COHORT STUDY
Advance Designs , Planning and Execution,

Dr SM Namayandeh
MD, PhD
Epidemiologist
Epid.ssu@gmail.com

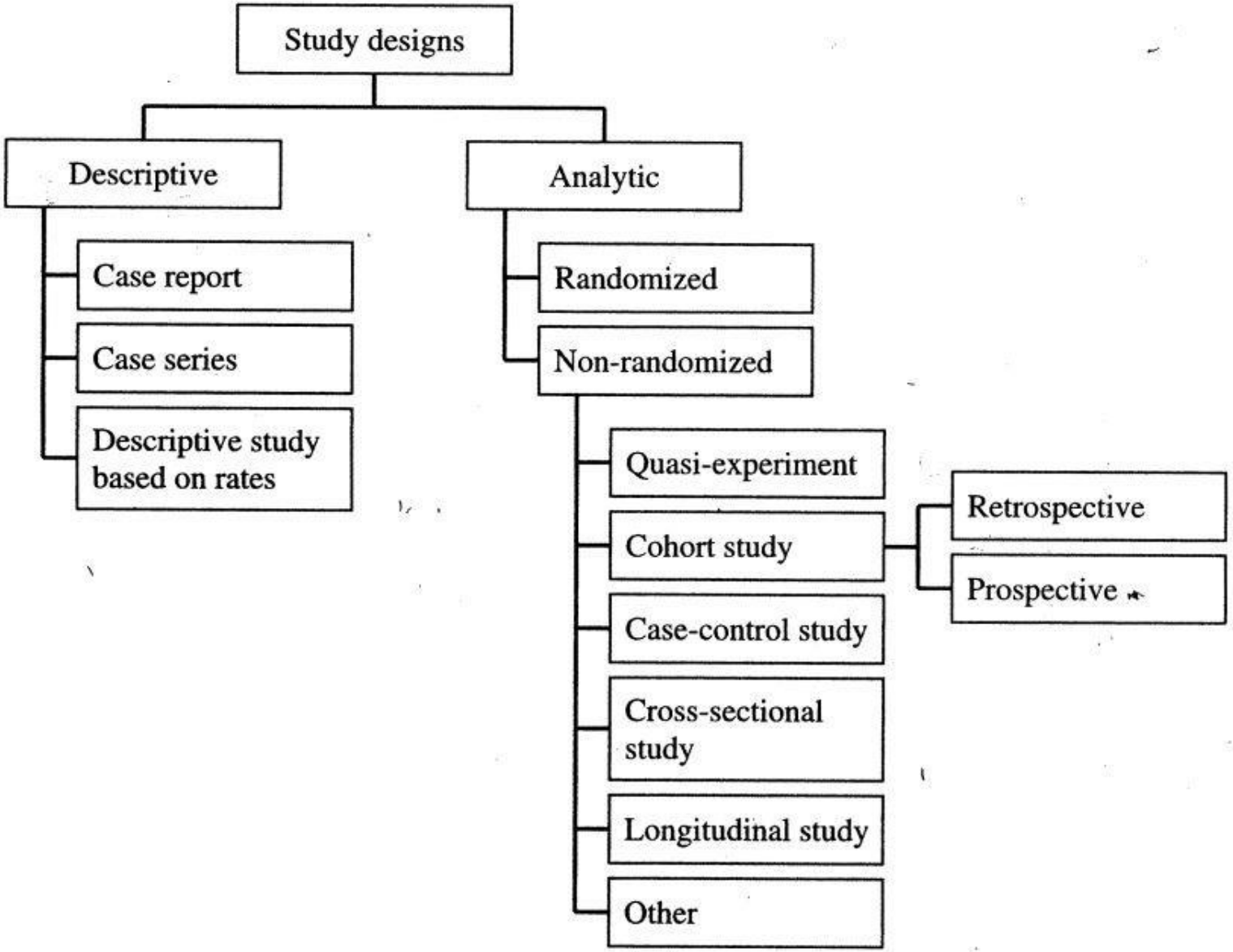
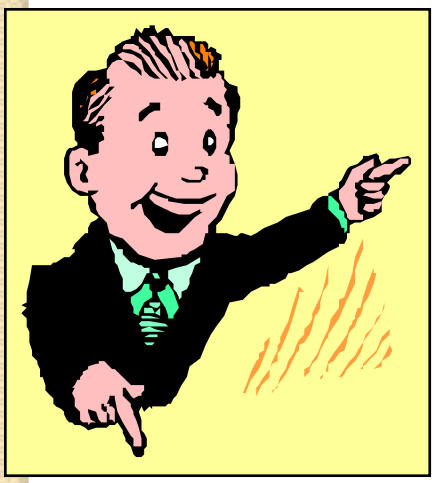


Figure 5–1. Major Epidemiologic Study Designs.

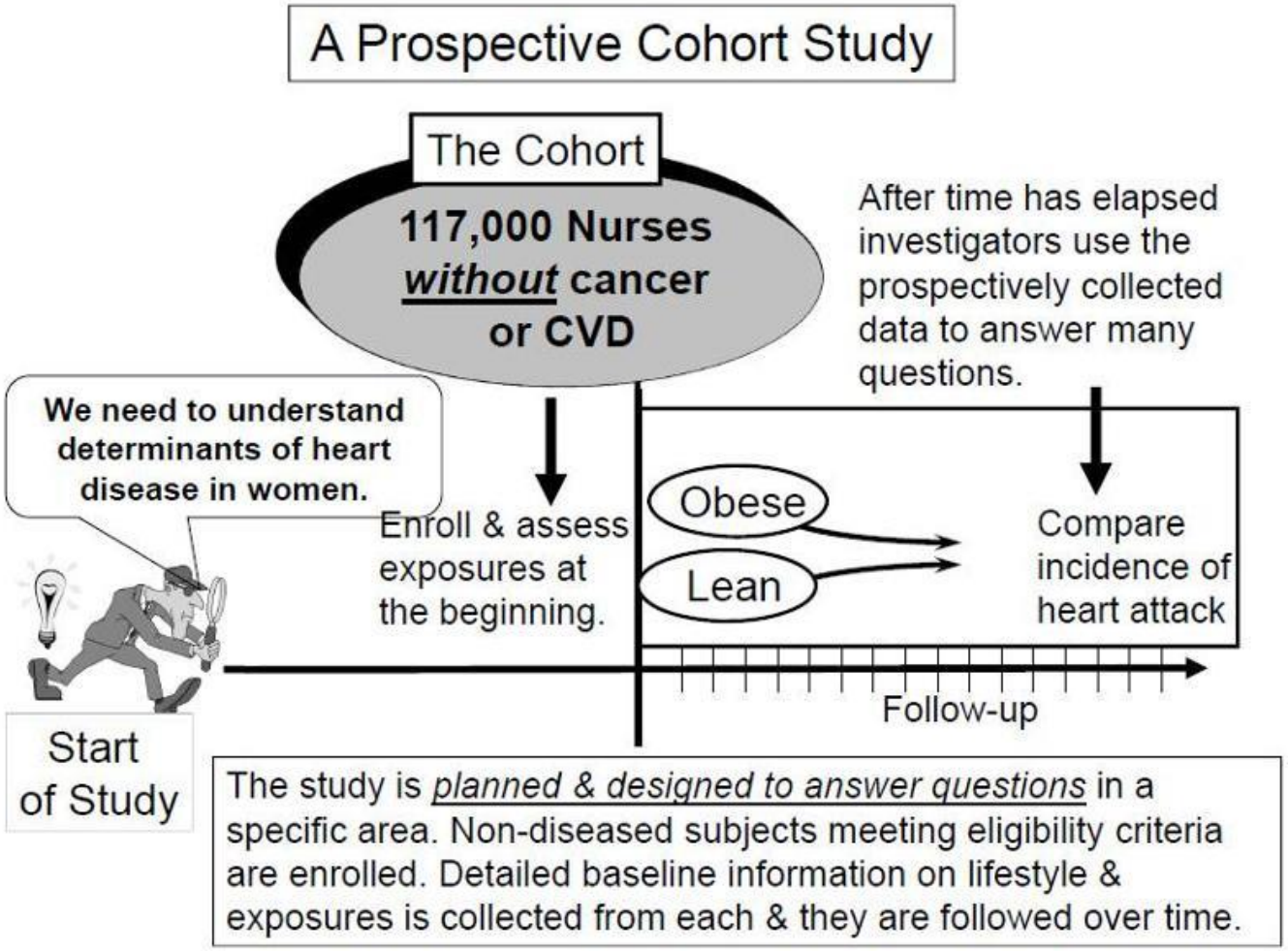
A Classification of Marketing Research Designs



- Research Design
  - Exploratory Research Design
  - Conclusive Research Design
    - Descriptive Research
      - Cross-Sectional Design
        - Single Cross-Sectional Design
        - Multiple Cross-Sectional Design
      - Longitudinal Design
    - Causal Research

Prospective cohort

- In a prospective study like the Nurses Health Study baseline information is collected from all subjects in the same way using exactly the same questions and data collection methods for all subjects.
- The investigators design the questions and data collection procedures carefully in order to obtain accurate information about exposures ***before*** disease develops in any of the subjects.
- After baseline information is collected, subjects in a prospective cohort study are then followed "longitudinally," i.e. over a period of time, usually for years, to determine if and when they become diseased and whether their exposure status changes.
- In this way, investigators can eventually use the data to answer many questions about the associations between "risk factors" and disease outcomes.


A Prospective Cohort Study
The Cohort
117,000 Nurses
without cancer
or CVD
After time has elapsed investigators use the prospectively collected data to answer many questions.
We need to understand determinants of heart disease in women.
Enroll & assess exposures at the beginning.
Obese
Lean
Compare incidence of heart attack
Follow-up
Start of Study
The study is planned & designed to answer questions in a specific area. Non-diseased subjects meeting eligibility criteria are enrolled. Detailed baseline information on lifestyle & exposures is collected from each & they are followed over time.

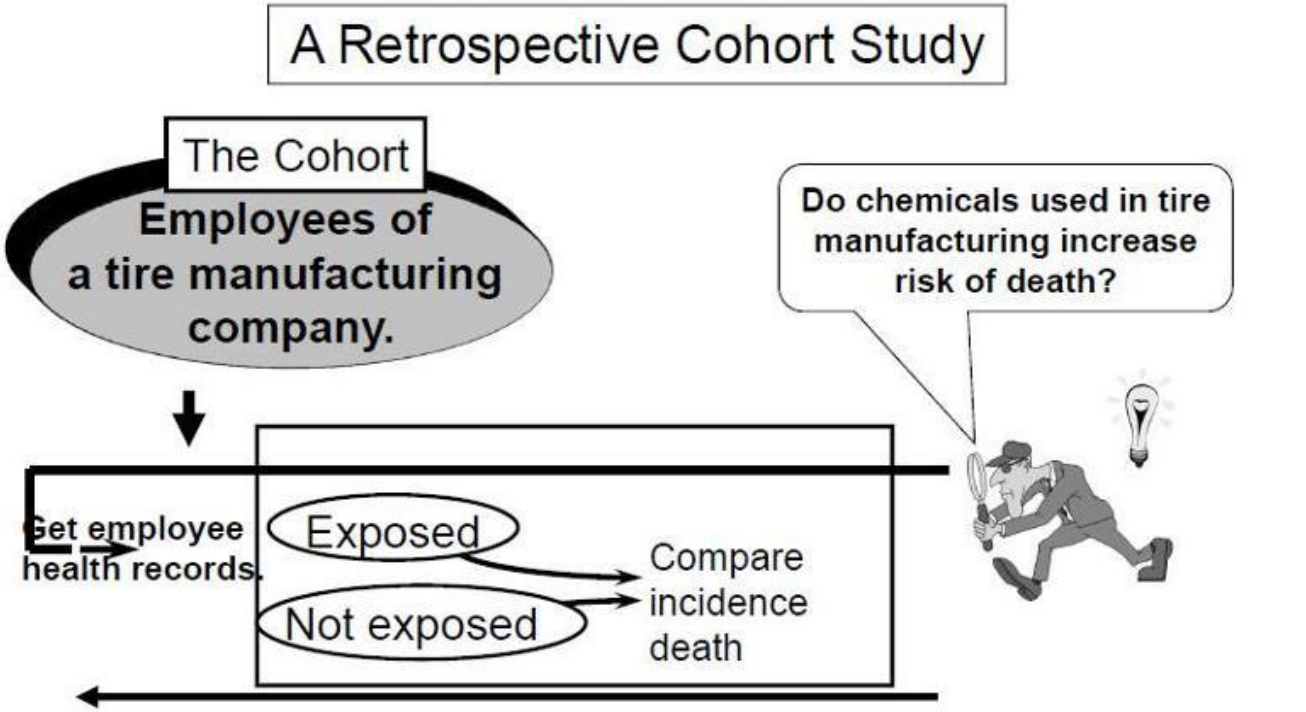
Key Concept:

- The distinguishing feature of a prospective cohort study is that at the time that the investigators begin enrolling subjects and collecting baseline exposure information, none of the subjects has developed any of the outcomes of interest.

Pitfall

- Note that in these prospective cohort studies a comparison of incidence between the groups can only take place after enough time has elapsed so that some subjects developed the outcomes of interest.
- Since the *data analysis* occurs after some outcomes have occurred, some students mistakenly would call this a retrospective study, but this is incorrect.
- The analysis always occurs after a certain number of events have taken place.
- The characteristic that distinguishes a study as prospective is that the subjects were enrolled, and baseline data was collected ***before*** any subjects developed an outcome of interest.

A Retrospective Cohort Study



This study was not preplanned. The investigator has to go back to pre-existing data that was not necessarily acquired in a precise, predetermined way. Follow up may have been incomplete.

Past

Start of Study

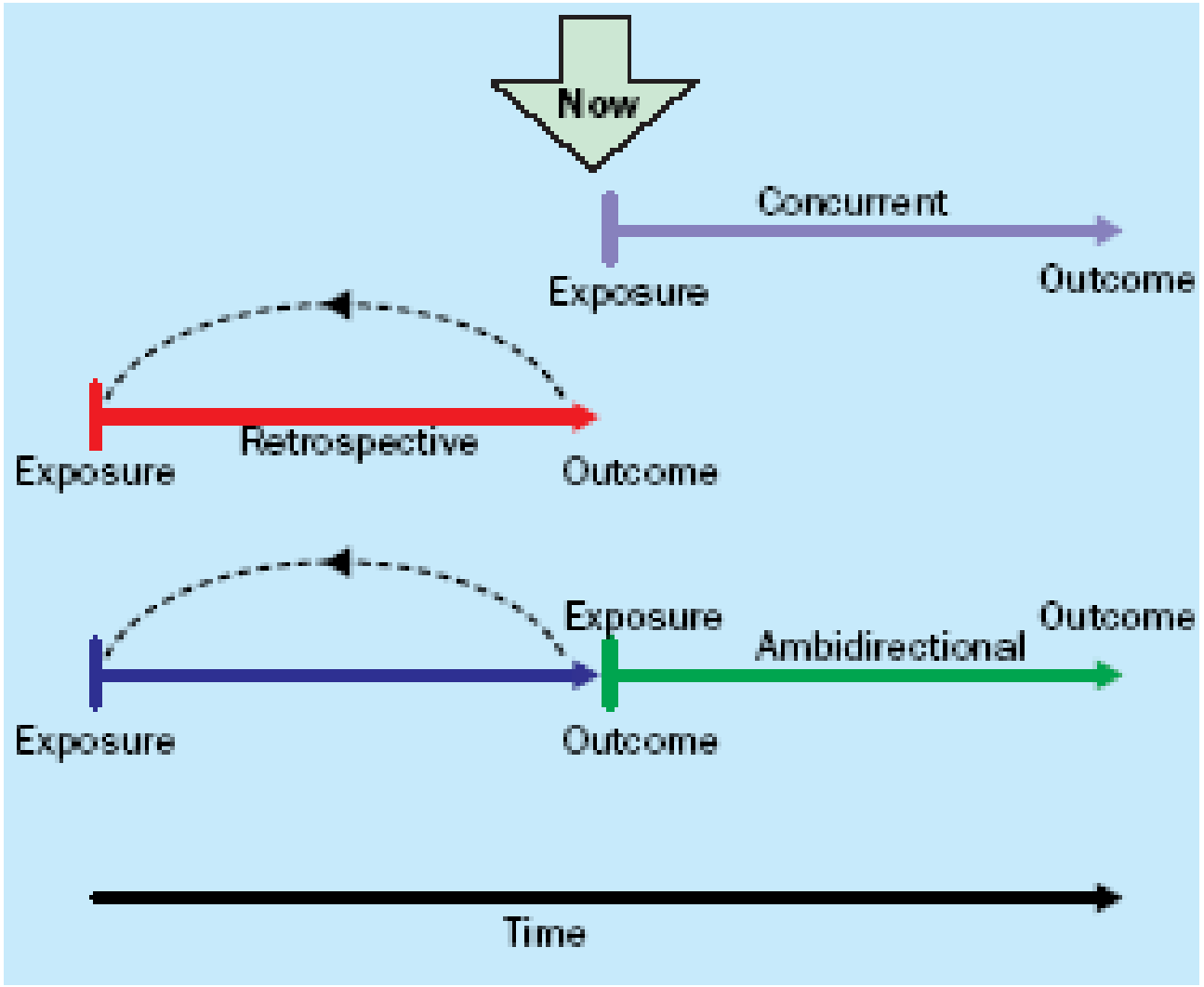
Key Concept:

- The distinguishing feature of a retrospective cohort study is that the investigators conceive the study and begin identifying and enrolling subjects ***after outcomes have already occurred***.

Key Concept:

- Common features of both prospective and retrospective cohort studies.

- None of the subjects have the outcome of interest at the beginning of the follow-up period. (In retrospective cohort studies the follow-up period begins in the past.)
- The groups being compared differ in their exposure status.
- One measures and compares the incidence of the outcome in order to determine whether there is an association between the exposure and the outcome.


Now
Concurrent
Exposure
Outcome
Retrospective
Exposure
Outcome
Exposure
Ambidirectional
Outcome
Exposure
Outcome
Time

Comparison (Control) Groups for Cohort Studies

1. **Internal controls**

   With a one-sample (population-based) cohort, exposure is unknown until after the first period of observation

   Example:

   a. Select the cohort (such as all residents of a given neighborhood)
   b. All members of the cohort are then given first round questionnaires, and/or clinical examinations, and/or testing to determine exposure
   c. The cohort is then divided into exposure categories based on those results

Comparison (Control) Groups for Cohort Studies (cont.)

2. **External controls**

- If everyone in a cohort is exposed (such as workers in an industry), a separate cohort as similar as possible to the exposed in terms of income, education, geography, and age should be sought

Example:

Workers in a neighboring but unexposed industry

Comparison (Control) Groups for Cohort Studies (cont.)

3. Known population rates

- If a comparison group cannot be assembled, known population rates of outcomes may be acceptable under some circumstances, if they are adjusted for the variables of interest
- **For lung cancer, however, rates are based on the population and are not adjusted for smoking**
- They are not, therefore, instructive to compare to populations with high smoking rates, such as miners

Outcome Definition

- Primary outcome - the main event that will be related to the exposure
  - Failure-time outcomes
    - Death
    - Disease occurrence
  - Repeated measures
- Secondary outcomes - other events that are of interest and may corroborate the findings of the main outcome

Cohort Study Design

Types of Cohorts

- Fixed Cohort
  - A group of individuals recruited and enrolled at a uniform point in the natural history of a disease or by some defining event
  - Cohort does not take on new members after it is assembled
  - Examples
    - Patients admitted to the ER with acute MI
    - Survivors of Hiroshima bombings
    - Children born to HIV-infected mothers

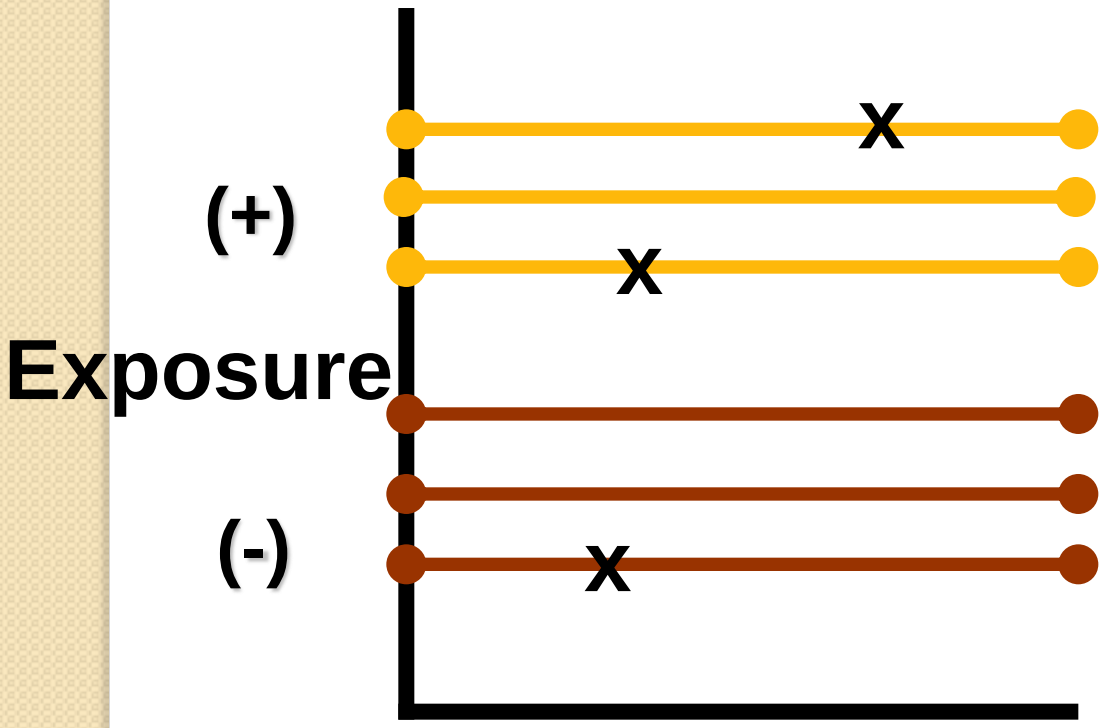
Cohort Study Design

Type of Cohorts

- Open cohort
  - A group of individuals recruited and enrolled through a mechanism that allows for in and out migration of people
  - Defined by characteristic other than disease, e.g., geographic location, administrative unit
  - Dynamic population
  - Examples
    - Framingham Study

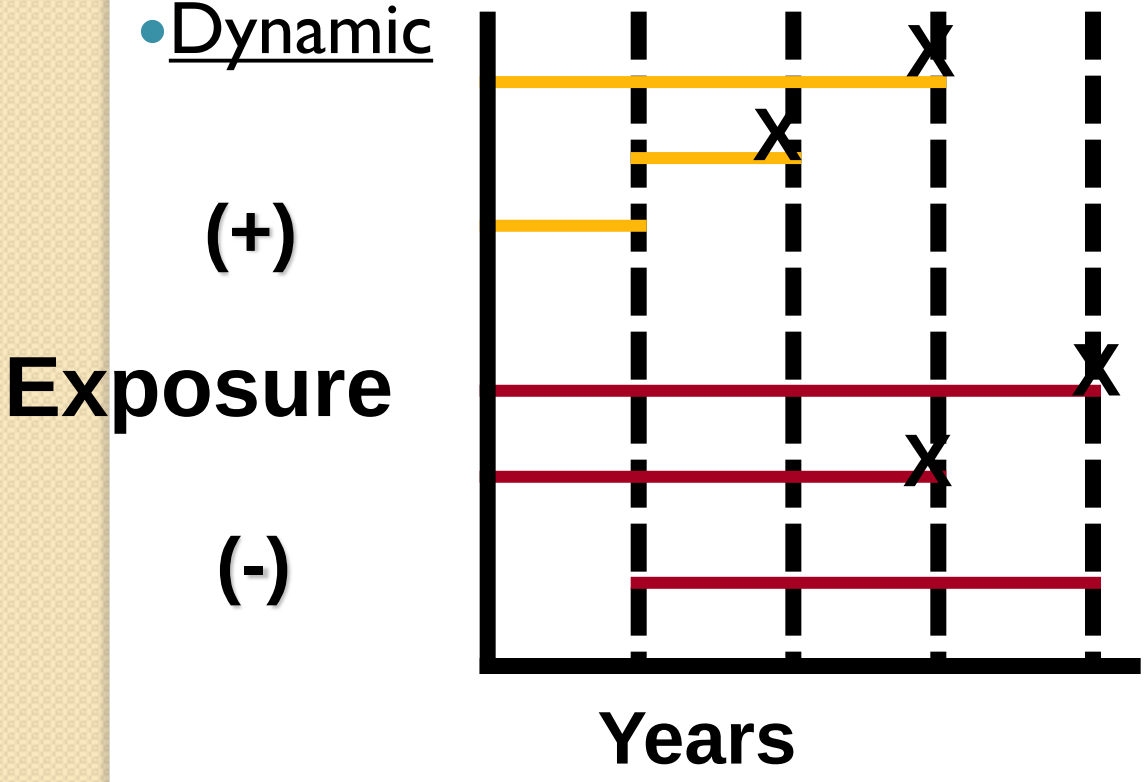
Cohort studies

- Fixed Cohort

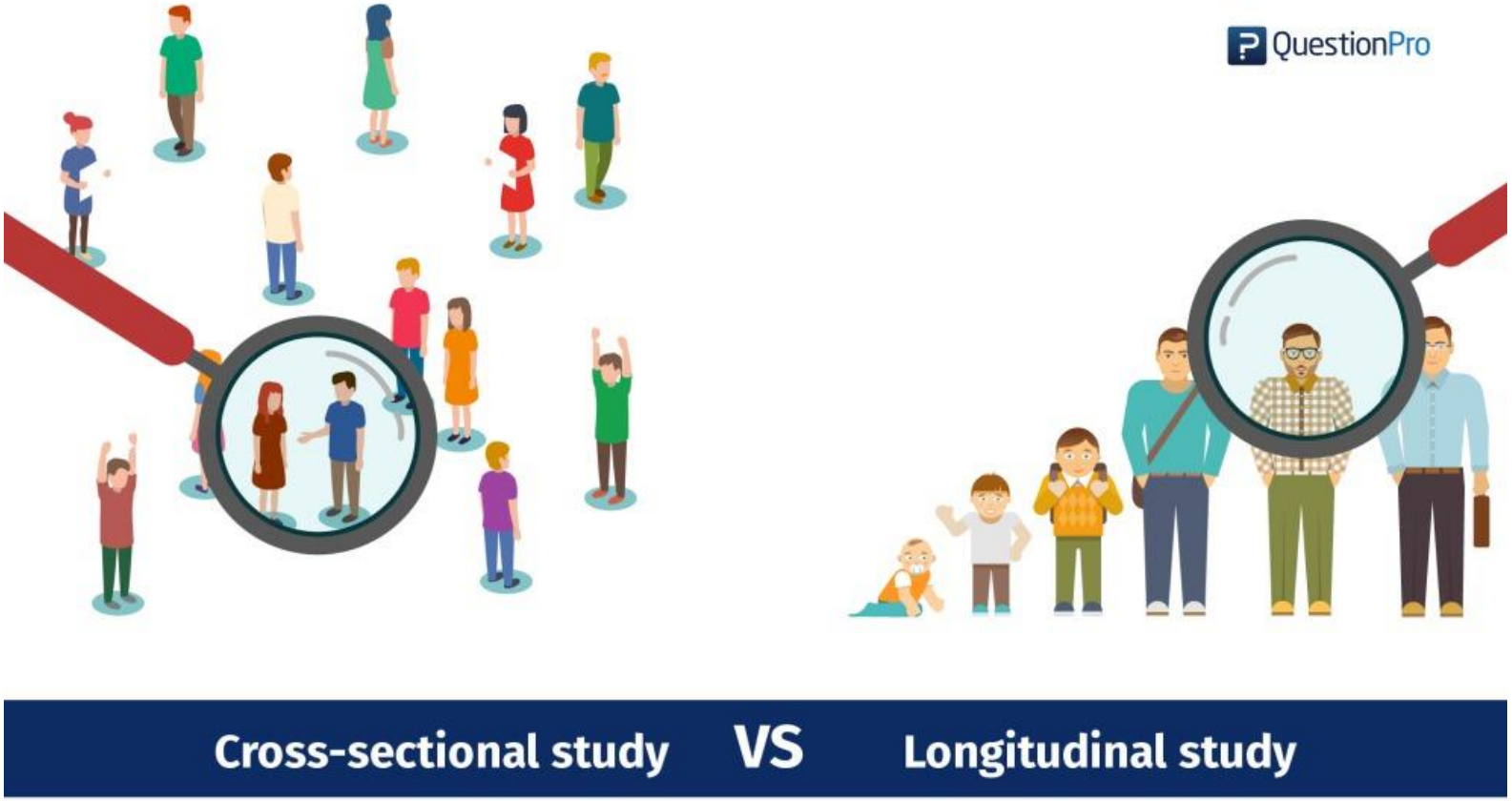


Cohort studies

- Dynamic



Analysing change over time: repeated cross sectional and longitudinal survey data



Cross-sectional Designs

- Involve the collection of information from any given sample of population elements only once.
- In **single cross-sectional designs,** there is only one sample of respondents and information is obtained from this sample only once.
- In **multiple cross-sectional designs**, there are two or more samples of respondents, and information from each sample is obtained only once. Often, information from different samples is obtained at different times.
- **Cohort analysis** consists of a series of surveys conducted at appropriate time intervals, where the cohort serves as the basic unit of analysis. A cohort is a group of respondents who experience the same event within the same time interval.

Inferences in two types of cohort study:

- Inferences from "life table-type“ cohort study types of cohort studies are restricted to population average effects.
- Longitudinal cohort studies, on the other hand, take explicit advantage of the repeated measures characteristics possible in cohort designs and make possible not only inferences on population average effects but on individual heterogeneity, changes in processes over time, and transitions between states of health and disease.

Longitudinal Designs

- A fixed sample (or samples) of population elements is measured repeatedly on the same variables
- A longitudinal design differs from a cross-sectional design in that the sample or samples remain the same over time

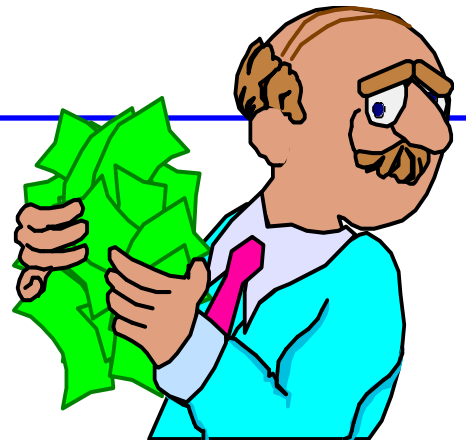
Relative Advantages and Disadvantages of Longitudinal and Cross-Sectional Designs

Evaluation Criteria	Cross-Sectional Design	Longitudinal Design
Detecting Change	-	+
Large amount of data collection	-	+
Accuracy	-	+
Representative Sampling	+	-
Response bias	+	-

Note: A "+" indicates a relative advantage over the other design, whereas a "-" indicates a relative disadvantage.

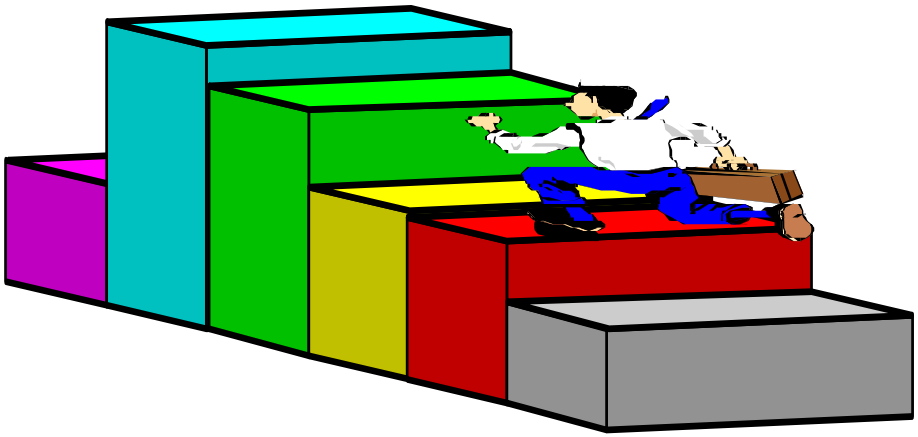
Cross-Sectional Data May Not Show Change

	Time Period	Brand Purchased
	Period 2 Survey	Period 1 Survey
Brand A	200	200
Brand B	300	300
Brand C	500	500
Total	1000	1000



Longitudinal Data May Show Substantial Change

Brand Purchased in Period 1	Brand Purchased in Period 2		
	Brand A	Brand B	Brand C Total
Brand A	100	50	50	200
Brand B	25	100	175	300
Brand C	75	150	275	500
Total	200	300	500	1000



Two broad cohort study categories:

- Tager (1998) has divide cohort studies into two broad categories :

1. "***life table-type***"
2. "***longitudinal***."

"life table-type" cohort study:

- The "***life table-type***" are characterized by their treatment of time and exposure in a manner that is closely tied to ***traditional life table*** methods of analysis.

- In general, *exposure* and *person-time* are ***summarized***, and ***incidence density***, ***cumulative incidence*** of a discrete disease outcome, and their respective ratios are the principal outcomes of the life table-type cohort study.

Definition of longitudinal cohort study:

- A ***fixed sample*** (or samples) of population elements is measured ***repeatedly*** on the ***same variables.***
- Sample or samples remain the ***same*** over time.

Inferential and analytic focus:

1. Estimation of effects at the ***individual level*** .

2. Estimation of the ***effects of changes in risk markers*** over time on disease (health state) outcome; average ***population change in risk with change in exposure***, also at ***individual*** heterogeneity.

3. ***Estimation of rate of change*** or change in level with time for some ***outcome*** that relates to ***disease natural history*** that can use for clear understanding of the ***behavior of biomarkers*** for a disease or state of health over time.

4. ***Natural history*** of states of health that either have ***multiple occurrences*** or can ***oscillate*** between different states.

5. ***Separation of cohort (or period) effects*** from the effects of age or chronologic calendar time.

Conclusion:

- Longitudinal cohort designs clearly offer the opportunity to ***explore*** a ***wider range of processes and outcomes*** that are of major interest to epidemiologists.

- They offer the means to explore the ***natural history*** of the ***biologic, social, and environmental processes*** that are important ***determinants*** of disease occurrence and preservation of health in a manner that is *not possible* with *life table cohort-type designs* or *case-control studies*.

- Moreover, ***advances*** in ***statistical methodology*** for longitudinal data , and the wide ***availability*** of ***software*** to implement these methods, make it ***feasible*** to ask a *wider range and complexity of questions* in the context of epidemiologic cohort studies than would have been feasible in the past.

Planning and Execution,

- **Selection of study groups**
- The aim of a cohort study is to select study participants who are identical with the exception of their exposure status. All study participants must be free of the outcome under investigation and have the potential to develop the outcome under investigation.
- **Measuring exposure**
- Levels of exposure (e.g. packs of cigarettes smoked per year) are measured for each individual at baseline at the beginning of the study and assessed at intervals during the period of follow-up. When several exposures are being considered simultaneously, the non-exposed group should comprise all those with none of the risk factors under investigation.
- A particular problem occurring in cohort studies is whether individuals in the control group are truly unexposed. For example, study participants may start smoking or they may fail to correctly recall past exposure. Similarly, those in the exposed group may change their behaviour in relation to the exposure such as diet, smoking or alcohol consumption.
- Exposure data may be obtained from a number of sources including medical or employment records, standardized questionnaires, interviews and by physical examination.
- **Measuring outcome**
- Outcome measures may be obtained from various sources, including routine surveillance of cancer registry data, death certificates, medical records or directly from the participant. Note that the method used to ascertain outcome must be identical for both exposed and unexposed groups.
- **Methods of follow-up**
- The follow-up of study participants in a cohort study is a major challenge. A great deal of cost and time is required to ensure follow-up of cohort members and to update measures of exposures and confounders, in addition to monitoring participants' health outcomes. The failure to collect outcome data for all members of the cohort will affect the validity of study results

How to run a cohort study

- If the data are readily available then a retrospective design is the quickest method.
- If high quality, reliable data are not available a prospective study will be required.
- The first step is the definition of the sample group. Each subject must have the potential to develop the outcome .
- Furthermore, the sample population must be representative of the general population if the study is primarily looking at the incidence and natural history of the condition (descriptive).
- If however the aim is to analyse the relation between predictor variables and outcomes (analytical) then the sample should contain as many patients likely to develop the outcome as possible, otherwise much time and expense will be spent collecting information of little value.

Estimates made from cohort studies

- Before discussing specific issues in study design, it will be useful to review briefly the estimates that the investigator usually wants to make from the data collected in a cohort study.
- Because, the nature of these estimates to a large extent dictates the information to be collected.
- It is useful to keep in mind from the outset the need to collect certain relevant information.
- For example, the investigator may want to estimate IR:

  - In just exposed group

  - In those exposed to various level and for various length of exposure

  - In those exposed to certain combination of exposure

Estimates made from cohort studies

- It is important to take any confounding variables into account.
- Determining whether exposure to one factor modifies the effect of exposure to another is also an important part of data analysis. Interaction of risk factors

Assembling the Cohort

- Before beginning a study, those who are susceptible and those who are immune or who for other reason are not at risk for the disease under study should be identified.
- However, such definitive separation into susceptible and immune cannot be done in certain situations like:

  - Diseases for which re-infection is common in the presence of antibody (Respiratory syncytial virus)

  - Diseases that result from reactivation of a latent infection(herpes viruses, TB, toxoplasmosis)

  - Diseases in which the height of Ab level is associated with it not only its presence (African Burkitt's lymphoma due to EBV)

Assembling the Cohort

- In non-infectious diseases definitive markers of susceptibility are not available.
- People with known or suspected history of the disease should be eliminated unless recurrences are of interest.
- Identifying those not at risk for a disease and those who have already had the disease may involve conducting a prevalence study at the beginning of a cohort study.
- The identification of persons with current past asymptomatic or subclinical disease is sometimes difficult.
- Various tests and procedures are useful, but in general these procedures must be simple, harmless, and inexpensive.

Assembling the Cohort

- In many situations, large sample sizes are needed for enough new cases of disease to develop.
- To avoid a large sample size, the cohort under study should be limited to a group at higher risk of the disease.
- If risk factors were identified from these selected populations, the next step is to try another cohort from a more general population.

Cohorts on special groups

- A commonly used cohort is people working in a particular industry, occupation or group:

  - They often have exposure of interest.

  - They are less likely to be lost to follow-up because of their lower mobility than the general population.

  - They have a certain amount of relevant information recorded in their medical and employment record.

  - In many instances, they undergo initial and then periodic medical examination.

- Lots of studies for identifying CHD risk factors have been done in special occupational groups.

Cohorts on special groups

- Prospective sero-epidemiologic studies of infectious diseases usually carry out in school, college, university and military populations.
- These studies contribute to the knowledge of:

  - Disease etiology and incidence

  - Ratio of inapparent to apparent infection

  - Biologic spectrum of clinical illness associated with a given infectious agent

  - Contribution of various agents to a given clinical syndrome

  - Level of antibody needed to protect against reinfection or recurrent clinical disease

  - Duration of immunity

Nurses' Health Study Cohort

- Nurses were selected for the cohort not because of any particular occupational exposure, but because it was believed that their cooperation would be at a high level and that they could report disease occurrence with a high degree of accuracy.
- The cohort was established in 1976.
- Validity and reliability of questionnaire showed a good quality of data.
- A questionnaire regarding all suspected risk factors of cancer and coronary heart diseases were sent to the participants.
- A follow-up questionnaire was mailed to the cohort members every 2 years to update data on major medical events and new area of interest.

Determining exposure status

- The techniques used to measure the possible risk factors of interest vary considerably from one study to another and from one risk factor to another.
- Exposures may be identified from existing records, as in studies aimed at identifying occupational exposures and medications as risk factors.
- In infectious diseases, the presence, duration, and intensity of exposure to an infectious agent depend on the source of infection and the means of transmission.
- When the source of infection and means of transmission are well defined and of one type, or are unique, then division into exposed and non-exposed group may be quite simple.

Determining exposure status

- The use of biologic markers in evaluation of exposure status of non-infectious diseases.
- Changes in exposure to putative risk factors during the course of follow-up often occur.
- Exposure levels may also change as a result of actions outside the control of either the study subjects or the investigator.
- Comparison group:
  - Internal
  - External
  - Combined
  - Self-selection

Measurement of Disease

- The procedures for disease identification should be comparable for exposed and unexposed.
- If possible, disease diagnosis should be done by persons unaware of the exposure status.
- People may not even know that they have a disease and may therefore have no reason to seek care.
- When the disease of interest needs hospitalization the best way is to monitor hospital records. (diversity in records or diagnostic criteria).
- The best way of ascertainment of infectious diseases is lab. Test.
- More active ascertainment must be employed for less severe infections for which medical care is not usually sought, such as many respiratory diseases, and intestinal diseases.

Measurement of Disease

- When measuring diseases with continuously contributed in the population a definite cut off level is needed.
- In measuring disease occurrence, seek for subclinical or asymptomatic conditions is very important.

Effects of nonparticipation

- Non-participants almost always differ in some way from participants.
- In most studies till 1991 a higher proportion of non-participants were smoker.
- In Framingham study a higher proportion of nonparticipants died within 5 years after the beginning of the study.
- The effect of nonparticipation on measures of association depends on both the size of the group omitted from the study and its specific characteristics.
- Also bias is likely to be greatest when the proportion of nonparticipants is high and when the participants differ greatly from nonparticipants in likelihood of developing the disease.

Effects of nonparticipation

- If the proportion participating differ according to likelihood of exposure but not disease, then the values of the rate ratio, risk ratio, and odds ratio are all unbiased.
- For example, if in a study of the association between smoking and CHD, a higher proportion of non-smokers participated than of smokers, but participants and nonparticipants were equally likely to develop CHD, then these measures of association would not be biased.
- If the proportions participating differ with respect to the disease but not exposure, then rate ratio and risk ratio are biased but the odds ratio is not.

Effects of nonparticipation

- In a prospective cohort study, it is more likely that the proportions participating will differ with respect to exposure than with respect to disease.
- The most serious impact of nonparticipation on study results occurs when the proportions participating not only vary with respect to exposure or disease but also vary according to specific combination of exposure and disease.
- For example, if persons who both have the exposure and develop the disease are less or more likely to participate in a study than is the remainder of the cohort, then the value of the rate ratio, risk ratio, and odds ratio can be quite different from the population values.

Open versus Closed Cohorts

- An open cohort – or dynamic cohort - is one where people can enter or leave
  - Examples: A workforce study that is ongoing
  - A city or other geographic location
- A closed cohort is where all persons in the cohort are defined at entry. No one enters, members can only exit.
  - Eg. McGill medical school class of 2004

Selection Bias

- Definition – selection bias occurs when there is a distortion in the estimate of effect (association) because the study or sample population is not truly representative of the underlying or source population in terms of the distribution of exposures and/or outcomes.
- Case control – detection or diagnosis or referral bias, and Berkson's bias
- Cohort studies – selection bias, healthy worker effect, drop-out bias

Avoiding Selection Bias – a representative sample

- In an un-biased sample we hope to have a representative sample as follows:

Truth – distribution of exposure and disease in source population		
	Exposed	Not Exposed
Diseased	A	B
Not Diseased	C	D

Odds Ratio = (A/B) / (C/D) •

$$= \frac{A \times D}{B \times C}$$

Example – Un-biased Sample

	Exposed	Not Exposed
Diseased	P_1A	P_2B
Not Diseased	P_3C	P_4D

- Odds Ratio = $\dfrac{(P_1 \times P_4)}{(P_2 \times P_3)} \times \dfrac{(A \times D)}{(B \times C)}$

IF $\dfrac{(P_1 \times P_4)}{(P_2 \times P_3)}$ ***THEN*** OR = $\dfrac{(A \times D)}{(B \times C)}$ = Truth!

To achieve Un-biased Sampling

- To achieve un-biased sampling the easiest is:
- $P_1 = P_2 = P_3 = P_4$
- This means the proportion sampled from each group is the same, i.e., 10% are sampled from each of the groups
- However if P_1 is higher than P_2 this can be okay as long as P_4 is also increased more than P_3

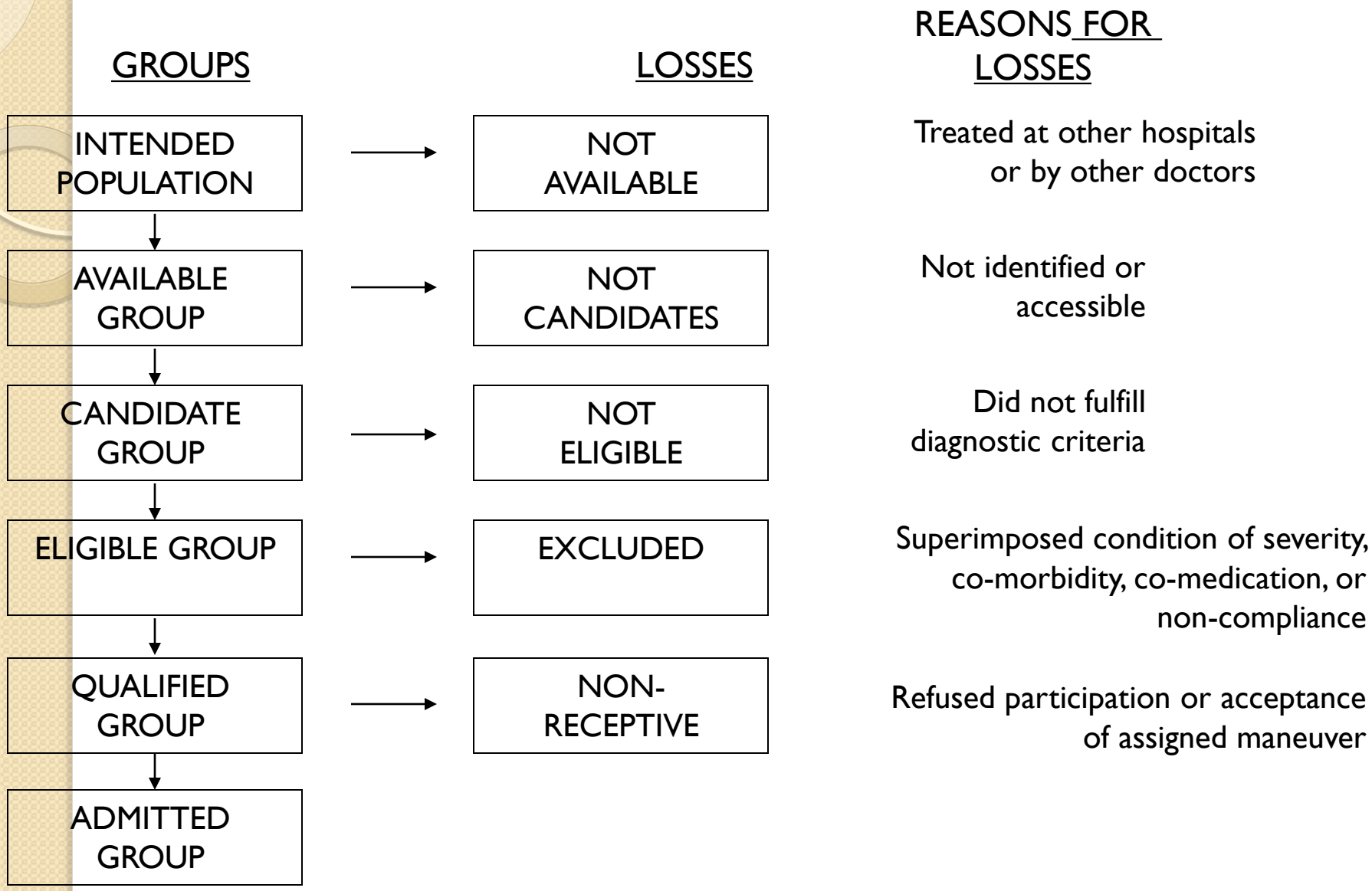


Figure 15-2. Diagram showing successive transfers from the intended population to the group admitted to a study of therapy

Volunteer Bias

- Another term for selection bias, when:
- Participants in a study are different from refuseniks
- Potential subjects who have the exposure and the outcome are more (or less) likely to participate
- Examples:
  - Fetal malformations and exposures.
  - Disease and occupational exposures, particularly if self-reported exposures.
  - (Both of these can also be affected by recall bias, because more likely to report possible exposures)
- What was the mortality of non-participants in the Framingham study, and why?

Susceptibility bias

- Just another term for selection bias
- Persons allocated to one form of treatment, or who who self-select to certain exposures are more, or less susceptible to develop health effects/ outcomes of interest.
  - Eg Cancer patients who have surgery vs medical or radiotherapy only. Surgical patients often appear to do better.

Healthy worker effect

- An important bias – found in work-force studies
  - Reflects medical screening (military, mining)
  - Or, physical requirements of job
- Results in better health status initially than general population, or certain control pop'n
  - Strongly affects results in cross-sectional studies
  - Reduces risk or delays occurrence of health outcomes of interest.
- Also occurs in smokers "healthy smoker effect"
  - Lung function in adolescent smokers > non-smokers

Selection Bias in Cohort Studies – Dropout's

- Losses to follow up occur in all cohort studies
- Generally will reduce power, and dilute results
- Particularly problematic if losses to follow up are greater in one of the exposure groups,
- REALLY important if due to development of disease

Drop-outs from a work-force - impact

- If a particular occupational exposure results in health effects quickly in a susceptible sub-group, and they then leave the work-force (quit) then this effect can be easily missed
  - In cross-sectional designs – none left
  - Even in cohorts – event rate appears low overall, because all outcomes of interest occur in small number of new workers (power problem)
- Example: Allergy to lab animals in researchers
  - Asthma in Grain workers
  - Latex allergy in health professionals

Selection Bias in Cohort Studies – Dropout's

- Example:
  - study of incidence of diabetes in obese persons.
  - Truth: IRR = 3.0
  - Losses – 33% in diabetes/obesity group (death/other)
    - 5% losses in all other groups
  - $\frac{(P_1 \times P_4)}{(P_2 \times P_3)}$ ***does not = 1***

Selection Bias from Dropout's in a Cohort Example

	At onset	Dropped Out		Detected at end with diabetes
		No DM	Diabetes	
Obese	227	10	9	18
Not Obese	773	35	3	30

- Incidence (biased):
  - In obese – 18/208 = 8.7%
  - In non-obese – 30/735 = 4.1%
- Biased incidence rate ratio – 8.7%/4.1% = 2.1

Controlling Selection Bias

- Most important strategy is prevention
  - Design strategies, particularly in case control
  - Recruitment – high % in all groups
  - Same recruitment in exposed/not exposed or cases/controls
  - In cohort studies close follow up to prevent dropouts
- Can assess impact in analysis
  - Comparing characteristics of dropouts with those who remained
  - Comparing those who participated with those who refused
  - Sensitivity analysis – best case/ worst case to assess impact of selection biases

Cohort Studies – Exposure Assessments

- Prospective - Measure exposures at outset
  - These can be one or many
  - Specific: cholesterol, obesity, smoking, blood pressure.
  - Proxies: occupation, housing
  - These can be measured repeatedly eg., every year or every six months, to account for changes in exposure over time (obesity, smoking, BP).
- Retrospective
  - Exposure is based upon past events
  - Usually exposures can not be directly quantified but proxies are taken (job description, distance from blast)
    - Sometimes records exist (transfusions, dust levels)

Pitfalls in exposure assessments

- Observer bias – if disease ascertained at same time
  - Blind observers to study hypothesis
  - Standardized protocols
- Are all exposures the same?
  - EG Thoracenteses (pleural taps) ?
  - Complications of pleural tap at MGH/RVH >> MCI
    - Why – patients, their diseases, or the 'tappers'?

Cohort Studies – Outcome Assessments

- Baseline measures – ensure that the cohort members are free of disease at the start.
  - Easy if prospective, harder if retrospective
- Outcomes then measured periodically
  - Through questionnaire, exam, labs (direct)
  - Through health service utilization (databases)
  - Through vital statistics (databases)
- Case definition is very important for outcome assessments
  - Due to enhanced case finding of milder disease among members of the cohort

Pitfalls in outcome assessments

- Ascertainment bias – if patients with Factor X are more likely to have testing to detect outcome.
  - Solution = standardized protocols, or blinding to exposures
- Observer bias – if patients with Factor X more likely to be Diagnosed with outcome of interest
  - Common with more subjective tests – eg CXR
  - Solution – independent reviewers, blinded to exposure status (Factor X)

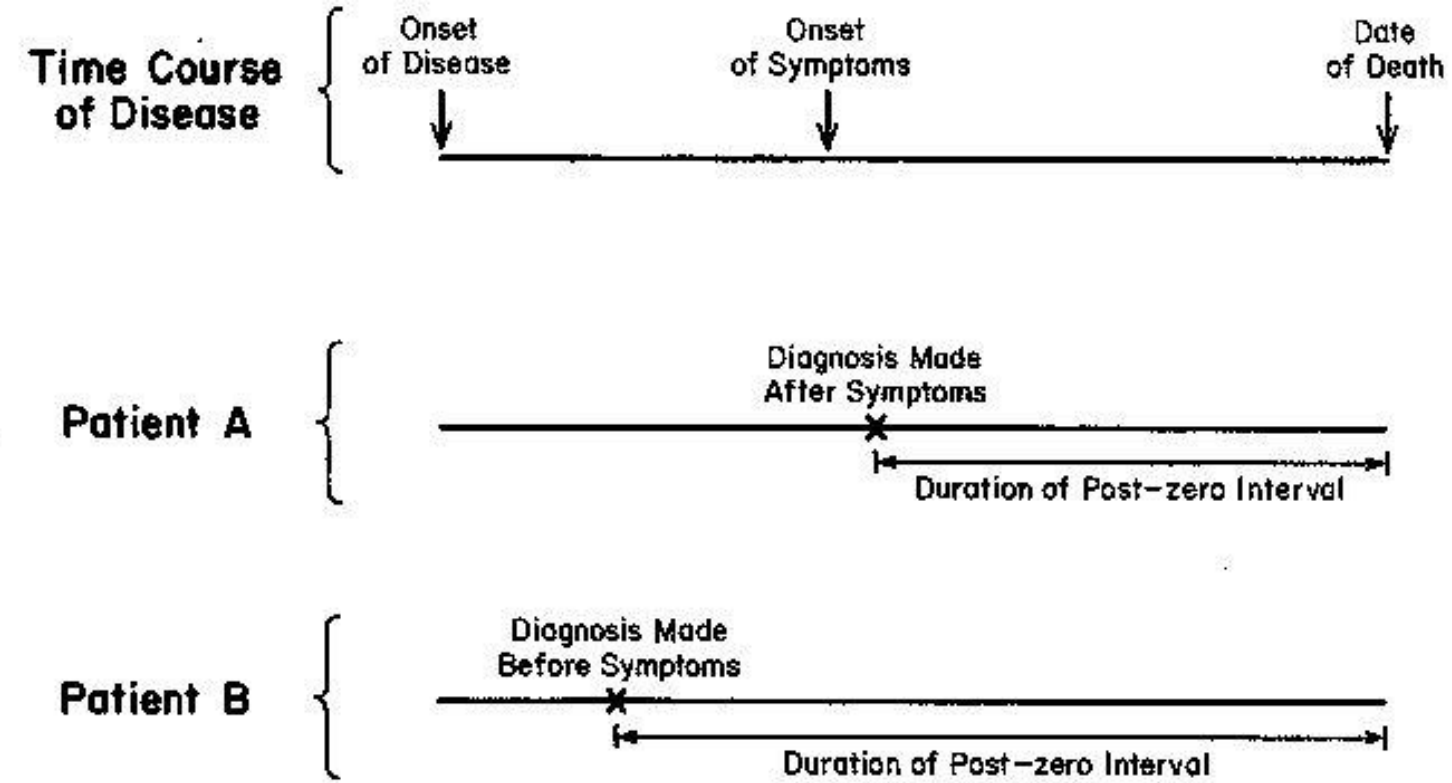


Figure 17–1. Illustration of lead-time bias. Because of presymptomatic detection of disease during screening, the post-zero interval of survival appears to be longer in Patient B than in Patient A, although no change has occurred in the total clinical course.

Cohort Studies – Measures of Incidence
Incidence rate

- Incidence rate =

$$\frac{\text{number developing disease}}{\text{Total number who entered cohort}}$$

per unit of time

- Incidence rate ratio = IRR =

$$\frac{\text{number with disease/number exposed}}{\text{number with disease/number unexposed}}$$

* *Note for IRR there is no unit of time but assumes that the amount of time was similar for those with and without disease and those exposed and unexposed*

Measuring Incidence in Cohort Studies

How to handle drop outs etc..?

- In a cohort study members drop out either because they are lost to follow up or die of other causes (or refuse to continue)
- How to count – keep them in or exclude them from analysis?
- It is better to use a method that allows variable length of follow up
- Otherwise in large long term cohort studies maybe only 50% of persons are still in the cohort at the end
- Also in a dynamic cohort have to be able to account for people who enter after the first year

Incidence Density Method - Example

Patient	Exposed	Follow up Years	Disease
1	YES	2	NO
2	YES	10	YES
3	NO	8	NO
4	NO	10	YES

- Incidence rate ratio = (1/2) / (1/2) = 1
- Density method = $\frac{(0/2 \text{ years}) + (1/10 \text{ years})}{(0/8 \text{ years}) + (1/10 \text{ years})}$
- Incidence density ratio = $\frac{(1/12)}{(1/18)}$

= 1.5

Incidence Rate Difference

- A patient asks "What is my risk because I smoke?" (or "how much will it go down if I quit smoking")
  - Can answer using incidence density ratio

$$= \frac{\text{incidence density if smoking}}{\text{incidence density non smoking}} = 1.5$$

In this example it would be one and a half times higher (or 50% more)

Incidence Rate Difference (cont'd)

- If a public health official asks you what is the impact of air pollution on cancer in Montreal?
- Incidence rate =

$$\frac{\text{number developing disease}}{\text{Total number who entered cohort per unit of time}}$$

- Incidence rate ratio =

$$\frac{\text{number with disease/number exposed}}{\text{number with disease/number unexposed}}$$

Cohort Studies – Survival Analysis

- Survival Analysis is a method of analysis is used if you have time to event for all event. Accounts for variable length of follow up. Survival analysis is advantageous when time to event is affected by the exposure.
- For example: A given cancer treatment increases survival at two years but five year mortality is unchanged. This would be an important advantage to patients.
- Survival analysis takes this into account by analysing time to disease.

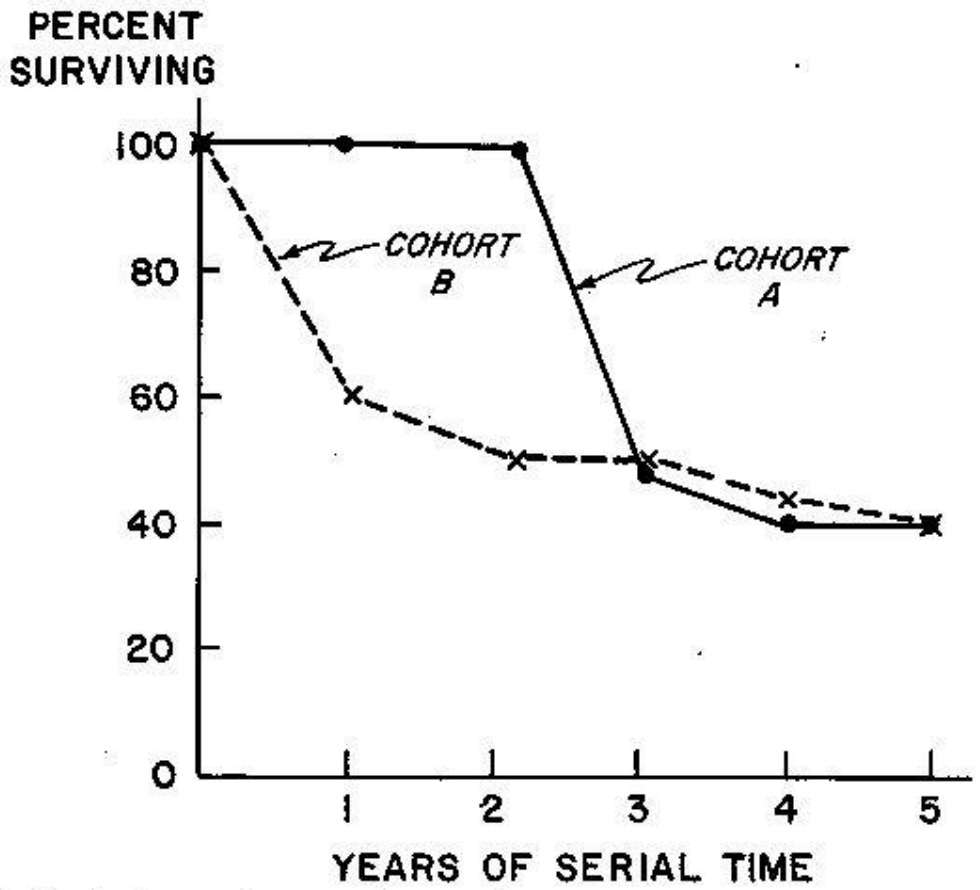


Figure 17–2. Survival curve in two cohorts having survival rates of 50% at three years and 40% at five years.

Cohort Studies – Survival Analysis Types

- Simplest – Direct
- Next simplest – actuarial or life-table
- Kaplan-Meier – still pretty simple. Calculates cumulative proportion free of outcome (survived) at each point in time when that outcome occurs. People who drop out or die of other causes are 'censored'. At each point numerator is all who have developed disease, while denominator is all without outcome in the interval just before
- Cox regression analysis – multivariate analysis with same basic principles

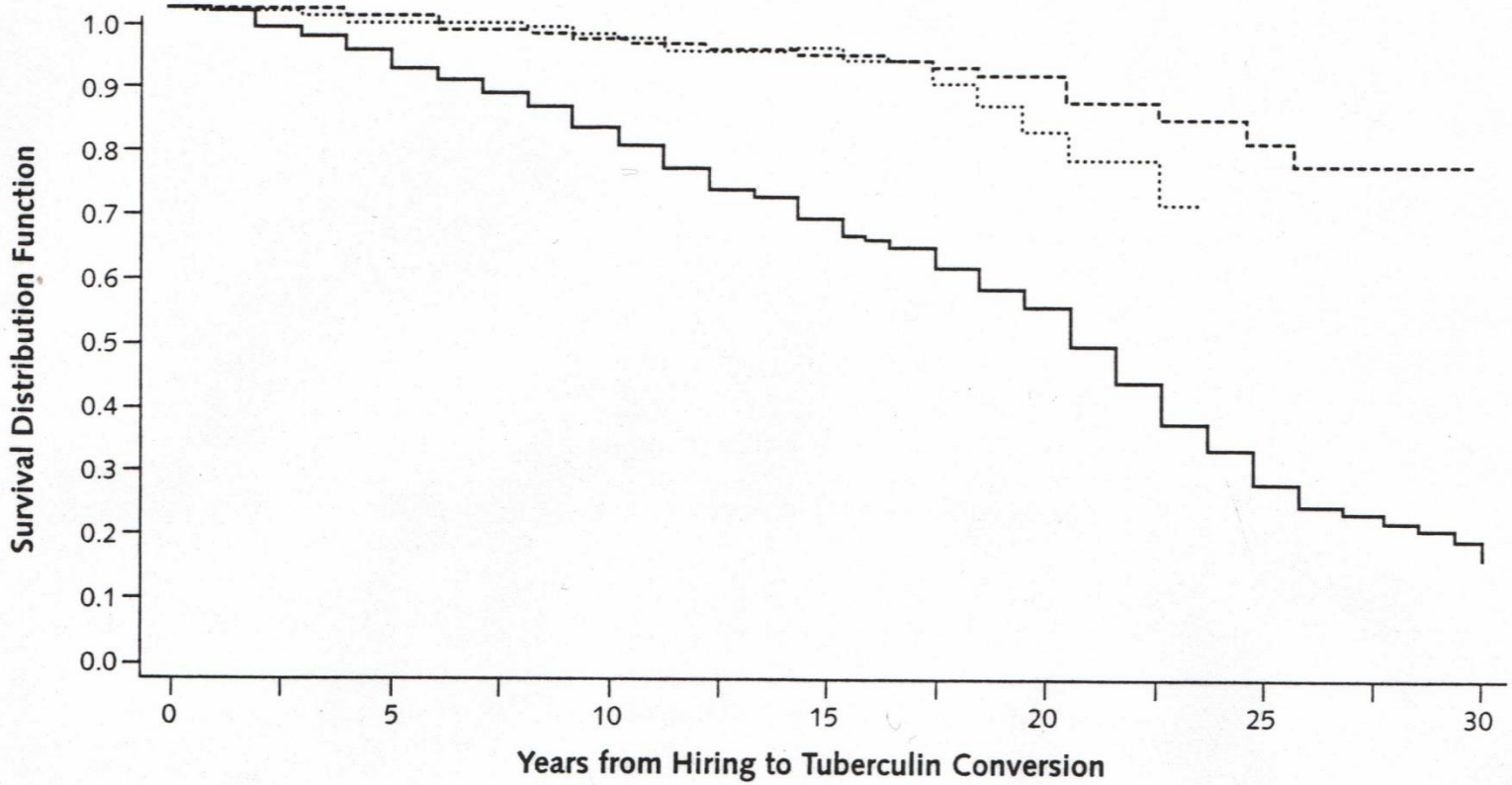
Table 17-1. DIRECT ARRANGEMENT OF SURVIVAL DATA FOR 50 PATIENTS

Interval	Censored During Interval	Cumulatively Followed from Onset Throughout This Interval	Died During Interval	Cumulative Deaths	Cumulative Mortality Rate	Cumulative Survival Rate
0-1 yr.	0	50	1	1	0.020	0.980
1-2 yr.	1	49	2	3	0.061	0.939
2-3 yr.	1	48	3	6	0.125	0.875

Table 17-3. VARIABLE-INTERVAL (KAPLAN-MEIER) ARRANGEMENT OF SURVIVAL DATA FOR 50 PATIENTS

Number Of Interval	Cumulative Survival Rate Before Death (s)	Time of Death (s) That End (s) Interval	Number Alive Before Death (s)	Number Of Death	Number Of Survivors	Interval Survival Rate	Censored Before Next Death
1	1.000	0.5	50	1	49	0.980	0
2	0.980	1.4	49	1	48	0.980	0
3	0.960	1.8	48	1	47	0.979	1
4	0.940	2.1	46	1	45	0.978	0
5	0.920	2.3	45	1	44	0.978	1
6	0.900	2.9	43	1	42	0.977	0
7	0.879	---	(42)	---	---	---	---

General Hospital Ventilation and time to TST conversion – Kaplan-Meier curves



Personnel, *n*						
< 2 ACPH	471	372	217	139	79	22
≥ 2 ACPH	651	518	237	104	42	20
Low-risk/ nonclinical	150	122	74	38	15	3