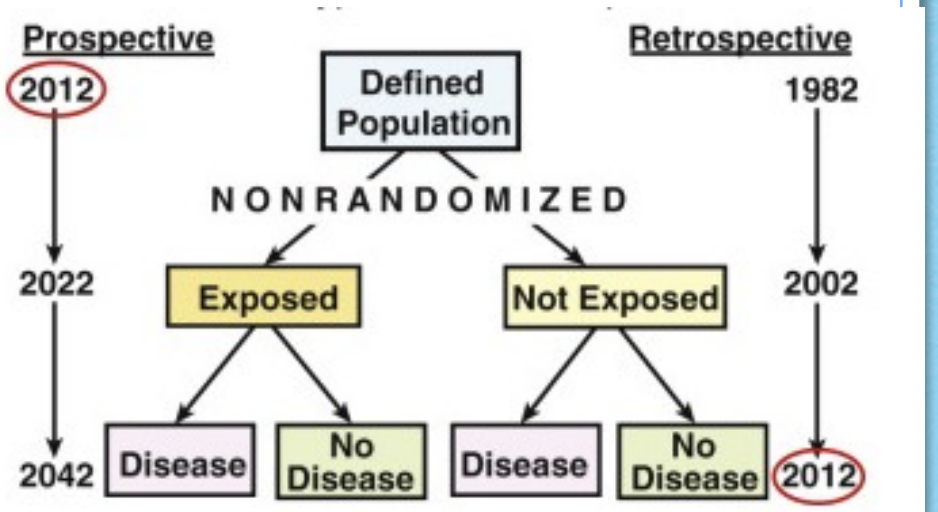
Retrospective Cohort Study

- Historical records used to identify cohort and obtain data on individual exposure status
  - Hospital records, administrative data



Comparison

- Design for both prospective and retrospective studies is the same
  - Comparing outcomes in exposed versus unexposed group
- ONLY difference is calendar time
  - *Prospective cohort design: exposure ascertained as it occurs during follow-up; participants followed over time to see who develops disease/outcome*
  - *Retrospective cohort design: exposure is ascertained from past records and outcome is ascertained in present day*

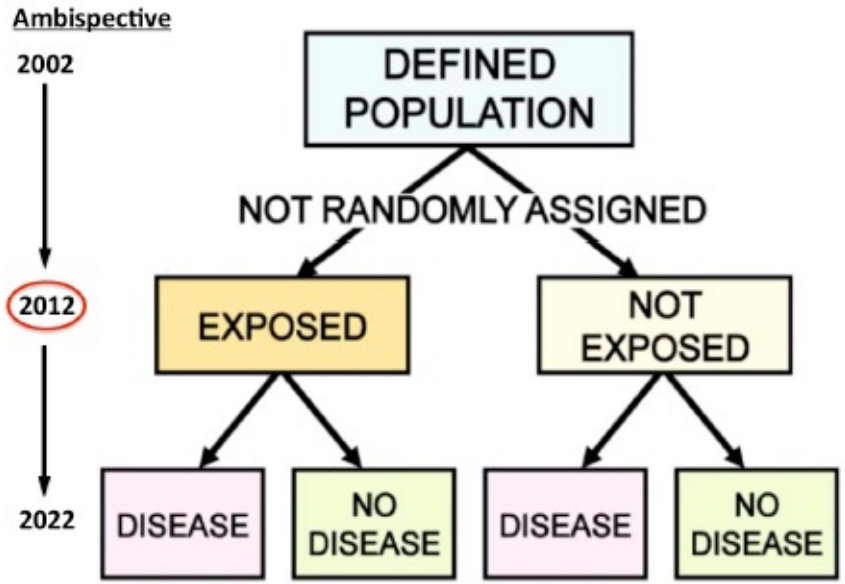


Comparing Cohort Studies

	PROSPECTIVE	RETROSPECTIVE
Pros	- Greater flexibility and better data quality - Low risk of bias in exposure measurement since outcome status not yet known	- Requires relatively less time to conduct study
Cons	- Requires relatively more time and resources to complete study	- Poorer flexibility and data quality due to reliance on historical data - Risk of bias in exposure measurement if outcome status known
Better when	- Disease has short latency period - Want high-quality data	- Disease has long latency period - Want to save time and/or $

Ambispective Cohort Study

- Also known as bidirectional cohort study design
- Mixture of prospective and retrospective studies



Strengths of Cohort Studies

- Establish temporal order between exposure and disease (evidence of causality)
- Can study multiple outcomes
- Useful for rare exposures
- Direct calculation of incidence in exposed and unexposed group which allows for calculation of risk difference/ratios

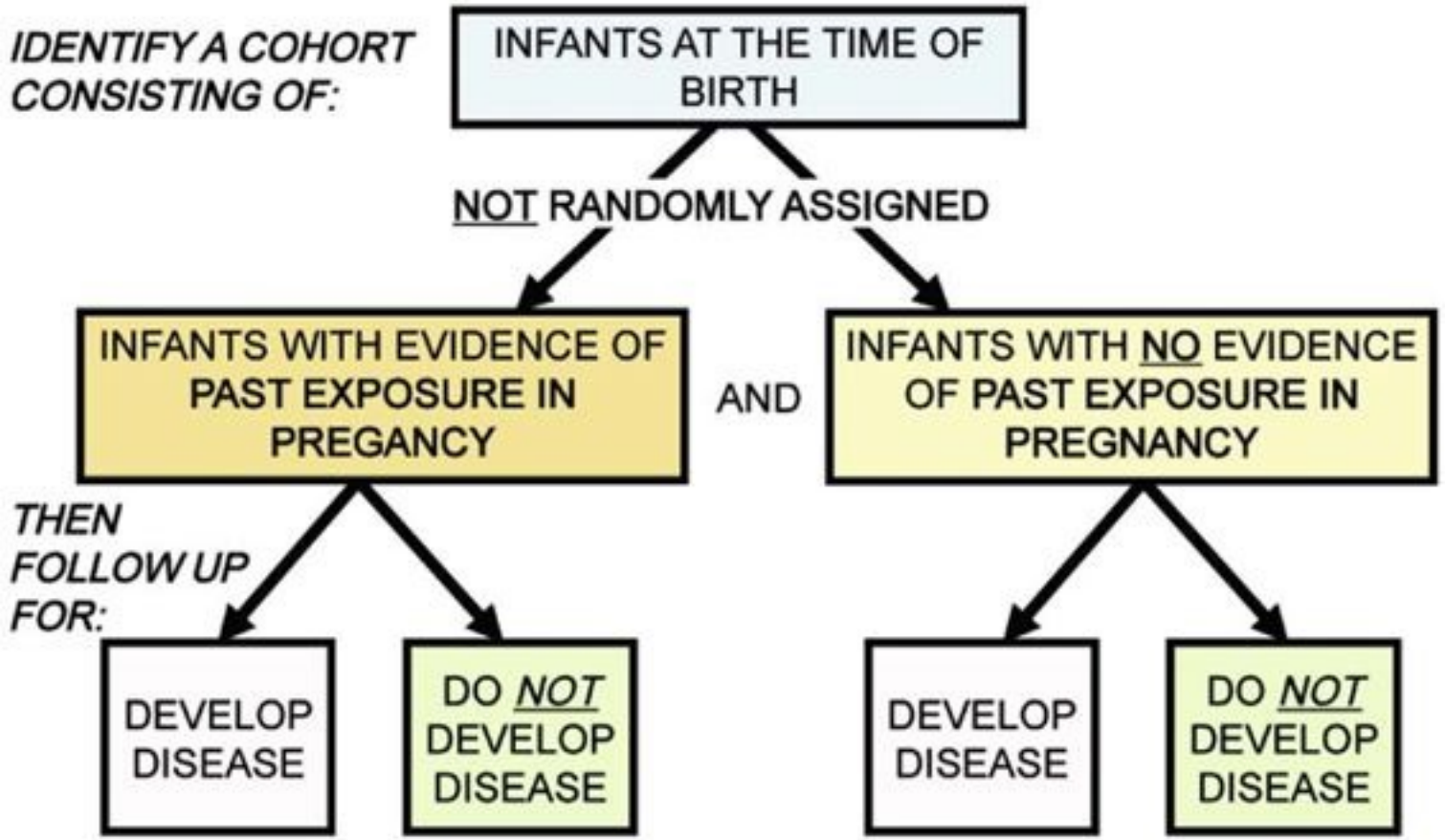
Weaknesses of Cohort Studies

- Costly and time-consuming
- Losses to follow-up
- Inefficient for rare diseases
- Requires large study population

Using cohort studies

- When sufficient evidence exists to suggest and association between disease and exposure
  - Exposure must be worth investigating
- When there is a reasonable latency period between exposure and development of disease
- When there are opportunities to minimize losses to follow-up

Example: Prenatal exposures and Parkinson's Disease



Cohort Studies vs. RCTs

Comparison of cohort studies and randomised controlled trials

Item	Cohort studies	Randomised controlled trials
Populations studied	Diverse populations of patients who are observed in a range of settings	Highly selected populations recruited on the basis of detailed criteria and treated at selected sites
Allocation to the intervention	Based on decisions made by providers or patients	Based on chance and controlled by investigators
Outcomes	Can be defined after the intervention and can include rare or unexpected events	Primary outcomes are determined before patients are entered into study and are focused on predicted benefits and risks
Follow-up	Many cohort studies rely on existing experience (retrospective studies) and can provide an opportunity for long follow-up	Prospective studies; often have short follow-up because of costs and pressure to produce timely evidence
Analysis	Sophisticated multivariate techniques may be required to deal with confounding	Analysis is straightforward

Rochon et al., 2005

Compare readings
RCT and cohort study

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting

Noa Dagan, M.D., Noam Barda, M.D., Eldad Kepten, Ph.D., Oren Miron, M.A., Shay Perchik, M.A., Mark A. Katz, M.D., Miguel A. Hernán, M.D., Marc Lipsitch, D.Phil., Ben Reis, Ph.D., and Ran D. Balicer, M.D.

ABSTRACT

BACKGROUND

As mass vaccination campaigns against coronavirus disease 2019 (Covid-19) commence worldwide, vaccine effectiveness needs to be assessed for a range of outcomes across diverse populations in a noncontrolled setting. In this study, data from Israel's largest health care organization were used to evaluate the effectiveness of the BNT162b2 mRNA vaccine.

METHODS

All persons who were newly vaccinated during the period from December 20, 2020, to February 1, 2021, were matched to unvaccinated controls in a 1:1 ratio according to demographic and clinical characteristics. Study outcomes included documented infection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), symptomatic Covid-19, Covid-19–related hospitalization, severe illness, and death. We estimated vaccine effectiveness for each outcome as one minus the risk ratio, using the Kaplan–Meier estimator.

RESULTS

Each study group included 596,618 persons. Estimated vaccine effectiveness for the study outcomes at days 14 through 20 after the first dose and at 7 or more days after the second dose was as follows: for documented infection, 46% (95% confidence interval [CI], 40 to 51) and 92% (95% CI, 88 to 95); for symptomatic Covid-19, 57% (95% CI, 50 to 63) and 94% (95% CI, 87 to 98); for hospitalization, 74% (95% CI, 56 to 86) and 87% (95% CI, 55 to 100); and for severe disease, 62% (95% CI, 39 to 80) and 92% (95% CI, 75 to 100), respectively. Estimated effectiveness in preventing death from Covid-19 was 72% (95% CI, 19 to 100) for days 14 through 20 after the first dose. Estimated effectiveness in specific subpopulations assessed for documented infection and symptomatic Covid-19 was consistent across age groups, with potentially slightly lower effectiveness in persons with multiple coexisting conditions.

POSTMENOPAUSAL ESTROGEN THERAPY AND CARDIOVASCULAR DISEASE

Ten-Year Follow-up from the Nurses' Health Study

MEIR J. STAMPFER, M.D., GRAHAM A. COLDITZ, M.B., B.S., WALTER C. WILLETT, M.D., JOANN E. MANSON, M.D., BERNARD ROSNER, PH.D., FRANK E. SPEIZER, M.D., AND CHARLES H. HENNEKENS, M.D.

Abstract *Background.* The effect of postmenopausal estrogen therapy on the risk of cardiovascular disease remains controversial. Our 1985 report in the *Journal,* based on four years of follow-up, suggested that estrogen therapy reduced the risk of coronary heart disease, but a report published simultaneously from the Framingham Study suggested that the risk was increased. In addition, studies of the effect of estrogens on stroke have yielded conflicting results.

Methods. We followed 48,470 postmenopausal women, 30 to 63 years old, who were participants in the Nurses' Health Study and who did not have a history of cancer or cardiovascular disease at base line. During up to 10 years of follow-up (337,854 person-years), we documented 224 strokes, 405 cases of major coronary disease (nonfatal myocardial infarctions or deaths from coronary causes), and 1263 deaths from all causes.

Results. After adjustment for age and other risk factors, the overall relative risk of major coronary disease in women currently taking estrogen was 0.56 (95 percent confidence interval, 0.40 to 0.80); the risk was significantly reduced among women with either natural or surgical menopause. We observed no effect of the duration of estrogen use independent of age. The findings were similar in analyses limited to women who had recently visited their physicians (relative risk, 0.45; 95 percent confidence interval, 0.31 to 0.66) and in a low-risk group that excluded women reporting current cigarette smoking, diabetes, hypertension, hypercholesterolemia, or a Quetelet index above the 90th percentile (relative risk, 0.53; 95 percent confidence interval, 0.31 to 0.91). The relative risk for current and former users of estrogen as compared with those who had never used it was 0.89 (95 percent confidence interval, 0.78 to 1.00) for total mortality and 0.72 (95 percent confidence interval, 0.55 to 0.95) for mortality from cardiovascular disease. The relative risk of stroke when current users were compared with those who had never used estrogen was 0.97 (95 percent confidence interval, 0.65 to 1.45), with no marked differences according to type of stroke.

Conclusions. Current estrogen use is associated with a reduction in the incidence of coronary heart disease as well as in mortality from cardiovascular disease, but it is not associated with any change in the risk of stroke. (N Engl J Med 1991; 325:756-62.)

PICO

Population, intervention (exposure), comparison, outcome

The Nurses' Health Study Cohort

The Nurses' Health Study began in 1976, when 121,700 female registered nurses in the United States completed questionnaires sent to them by mail about their medical history, including previous cardiovascular disease, menopause, diabetes, hypertension, high serum cholesterol levels, and parental myocardial infarction. We included questions on height, weight, smoking, the use of postmeno-

Identification and Confirmation of Cardiovascular End Points

The study end points included nonfatal myocardial infarction, fatal coronary heart disease, coronary-artery bypass grafting or angioplasty, fatal and nonfatal stroke, total cardiovascular mortality, and deaths from all causes after the return of the 1976 questionnaire but before June 1, 1986. Nurses who reported having a nonfatal myocardial infarction or stroke on a follow-up questionnaire were asked for permission for a study investigator to review their medical records. Nonfatal myocardial infarctions were considered confirmed by hospital records if they met the World Health Organization criteria[7] (i.e., symptoms plus either cardiac-enzyme elevations or diagnostic electrocardiographic changes). Myocardial infarctions that required hospitalization and for which confirmatory information was obtained by interview or letter, but for which no medical records were obtainable, were designated as probable. Thus, infarc-

Ascertainment of Estrogen Use

In 1976 the women were asked whether they had taken hormone supplements after menopause, and if so, for how long. Information on hormone use, including the type taken, was updated in the subsequent questionnaires sent every two years through 1986, with explicit questions about current use and duration of use in the intervening period. Because no information on current use was explicitly requested on the 1976 questionnaire, we considered women to have been current estrogen users for the 1976–1978 period if the duration of their estrogen use was equal (within 12 months) to the interval between menopause and the date of completion of the questionnaire. Women whose duration of hormone use was more than 12 months shorter than this interval were considered former users. The daily dose of conjugated estrogens was obtained beginning in 1980.

Statistical Analysis

For each participant, person-months were allocated to the categories of hormone use according to the data reported in 1976 and updated at each two-year interval according to information obtained subsequently. Follow-up for a participant ended with a diagnosis of cardiovascular disease or death. If no questionnaire was returned for a two-year follow-up period, the most recent data were applied to the subsequent follow-up interval. If a woman's previous status had been current hormone use, however, she was classified in the update as having used hormones at some time, but current or former use was not specified.

RCT: flow diagram

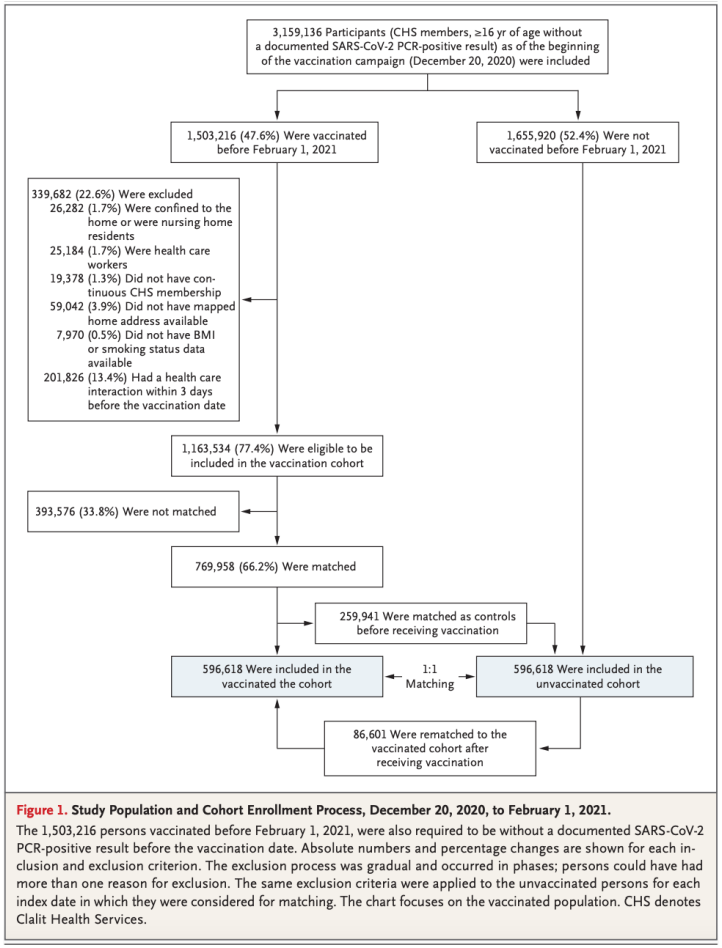


Figure 1. Study Population and Cohort Enrollment Process, December 20, 2020, to February 1, 2021.
The 1,503,216 persons vaccinated before February 1, 2021, were also required to be without a documented SARS-CoV-2 PCR-positive result before the vaccination date. Absolute numbers and percentage changes are shown for each inclusion and exclusion criterion. The exclusion process was gradual and occurred in phases; persons could have had more than one reason for exclusion. The same exclusion criteria were applied to the unvaccinated persons for each index date in which they were considered for matching. The chart focuses on the vaccinated population. CHS denotes Clalit Health Services.

Also called a CONSORT diagram

CONSORT
http://www.consort-statement.org/

-25 item checklist for reporting of RCT

Measuring Exposures and Outcomes

Exposures

Time-fixed

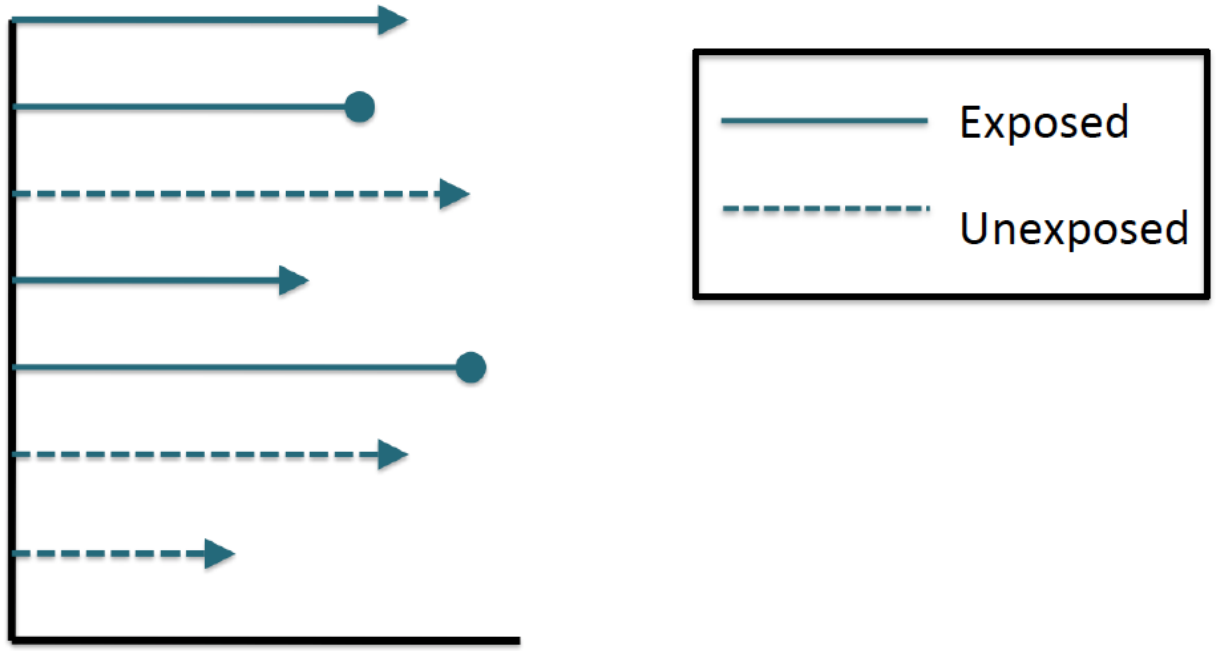
- Exposure stays constant over time
- Genotype, sex, birth-weight, environmental exposure
- Once you're exposed, you can't get 'un-exposed'

Time-varying

- Exposure status changes over time
- Health behaviours, medication use, employment status
- Can transition between exposed and unexposed groups

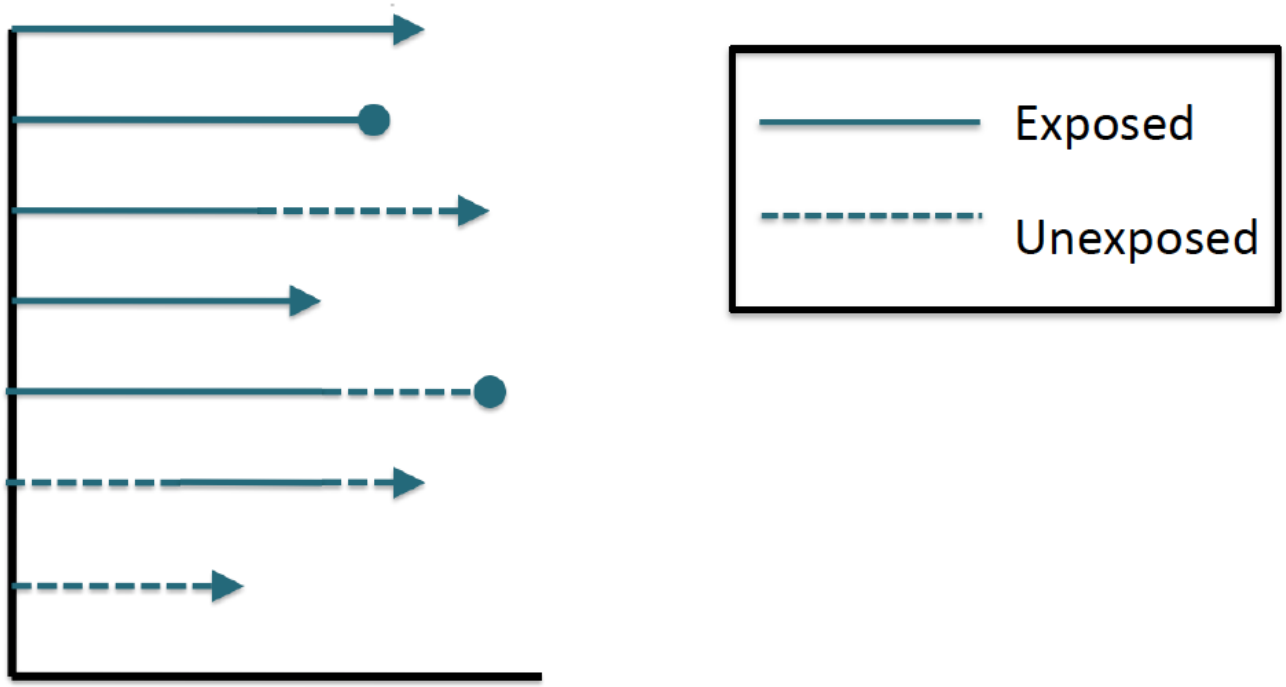
Time-fixed variables

All person time for an individual is categorized as either exposed or unexposed



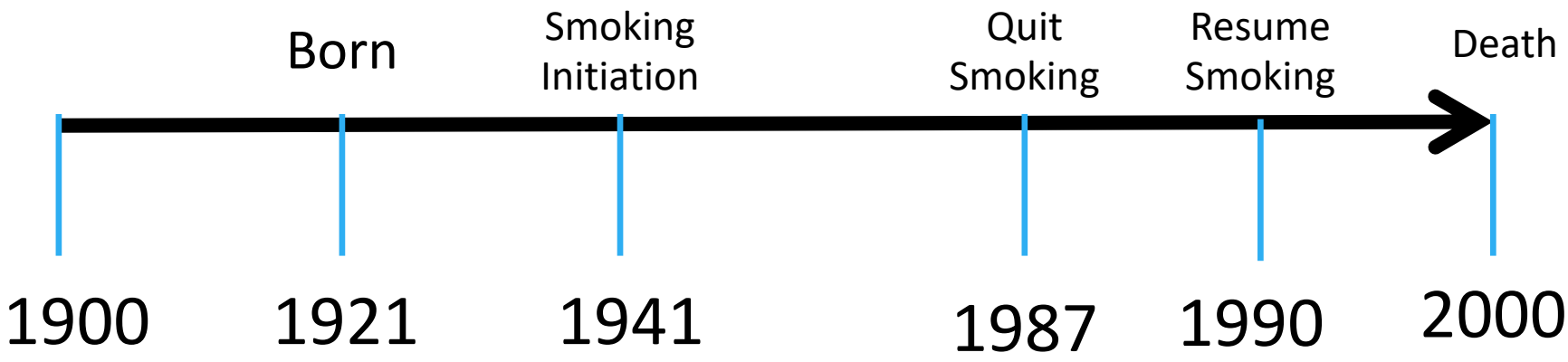
Time-varying variables

Exposure status can vary within individuals according to their actual levels of exposures



Measuring Exposure

- Not always an easy or straightforward task
  - Especially for time-varying exposures with cumulative effects



Exposure Metrics

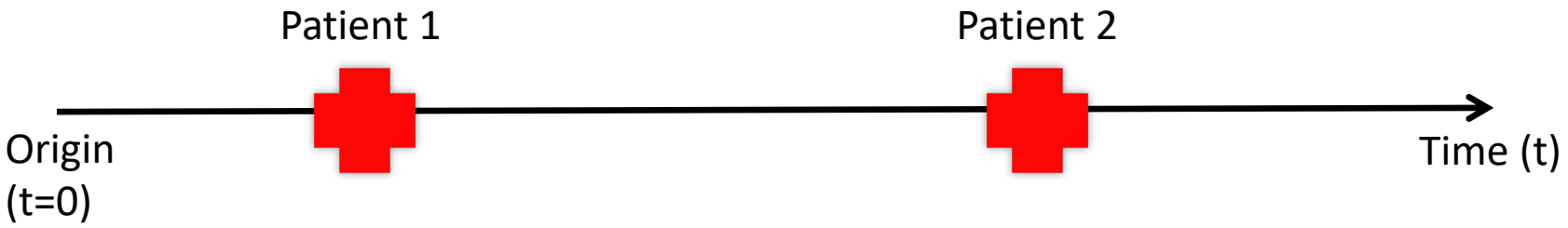
- Current value
- Threshold reached (ever/never)
- Cumulative amount
- Average amount (packs/day)
- Time since initiation
- Exposure during specific time window
- Lagged values (to allow for latency period)

Measuring Outcomes

- Different types:
  1. Non-repeatable events (e.g., birth, death, first MI)
  2. Repeatable events (e.g., influenza, pregnancy)
  3. Percent change/level of biomarker (e.g., CD4 count, cortisol levels).

Time-Scales

Often we are interested in not only whether an event occurs, but also when it occurs



Time Scales and Origins

Time Scale	Origin
Age	Birthdate (or year)
Calendar time	Calendar date (or year)
Time in study	Start of follow-up
Time at risk	Start of relevant exposure (e.g., smoking, pregnancy)

Time scale and origin chosen depend on the objective of study

Analyzing Cohort Data

Effect Estimates

From a cohort study we can calculate:

- Risk Ratios
- Risk Differences
- Rate Ratios
- Rate Differences
- Odds Ratios

		THEN FOLLOW TO SEE WHETHER	
		CHD Develops	CHD Does Not Develop
First Select {	Smoke cigarettes	84	2,916
	Do not smoke cigarettes	87	4,913

		THEN FOLLOW TO SEE WHETHER			
		CHD Develops	CHD Does Not Develop	Totals	Incidence per 1,000 per Year
First Select	Smoke cigarettes	84	2,916	3,000	28.0
	Do not smoke cigarettes	87	4,913	5,000	17.4

Risk in the exposed group= 84/3000 = 0.028 (28.0 per 1000)

Risk in the unexposed group = 87/5000= 0.0174 (17.4 per 1000)

What is the risk ratio? = 0.028-0.0174= 1.61

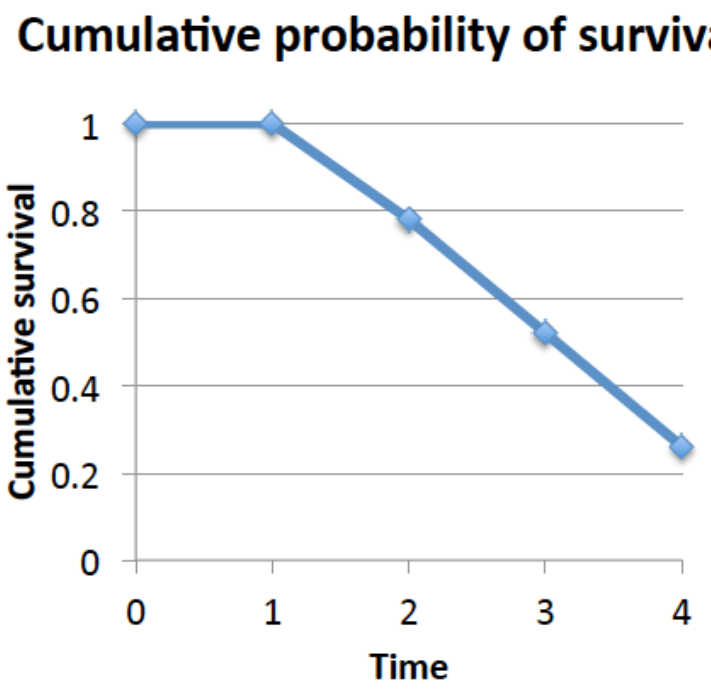
What is the risk difference?= 0.028-0.0174= 0.0106

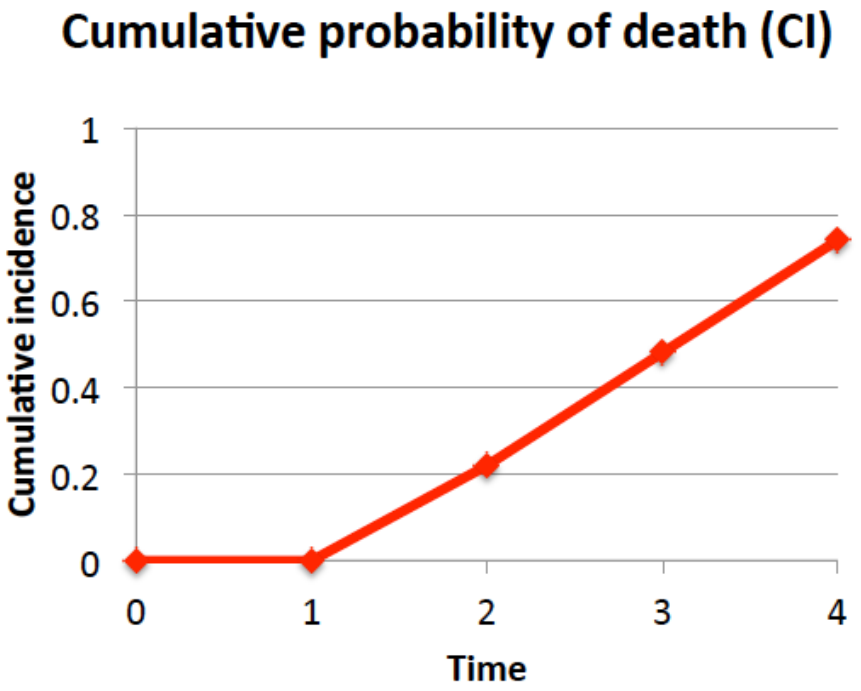
What is the odds ratio? = (84*4913)/(87*2916)= 1.63

Assuming **zero** losses to follow-up and drop out, what would the rate ratio and rate difference be?

Would be equal to the risk ratio and risk difference because total number of individuals = person time contribution if zero losses or drop-outs

Survival & Mortality





Survival is the complement of mortality

P= 1-q

Survival Analysis

- Want to compare average disease risk in exposed and unexposed groups but there are losses to follow-up
- When follow-up is long, need to account for losses to follow-up (e.g., withdrawals from study)
  - Moving away
  - No longer want to participate
  - Not able to participate any longer- proxies?
- Survival analysis is a set of analytic techniques that explicitly accounts for losses to follow-up

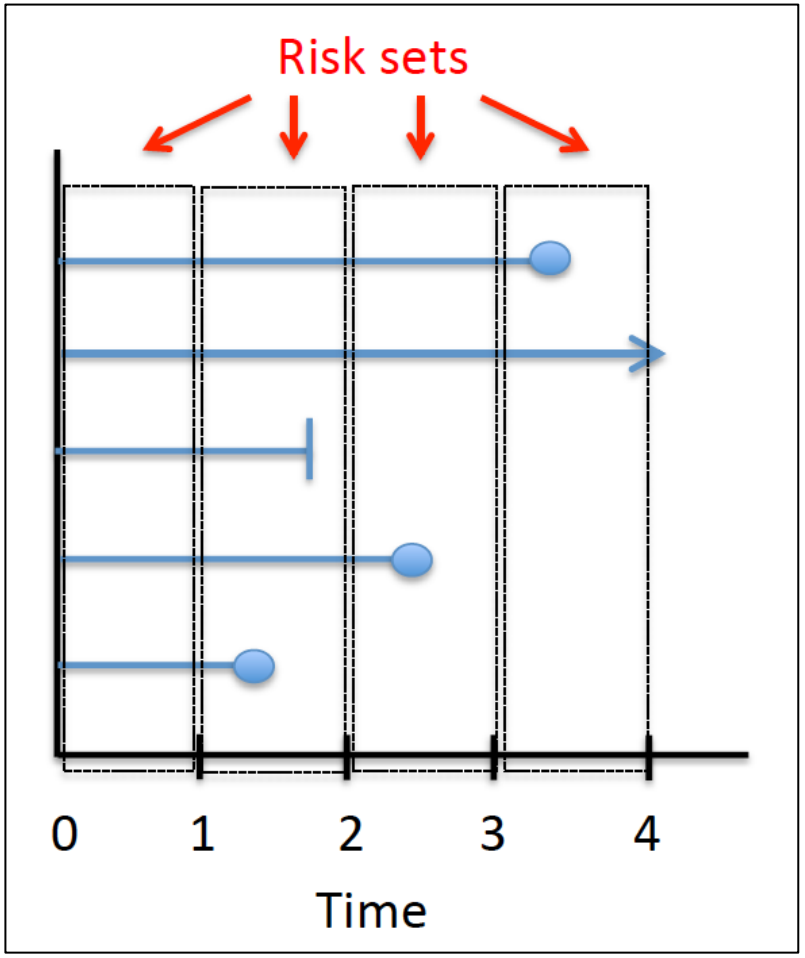
Kaplan Meier (KM) Method

- See text p. 134
- Method for examining survival over time in a cohort
- Rather than using pre-determined intervals (1 yr, 5 yr etc.) as with lifetables, calculate survival probabilities each time a death occurs
- If exact death times are available, KM makes the best use of available data

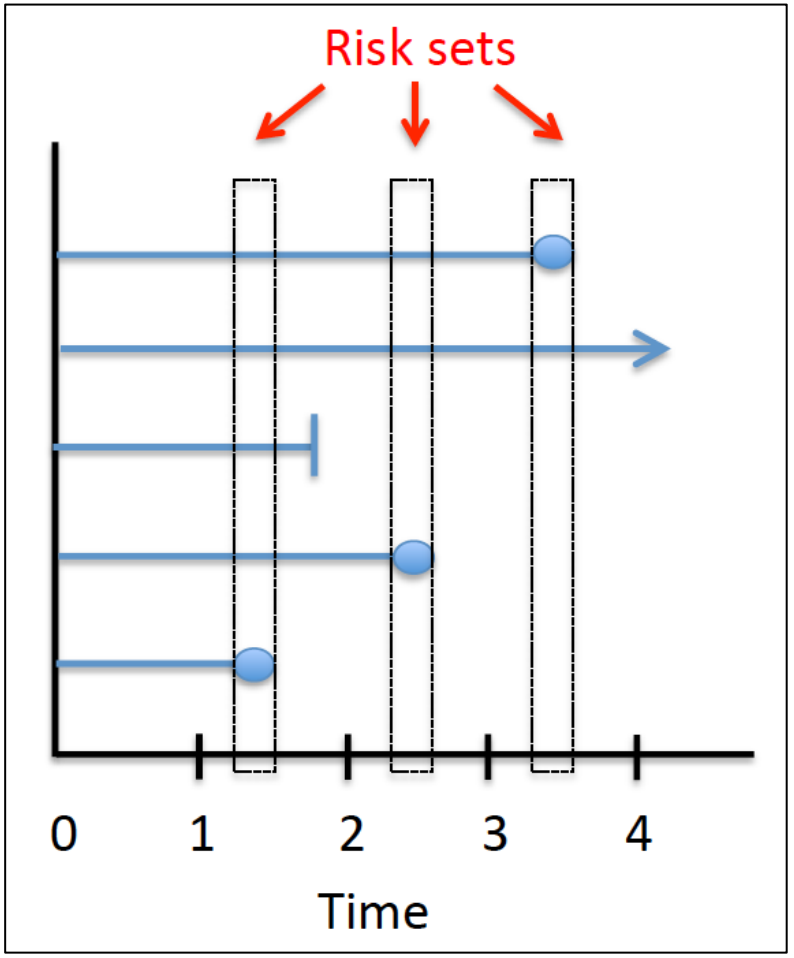
Kaplan-Meier

- Uses exact times that events occurs rather than time intervals
- Identify the exact point in time when each death occurred so that each death terminates the previous interval and a new interval (= new row in KM table) is started
- The number of persons who died at that point is used as the numerator, and the number alive up to that point (including those who died at that time point) is used as the denominator,
- Any withdrawals that occurred before that point are subtracted.

Life Tables



Kaplan Meier



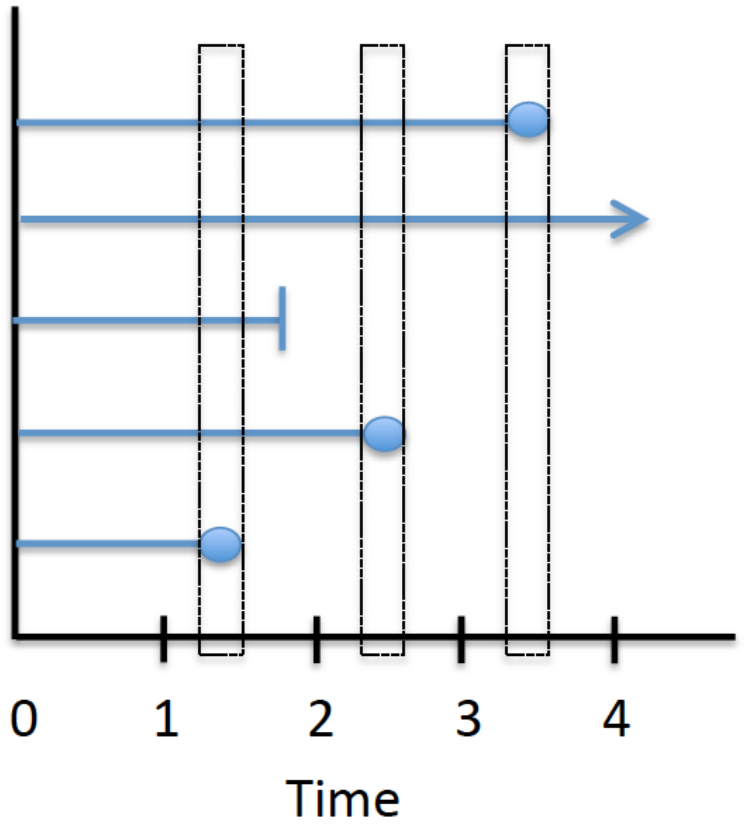
General Idea

1. Split up observation time into risk sets
2. For each risk set, calculate probability of survival among only those individuals in the risk set

P= conditional survival probability

3. Then at each time, t:
   - Estimate cumulative survival, $CS(t) = P_1 \times P_2 \times P_3 \ldots \times P_t$
   - Cumulative incidence, CI(t)

KM Example

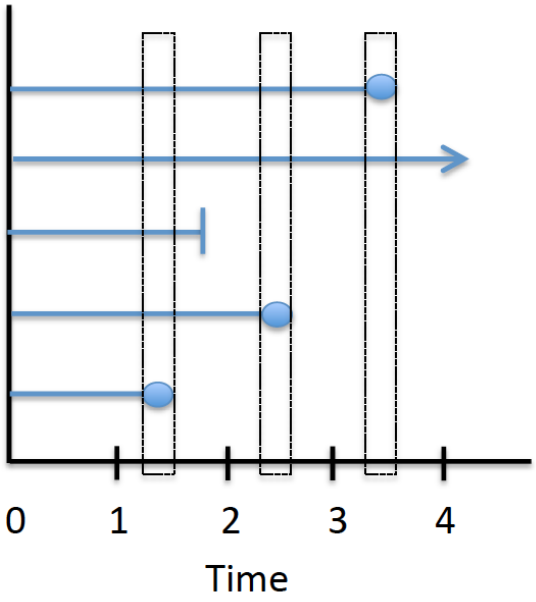


Probability of survival at event 1 (P_1):
- Probability of death = 1/5 = 0.2
- Probability of survival = 1-0.2 = 0.8

Probability of survival at event 2 (P_2):
- Probability of death = 1/3 = 0.33
- Probability of survival = 1-0.33 = 0.67

Probability of survival at event 3 (P_3):
- Probability of death = 1/2 = 0.50
- Probability of survival = 1-0.50 = 0.50

KM Example



Cumulative probability of survival:

CS(0) = 1.0

CS(event 1) = P_1 = 0.8

CS(event 2) = P_1 x P_2 = 0.8 x 0.67 = 0.54

CS(event 3) = P_1 x P_2 x P_3 = 0.8 x 0.67 x 0.50 = 0.27

Cumulative probability of death:

CI(0) = 1 – CS(0) = 1 – 1 = 0

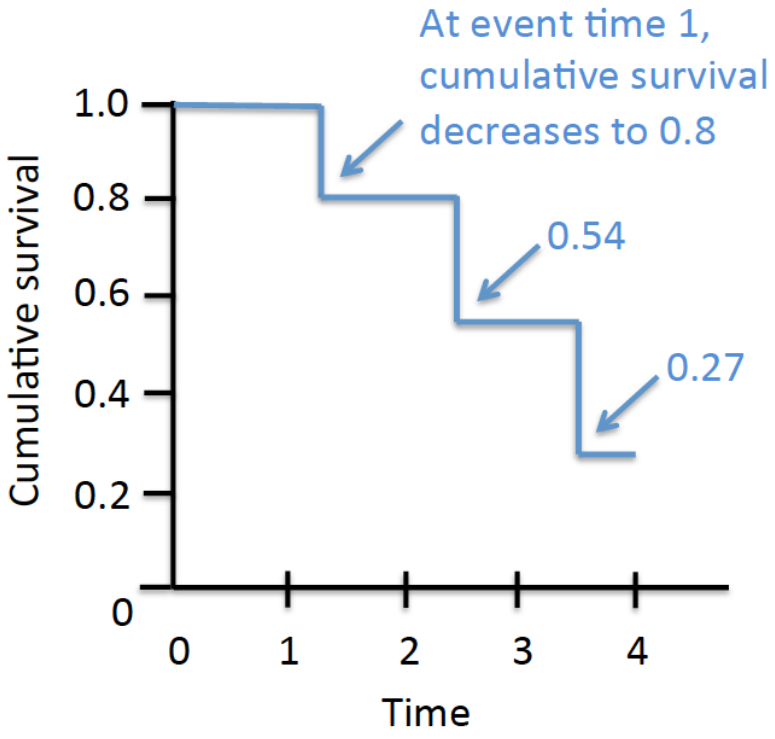
CI(event 1) = 1 – CS(event 1) = 1 – 0.8 = 0.2

CI(event 2) = 1 – CS(event 2) = 1 – 0.54 = 0.46

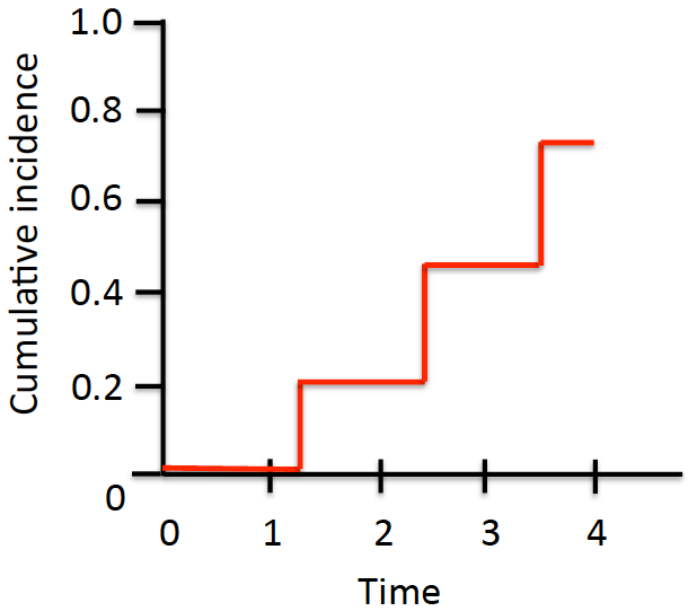
CI(event 3) = 1 – CS(event 3) = 1 – 0.27 = 0.73

KM Graphs

Cumulative probability of survival



Cumulative probability of death (CI)



KM Textbook Example (p.134)

- Six patients were studied, of whom four died and two were lost to follow-up ("withdrawals").
- The deaths occurred at 4, 10, 14, and 24 months after enrollment in the study.

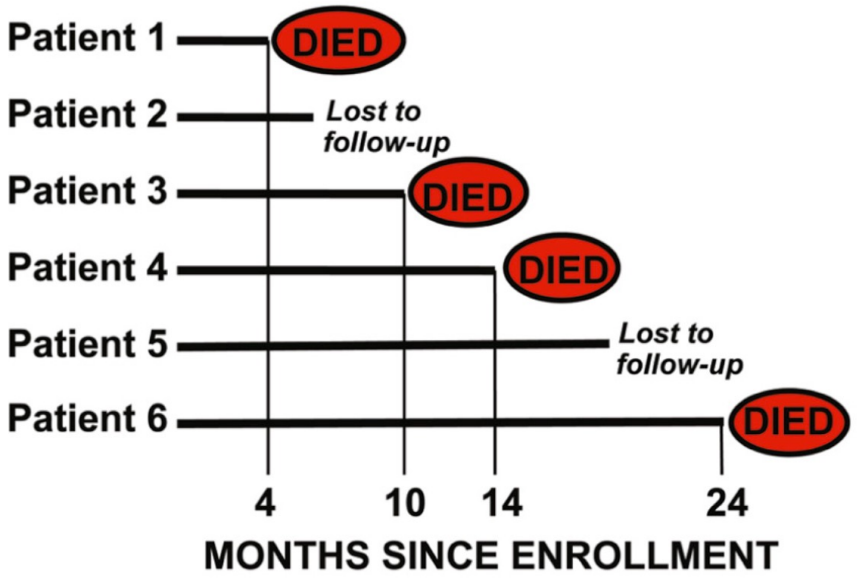


Fig. 6.12

Hypothetical example of a study of six patients analyzed by the Kaplan-Meier method.

Calculating survival via KM

(1) Times to Deaths From Starting Treatment (Months)	(2) No. Alive at Each Time	(3) No. Who Died at Each Time	(4) Proportion Who Died at That Time: $\frac{\text{Col (3)}}{\text{Col (2)}}$	(5) Proportion Who Survived at That Time: 1 – Col (4) (p)	(6) Cumulative Proportion Who Survived to That Time: Cumulative Survival
4	6	1	0.167	0.833	0.833
10	4	1	0.250	0.750	0.625
14	3	1	0.333	0.667	0.417
24	1	1	1.000	0.000	0.000

Cumulative survival proportion

Row 1: 1-0.167= 0.833

Row 2: 0.833*0.750= p1*p2=

Row 3: 0.625*0.667= p1*p2*p3=

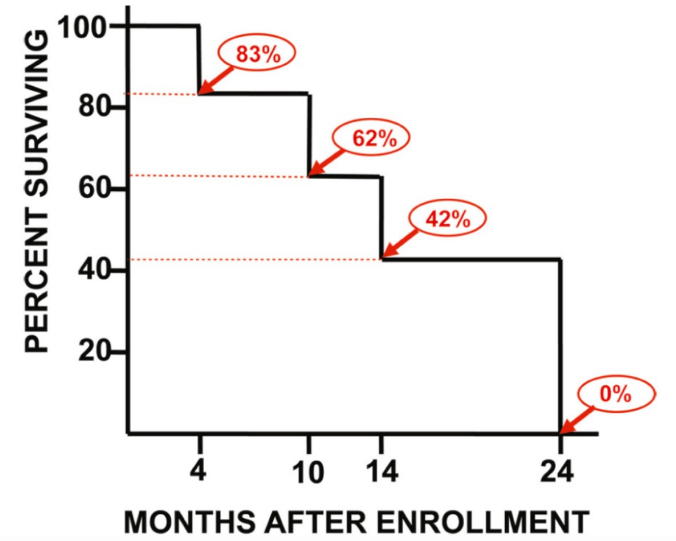
Row 4: 0.417*0.000= p1*p2*p3*p4=

KM Function

Graph of the cumulative survival probability

- Stepwise function
- After the drop in survival that accompanies each death, survival remains constant until the next death occurs

(6) Cumulative Proportion Who Survived to That Time: Cumulative Survival
0.833
0.625
0.417
0.000



Survival Analysis

- Target of inference: amount of time from an origin to the event
- Sometimes called time-to-event analysis
- Requires investigators to choose a time-axis and origin (t=0) as a common start to follow-up for all subjects
- Assumes that given enough time (and no competing risks), the event of interest will eventually occur

Hazard ratios

- Measure of association used in time to event analyses
- Hazard= Represents the instantaneous incidence rate at time t conditional on not having experienced the event yet
- Hazard ratio= ratio of two hazards (e.g., hazard in exposed group/hazard in unexposed group)
- In contrast to KM, the hazard rate cannot be calculated by hand, because it is defined for an infinitely small time interval, but the hazard function over time can be estimated using statistical modeling techniques