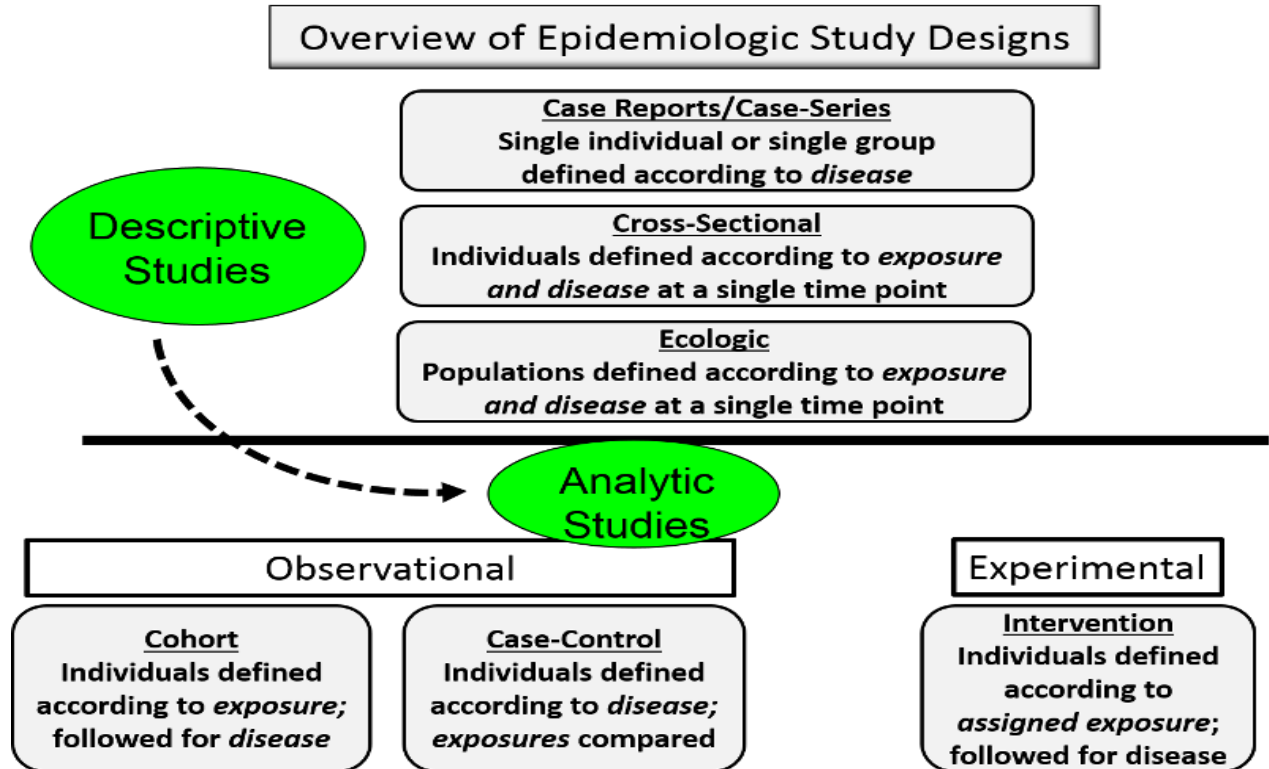


## Overview of Epidemiologic studies

The descriptive studies: case reports and case series, cross-sectional studies, and ecologic studies. In addition to identifying new problems and keeping track of trends in a population, they also generate hypotheses that can be tested using one of the analytic studies shown at the bottom.

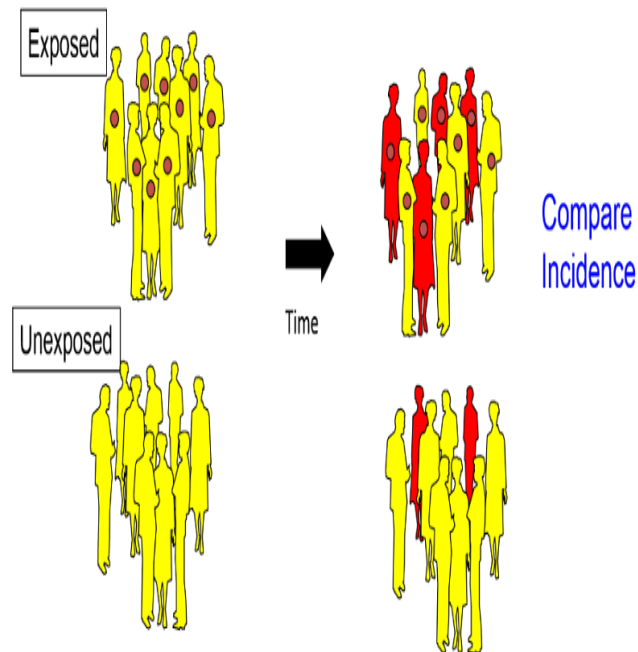


Note that cohort studies and case-control studies are observational studies, because investigators do not allocate exposure status. Some exposures are constituent (e.g., one's genome), some are behaviors and life style choices, and others are circumstantial, such as social, political, and economic determinants that affect health. None of these exposures are controlled by the investigators in observational studies; the investigators literally observe, collecting data on these exposures and on a variety of health outcomes. In contrast, intervention studies (also called clinical trials or experimental studies) are more like a true experiment in that the investigators assign subjects to a specific exposure (e.g., one or more treatment groups), and they are followed forward in time to record health outcomes of interest. Each of these analytic studies is useful in particular circumstances. Let's begin by discussing cohort studies.

### Cohort Studies

#### Key Features of Cohort Studies

In cohort studies investigators enroll individuals who do not yet have the health outcomes of interest at the beginning of the observation period, and they assess exposure status for a variety of potentially relevant exposures. The enrollees are then followed forward in time (i.e., these are longitudinal studies rather than cross-sectional) and health outcomes are recorded. With this data investigators can sort the subjects according to their exposure status for one of the exposures of interest and compare the incidence of disease among the exposure categories.



For example, in 1948 the Framingham Heart Study enrolled a cohort of 5,209 residents of Framingham, MA who were between the ages of 30-62 and who did not have cardiovascular disease when they were enrolled. These subjects differed from one another in many ways: whether they smoked, how much they smoked, body mass index, eating habits, exercise habits, sex, family history of heart disease, etc. The researchers assessed these and many other characteristics or "exposures" soon after the subjects had been enrolled and before any of them had developed cardiovascular disease. The many "baseline characteristics" were assessed in a number of ways including questionnaires, physical exams, laboratory tests, and imaging studies (e.g., x-rays). They then began "following" the cohort, meaning that they kept in contact with the subjects by phone, mail, or clinic visits in order to determine if and when any of the subjects developed any of the "outcomes of interest," such as myocardial infarction (heart attack), angina, congestive heart failure, stroke, diabetes and many other cardiovascular outcomes. They also kept track of whether their risk factors changed.

Over time some subjects eventually began to develop some of the outcomes of interest. Having followed the cohort in this fashion, it was eventually

possible to use the information collected to evaluate many hypotheses about what characteristics were associated with an increased risk of heart disease. For example, if one hypothesized that smoking increased the risk of heart attacks, the subjects in the cohort could be sorted based on their smoking habits, and one could compare the subset of the cohort that smoked to the subset who had never smoked. For each such comparison that one wanted to make, the cohort could be grouped according to whether they had a given exposure or not, and one could measure and compare the frequency of heart attacks (i.e., the cumulative incidence or the incidence rates) between the groups.

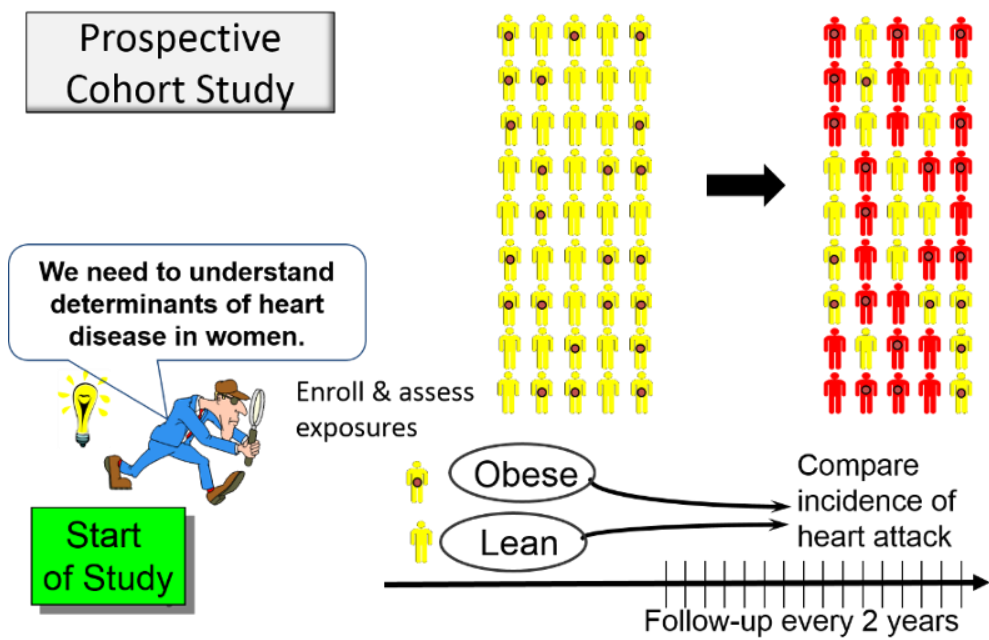
### **The Population "At Risk"**

From the discussion above, it should be obvious that one of the basic requirements of a cohort type study is that none of the subjects have the outcome of interest at the beginning of the follow-up period, and time must pass in order to determine the frequency of developing the outcome.

For example, if one wanted to compare the risk of developing uterine cancer between postmenopausal women receiving [hormone-replacement therapy](#) and those not receiving hormones, one would consider certain eligibility criteria for the members prior to the start of the study: 1) they should be female, 2) they should be post-menopausal, and 3) they should have a uterus. Among post-menopausal women there might be a number who had had a hysterectomy already, perhaps for persistent bleeding problems or endometriosis or prior uterine cancer. Since these women no longer have a uterus, one would want to exclude them from the cohort, because they are no longer at risk of developing this particular type of cancer. Similarly, if one wanted to compare the risk of developing diabetes among nursing home residents who exercised and those who did not, it would be important to test the subjects for diabetes at the beginning of the follow-up period in order to exclude all subjects who already had diabetes and therefore were not "at risk" of developing diabetes.

### **Prospective Cohort Studies**

Cohort studies can be classified as prospective or retrospective based on when outcomes occurred in relation to the enrollment of the cohort. The [Framingham Heart Study](#) is an example of a prospective cohort study. Another well-known prospective cohort study is the [Nurses' Health Study](#). The original Nurses' Health Study (NHS) began in 1976 by enrolling about 121,000 female nurses from across the United States who were initially free of known cardiovascular disease or cancer. (The Nurses' Health Study is now enrolling the third generation cohort, which includes male and female nurses).



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.

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.

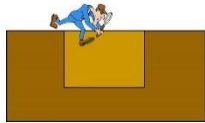
After baseline information is collected, the participants are followed "longitudinally," i.e. over a period of time, usually for years, to determine if and when they become diseased and whether their exposure status changes. Most studies of this type contact the participants periodically, perhaps every two years, to update information on exposures and outcomes. In this way, investigators can eventually use the data to answer many questions about the associations between exposures ("risk factors") and disease outcomes. For example, one NHS study examined the association between smoking and breast cancer and found that there was no significant association.

Another NHS study examined the association between obesity and myocardial infarction. They used reported height and weight to calculate BMI and categorized women into five categories of BMI. The table below summarizes their findings with respect to non-fatal myocardial infarction.

BMI	# non-fatal MIs	Person-Years	Inc. Ra
$\geq 30$	85	99,573	85.4

25.0-29.9	67	148,541	45.1
23.0-24.9	56	155,717	36.0
20.0-22.9	57	194,243	29.3
<20	41	177,356	23.1

The data above are from Willett WC, Manson JE, et al.: [Weight, weight change, and coronary heart disease in women. Risk within the 'normal' weight range.](#) JAMA. 1995 Feb 8;273(6):461-5.



**Potential Pitfall:** Analysis of prospective cohort studies can take place only after enough time has elapsed so that a sufficient number of subjects have 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 were collected *before* any subjects developed an outcome of interest.

### Follow Up in Prospective Cohort Studies

Ideally, investigators want to have complete follow-up on all subjects, but in large cohort studies that run for years, there are inevitably people who become lost to follow up as a result of death, moving, or simply loss of interest in participating. When this occurs, the investigators know the subject's exposure status prior to losing them, but not their outcome.

The biggest problem with substantial loss to follow up (LTF) is that it can bias the results of the study if the losses are different for one of the exposure-outcome categories. This will be illustrated in the module on bias.

There is no way to know if the losses are different for one of the exposure-outcome categories, so the only strategy to minimize bias from loss to follow up is to keep follow up high (in both prospective cohort studies and clinical trials).

### Strategies to Maintain Follow Up

- Choosing subjects who are motivated

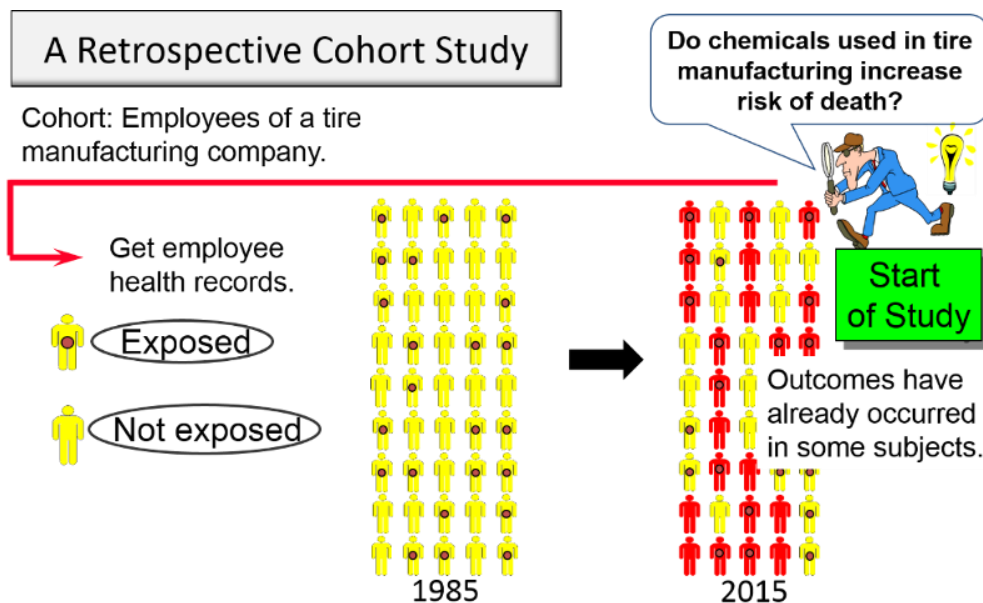
- ❑ Choosing subjects who are easy to track (e.g., registered nurses or physicians)
- ❑ Keeping subjects interested with newsletters and incentives;
- ❑ Being courteous and making them feel that they are members of a research "family"
- ❑ Frequent phone calls
- ❑ Making questionnaires easy to fill out

## Retrospective Cohort Studies

In contrast to prospective studies, retrospective studies are conceived *after* some people have already developed the outcomes of interest. The investigators *jump back in time to identify a cohort of individuals at a point in time before they had developed the outcomes of interest*, and they try to establish their exposure status at that point in time. They then determine whether the subjects subsequently developed the outcome of interest.

In essence, the investigators jump back in time to identify a useful cohort which was *initially free of disease and 'at risk'* of developing the outcome. They then use whatever records are available to determine each subject's exposure status at the beginning of the observation period, and they then ascertain what subsequently happened to the subjects in the two (or more) exposure groups. Retrospective cohort studies are also 'longitudinal,' because they examine health outcomes over a span of time. The distinction is that in retrospective cohort studies some or all of the cases of disease have already occurred before the investigators initiate the study. In contrast, exposure information is collected at the beginning of prospective cohort studies before any subjects have developed any of the outcomes of interest, and the 'at risk' period begins after baseline exposure data is collected and extends into the future.

Suppose investigators wanted to test the hypothesis that working with the chemicals involved in tire manufacturing increases the risk of death. Since this is a fairly rare exposure, it would be advantageous to use a special exposure cohort such as employees of a large tire manufacturing factory and conduct a retrospective cohort study.





The employees who actually worked with chemicals used in the manufacturing process would be the exposed group, while clerical workers and management might constitute the "unexposed" comparison group. Instead of following these subjects for decades, it would be more efficient to use employee health and employment records over the past two or three decades as a source of data. In essence, the investigators are jumping back in time to identify the study cohort at a point in time before the outcome of interest (death) occurred. They can classify them as "exposed" or "unexposed" based on their employment records, and they can use a number of sources to determine subsequent outcome status, such as death (e.g., using health records, next of kin, National Death Index, etc.).

Retrospective cohort studies are less expensive and more efficient than prospective cohort studies, because subjects don't need to be followed for years. However, the disadvantage is that the quality of the data is generally inferior to that of a prospective study. In the study of mortality and tire manufacturing chemicals the clerical staff may be much less exposed to the chemicals, but there are likely to be important differences in other factors that influence mortality (confounding factors), such as sex, age, socioeconomic status, education, diet, smoking, alcohol consumption, etc. Employee health records are unlikely to capture this information in sufficient detail to enable the investigators to adjust for differences in these other factors. (We will discuss adjusting for confounding later in the course.)

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* in some of the subjects.

## Strengths and Disadvantages of Cohort Studies

### Prospective Cohort Studies

#### Strengths of Prospective Cohort Studies

1. They can provide better quality of data on the primary exposure and also on confounding variables
2. Since exposures are assessed before outcomes occur, they are *less prone to bias*.

#### Disadvantages to Prospective Cohort Studies

1. They are more expensive and time consuming.
2. They are not efficient for diseases with long latency.
3. Losses to follow up can bias the measure of association.

### Retrospective Cohort Studies

#### Strengths of Retrospective Cohort Studies

1. They are useful for rare exposures, e.g., unusual occupational exposures

2. They are cheaper and faster than prospective cohort studies
3. They are more efficient for diseases with a long latency period

### **Limitations of Retrospective Cohort Studies**

1. Exposure data may be inadequate and there may be inadequate data on confounding factors, such as exercise, other health problems, etc.; old records were not designed to be used for future studies

### **Test Yourself**



## **Selection of Subjects for Cohort Studies**

The selection of subjects for a study is primarily dictated by the research questions and by feasibility.

### **General Cohorts**

For relatively common exposures and health outcomes a general cohort, such as residents of Framingham, MA, can be enrolled. The Framingham Heart Study, which began in 1948, enrolled 5,209 men and women 30-62 years old. At the time little was known about the determinants of heart disease and stroke, devastating health problems that had steadily increased in frequency throughout the 20th century. The investigators gathered extensive baseline information with questionnaires, lab tests, and imaging studies. They then followed the subjects, and had them return to the study office every two years for a detailed medical history, physical examination, and repeat lab tests. The Framingham study has been enormously successful in providing information about the most important determinants of cardiovascular diseases (e.g., hypertension, high cholesterol, smoking, obesity, diabetes, and physical inactivity). Framingham investigators also collaborate with leading researchers throughout the world on studies of stroke and dementia, osteoporosis and arthritis, nutrition, diabetes, eye diseases, hearing disorders, lung diseases, and genetic patterns of common diseases.

The Nurses' Health Study and the Black Women's Health Study would also be considered general cohorts, because they both provide the opportunity to study many exposures and many health outcomes among residents with a wide variety of occupations and circumstances. These studies enable investigators to collect exposure information on many common exposures (e.g., high blood pressure, smoking, alcohol



use, diet, exercise, etc.), and, after sufficient follow up time, many health outcomes can be studied. When conducting studies using data from a general cohort, the reference group comes from within the cohort, i.e., an internal comparison group. For example, when the Nurses' Health Study examined the association between exercise and heart disease, they carefully assessed physical activity and computed an overall "MET" [score](#) that takes into account the frequency, duration, and intensity of many activities. They then sorted them by MET score, divided the cohort into quintiles (i.e., five more or less equal numbers of subjects), and used the quintile with the lowest MET scores as the reference group against which they compared each of the other quintiles. [Manson JE, Hu FB, et al.: [A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women](#). N Engle J Med 1999;341:650-8].

## Special Cohorts

For rare or unusual exposures the obvious choice would be a special cohort that provides a sufficient number of subjects with the exposure of interest. Examples might include occupational exposures (e.g., asbestos, radiation, and pesticides), unusual diets, drug exposures (e.g., pregnant women treated with diethylstilbesterol in the 1960s), or rare events (e.g., Hurricane Katrina, the bombing of Hiroshima, exposure of responders to the attack on the World Trade Center on 9/11). With special cohorts there is obviously a focus on a single exposure, but many potential health outcomes can be studied. Another major difference from general cohorts is that selection of an appropriate comparison group can be challenging.

A good example of a special cohort study is the [US Air Force Health Study](#) on the effects of exposure to dioxin. During the Vietnam War, the U.S. military sprayed the herbicide dioxin ("agent orange") over Vietnam to expose enemy supply lines and bases. Airmen were exposed during spraying flights, while loading the chemical and while performing maintenance on the planes that were used. After the war, combat veterans who had been in Vietnam complained of a variety of health problems. In 1979, the US Congress directed that an epidemiologic study be conducted to evaluate adverse health effects associated with exposure to dioxin and other herbicides used during the Vietnam conflict. The study (informally called the "Ranch Hand Study") enrolled a special cohort consisting of US Air Force pilots who had flown missions to spray dioxin. The comparison group consisted of Air Force flight crews and maintenance personnel who served in Southeast Asia but had not been involved in herbicide spraying operations. Subjects have been followed for many years, and several analyses have found increased all-cause mortality and cardiovascular mortality in those exposed to dioxan. There was also evidence of an association with obesity and possibly diabetes. There were conflicting reports regarding the association between dioxan and cancers.

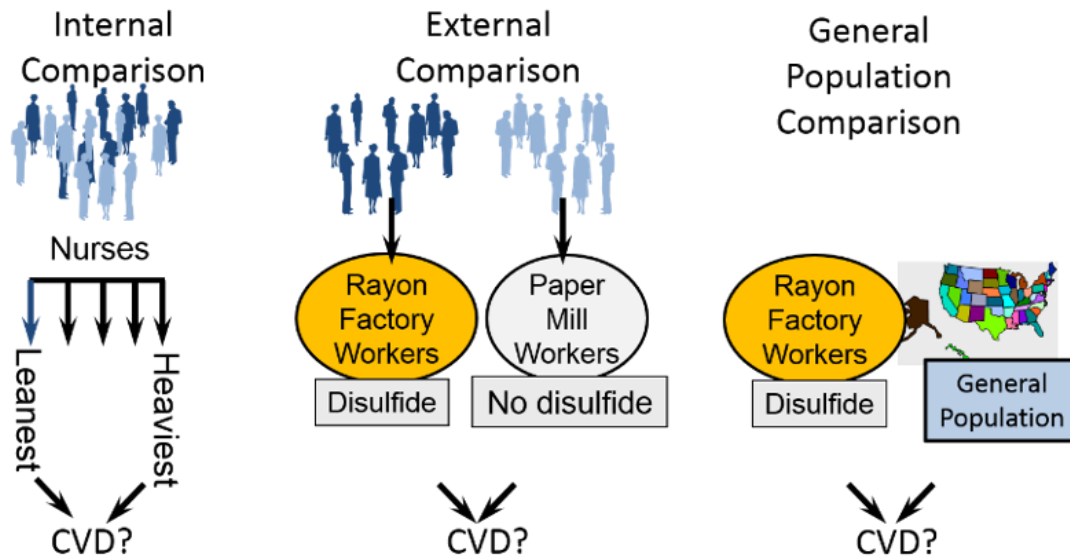
## Selection of a Comparison Group

The major challenge for the Air Force Health Study (AFHS) and other special cohort studies is selection of an appropriate comparison group. The goal of analytic studies is to compare health outcomes in exposed and unexposed groups that are otherwise as similar as possible, i.e., having the same distributions of all other factors that could have any association with health outcomes. We will see that intervention studies with large numbers of subjects randomly assigned to two or more treatment groups (exposures) can usually achieve this so that the groups being compared have similar distributions of age, sex, smoking, physical activity, etc., but random assignment does not occur in cohort studies. Suppose that a cohort study had smokers who were older than the non-smokers. It is well established that the risk of heart disease increases with age, i.e., it is an independent risk factor for heart disease, and if the smokers are older, they have an additional risk factor that will cause an overestimate of the association between smoking

and heart disease. This phenomenon, called **confounding**, occurs when the exposure groups that are being compared differ in the distribution of other determinants of the outcome of interest. Another concern is that the exposure groups being compared may differ in the quality or accuracy of the data that is being collected, and this can also bias the results (so-called **information bias**). Confounding and bias will be discussed later in the course, but for now, it is important to recognize the importance of selecting a comparison group that differs in exposure status but is as similar as possible to the exposed group in all other ways including:

1. Other factors that can influence the health outcome
2. The quality and accuracy of their data

The figure below depicts three studies of cardiovascular disease illustrating the general approaches to selecting a comparison group for a cohort study.



As noted earlier, general cohorts employ an **internal comparison group**, e.g., dividing the cohort into quintiles of BMI or quintiles of activity and using the quintile with the lowest BMI or the lowest activity as the reference group. This is the best comparison group for a general cohort study, because the subjects are likely to be similar in some ways, but they may still differ with respect to potentially confounding factors. For example, nurses who exercise regularly may be generally more health conscious (e.g., less likely to smoke; more likely to eat a healthier diet; more likely to take vitamins, etc.).

The second method is to use an **external comparison group**. A special exposure cohort consisting of workers in a rayon factory, was selected to study the association between disulfide exposure and risk of cardiovascular disease, and the comparison group consisted of workers in a paper mill. These two groups may be similar in age distribution, socioeconomic status, and other factors, but they may also differ with respect to other confounding factors. In addition, paper mills have their own mix of occupational exposures, which might also affect the likelihood of cardiovascular disease and bias the results.

The third approach is to use the **general population as a comparison group**, for example, if trying to determine whether workers in a rayon factory had higher mortality rates. This approach is less costly, and it is sometimes used for studies of occupational exposures when it is difficult to find an appropriate internal or external comparison

group. However, using rates of death or disease in the general population has a number of limitations:

- General population data are frequently limited to studies of mortality since accurate rates on specific health outcomes may not be available.
- General population rates include exposed and unexposed individuals.
- The general population is not really comparable because there are many confounding variables that cannot be controlled for.
- The general population includes people who are unable to work because of disease or disability (the "healthy worker effect" which is discussed in the module on bias).

### Test Yourself



## Basic Analysis of Cohort Study Data

One of the first steps in the analysis of an epidemiologic study is to generate simple descriptive statistics on each of the groups being compared. This helps characterize the study population, and it also alerts you and your readers to any differences between the groups with respect to other exposures that might cause confounding.

The illustration below is Table 1 from the study by Manson et al. on exercise and prevention of cardiovascular disease. Recall that they calculated each subject's MET score to estimate their overall activity level and then divided the cohort into quintiles based on the MET scores.

**TABLE 1. DISTRIBUTION OF INDICATORS OF CORONARY RISK ACCORDING TO QUINTILE GROUP FOR TOTAL PHYSICAL-ACTIVITY SCORE AT BASE LINE (1986).\***

VARIABLE	QUINTILE GROUP FOR TOTAL PHYSICAL ACTIVITY†				
	1	2	3	4	5
No. of women	13,859	15,065	14,598	14,326	14,640
Total physical-activity score (MET-hr/wk)					
Median	0.8	3.2	7.7	15.4	35.4
Range	0–2.0	2.1–4.6	4.7–10.4	10.5–21.7	>21.7
	percentage of group				
Risk indicator					
Currently smoking	28.2	23.7	19.6	17.4	17.5
History of hypertension	26.1	25.1	24.0	22.4	21.0
History of diabetes	4.2	3.3	3.7	2.8	2.6
History of hypercholesterolemia	12.0	11.4	11.7	11.6	10.6
Parent with myocardial infarction before age of 60 yr	14.0	14.4	13.9	14.2	14.0
Current postmenopausal hormone-replacement therapy‡	19.5	21.5	23.1	23.8	24.1
Use of multivitamin supplement	36.9	39.9	43.0	45.0	47.2
Use of vitamin E supplement	12.8	13.9	15.8	17.5	19.4
	mean				
Age (yr)	52.1	52.3	52.2	52.2	52.3
Alcohol consumption (g/day)	5.9	5.8	6.0	6.4	7.0
Body-mass index	25.1	24.6	24.2	23.9	23.5
Waist-to-hip ratio	0.79	0.79	0.78	0.78	0.77
Saturated fat (% of caloric intake)	12.3	11.9	11.7	11.5	11.2
Polyunsaturated fat (% of caloric intake)	6.2	6.1	6.2	6.1	6.1
Trans fat (% of caloric intake)	1.8	1.7	1.7	1.6	1.5
Dietary cholesterol (mg/1000 cal)	154	152	151	149	148

There are columns for each of the five quintiles in order from the least active to the most active. The rows list many variables that characterize the subjects and could also be confounders. Note that dichotomous variables are listed first and the percent with a given characteristic is listed for each quintile. For example, 28.2% of quintile 1 were current smokers, and this decreased steadily to 17.5% in the most active group (quintile 5). Therefore, smoking will be a potential confounding factor, because it is a risk factor for cardiovascular disease, and it differs among the exposure groups. Other possible confounding factors in table 1 include history of hypertension, history of diabetes, history of hypercholesterolemia (high blood levels of cholesterol), current use of hormone replacement therapy, use of multivitamins, and use of vitamin E supplements.

Continuous variables are listed in the lower half of Table 1, showing the mean value for each quintile of activity. Age is a risk factor for cardiovascular disease, but it is unlikely to cause confounding in this particular study, because the mean age is 52.1-52.3 years in all five quintiles. However, some of the other continuous variables do differ across the exposure groups, e.g., body mass index, alcohol consumption, and dietary cholesterol. Overall, increasing activity seems to be associated with trends in characteristics associated with a healthier lifestyle. If our goal is to understand the independent effect of exercise on risk of heart disease, then one must adjust for as many of these confounding factors as possible in the subsequent analysis. You will learn how to do this later in the course when we discuss confounding more completely.

You learned how to use R to generate descriptive statistics in the introductory module on R, and you have the tools to generate a table like Manson's Table 1 from a data set. The only other tool that you need is how to generate descriptive statistics in subsets of the data, e.g., the quintiles in the study by Manson et al. Methods for sub-setting are presented on the next page.

## Analyzing Data in Subsets Using R

### The `tapply()` command

The `tapply()` function is useful for performing functions (e.g., descriptive statistics) on subsets of a data set. In effect this enables you to subset the data by one or more classifying factors and then performing some function (e.g., computing the mean and standard deviation of a given variable) by subset. Note that `tapply()` is used for descriptive statistics (e.g., mean, sd, summary) for continuously distributed variables. For categorical variables you should use the `table()` function to get counts of categorical variables and use the `prop.table()` function to get proportions. The basic structure of the `tapply` command is:

```
tapply(<var>, <by.var>, <function>)
```

where `<var>` is the variable that you want to analyze, `<by.var>` is the variable that you want to subset by, and `<function>` is the function or computation that you want to apply to `<var>`.

For example, suppose I have a data set with continuous variables `Dubow` (Dubow Score), `DrugExp` (Drug Exposure) and `Ppregwt` (Pre-pregnancy weight). My goal is to sort the data set by `DrugExp` and then compute the mean and standard deviation of `Dubow` Scores and `Pre-pregnancy weights` for each category of `DrugExp`.

```
> tapply(Dubow, DrugExp, mean) # Gives means of
Dubowitz score by drug exposure
> tapply(Dubow, DrugExp, sd) # Gives the standard
deviations of Dubowitz score by drug exposure
> tapply(Ppregwt, DrugExp, mean) # Gives the means
of pre-pregnancy weight by drug exposure
> tapply(Ppregwt, DrugExp, sd) # Gives the
standard deviations of pre-pregnancy weight by
drug
> tapply(Birthwt, DrugExp, t.test) # Gives 95%
confidence interval for exposed and unexposed
in one output
```

### An Alternate Method of Subset Analysis

Getting descriptive statistics by category can also be achieved as follows:

```
> mean(Birthwt[DrugExp==1]);
mean(Birthwt[DrugExp==0]) # means for each
```

```
exposure group
> sd(Birthwt[DrugExp==1]);
sd(Birthwt[DrugExp==0])# standard deviation for
each exposure group
> t.test(Birthwt[DrugExp==1]) # 1-sample t-test
to get 95% CI for those exposed to drugs
> t.test(Birthwt[DrugExp==0]) # 1-sample t-test
to get 95% CI for those unexposed to drugs
```

Using the double equal sign (==) basically means "only if DrugExp equals 1".

## Creating a Dichotomous Variable from a Continuous Variable

Suppose my data set has a continuously distributed variable called "birthwt", which is each child's weight in grams at birth, but I wish to create a new variable that categorizes children as having Low Birth Weight (lowBW), i.e. less than 2500 grams or not. I can do this using the `ifelse()` function, which has the following format:

```
> ifelse(<logical statement>, <if true>, <if
false>)
```

Example:

```
> lowBW <-ifelse(Birthwt<2500,1,0)
```

If the variable `birthwt` is less than 2500, then the new variable `lowBW` will have a value of 1, meaning "true"; if not, it will have a value of 0 meaning "false". When this command is executed, you should see the new variable show up in the global environment window at the upper right corner of RStudio. Note that you should reattach your data set so that the new variable will be recognized.

If you want the `lowBW` category to include those whose weight was exactly 2500 grams, then use `<=` (less than or equal to) as below.

```
> lowBW <-ifelse(Birthwt<=2500,1,0)
```

## Crude Measures of Association in a Cohort Study (or Intervention Study)

After generating the descriptive statistics for an epidemiologic study, the next step is to generate estimates for the magnitude of association between the primary exposure of interest (e.g., physical activity level in the Manson study) and the primary outcome of interest (e.g., development of cardiovascular disease). As noted above, there may be confounding factors that can distort the estimated measure of association, but one still begins by generating crude measures of association, i.e., estimates that have not yet been adjusted for confounding factors.

### Test Yourself





The table below shows data from the top portion of Figure 2 from the study by Manson et al.

Table – Relative Risk of Coronary Events According to Quintile Group for Total Physical Activity

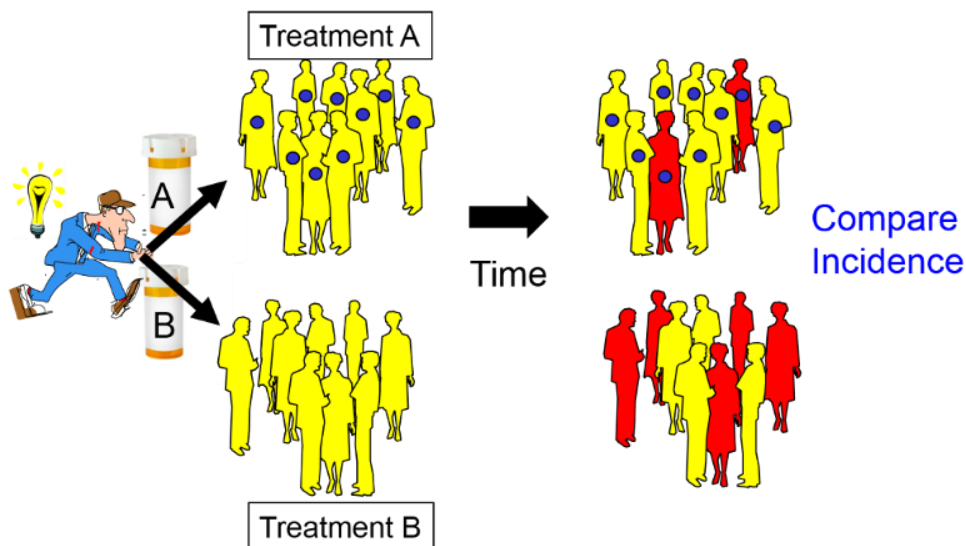
Variable	Quantile Group Based on Phys			
	1	2	3	
MET-hours/week	0-2.0	2.1-4.6	4.7-10.4	
Number of coronary events	178	153	124	
Person-years of follow up	106,252	116,175	112,703	

Using the data in the table above, a) compute the incidence rate ratio and the incidence rate difference for moderate activity compared to the least active subjects, and b) write an interpretation of your findings. Complete both parts before comparing your answers to those at the link below.

Answers

## Intervention Studies (Clinical Trials, Experimental Studies)

Intervention studies (clinical trials) are similar to prospective cohort studies in design in that subjects with or without a given exposure are followed over time to compare incidence of the outcome of interest. The key difference is that prospective cohort studies are observational, but in clinical trials the investigators assign subjects to the exposure groups



While this design is frequently used to evaluate new drugs, it can be used to evaluate the efficacy of

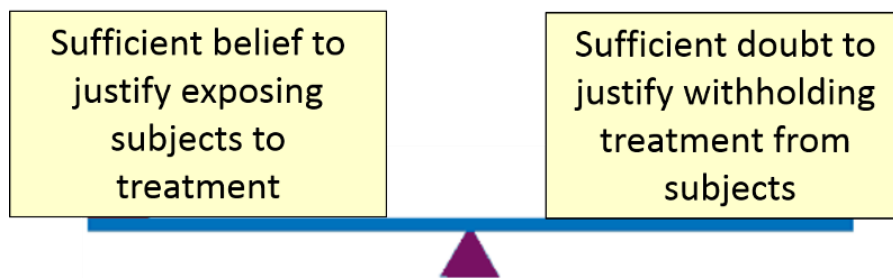
- Diets (e.g., primary prevention of cardiovascular disease with a Mediterranean diet)
- Exercise regimens (e.g., a clinical trial of exercise to alleviate post-partum depression)
- New programs (e.g., pre-natal care in groups of 8-10 women versus usual one-on-one pre-natal care)
- New clinical management schemes (e.g., a protocol to reduce post-operative complications)

However, unlike prospective cohort studies in which investigators record exposures that subjects already have, in clinical trials the investigators assign patients to one of the exposure groups being compared. Ideally, this assignment is done with random allocation, meaning that each subject has an equal chance of being assigned to any one of the "exposures."

## Ethical Considerations

Investigators assign patients to competing treatments in clinical trials, and this raises the question of whether it is ethical to do this. Certainly, it is not ethical to test all exposures in this fashion. It would be unethical, for example, to conduct a clinical trial on the effects of smoking, particularly since we know that the harm caused by smoking far outweighs any potential benefits, such as relaxation or weight control.

On the other hand, consider a situation in which a new drug has been developed to treat breast cancer. Perhaps it has been found to be effective in cell cultures and in animal models, and perhaps preliminary studies in small groups of human volunteers have shown some evidence of effectiveness with minimal side effects. In other words, there is reason to believe that it might be a beneficial new treatment, but there is also doubt about effectiveness and possible side effects. Testing on a large scale with a comparison group may show that it is not so effective or that its side effects are unacceptable. This is what is referred to as  **equipoise** , i.e., the balance between sufficient belief in its potential benefit and safety that one can justify exposing some subjects to it and sufficient doubt about its benefit and safety that one can justify withholding it from some subjects.



It is unethical to conduct a clinical trial in the absence of equipoise, and if equipoise ceases to exist during the course of a clinical trial, the trial must be discontinued.

Before research on living humans is conducted, a detailed protocol must be submitted to an **Institutional Review Board (IRB)** for review and approval. This is true not only of clinical trials, but also all other types of human research including case-series, cross-sectional surveys, prospective and retrospective cohort studies, and case-control studies. [For a more detailed overview of the ethical considerations for human research, see our [online module on Research Ethics](#).]

**"Human Research"** is defined as any systematic investigation involving living humans (including research development, testing and evaluation), designed to develop or contribute to generalizable knowledge.

## Informed Consent

One of the key things that an IRB will consider is whether potential subjects have provided informed consent, which is the process by which study participants consent to be subjects only after becoming fully informed and understand all aspects of the research including the purpose, risks, type of information to be collected, potential benefits, and alternatives to the research. Informed consent should allow people to make a fully informed decision about whether to participate in a study or not based on their own goals and values. Informed consent must be obtained before assignment to a treatment group, and consent can be withdrawn at any time during the study.

Potential participants must be fully informed about:

- The purpose of the study
- The treatment options (including alternatives) and the potential outcomes
- The risks and potential benefit of the study
- Randomization, i.e., that the treatment they receive is not their choice
- What will be required of them (questionnaires, visits, samples collected, etc.)
- Their ability to withdraw from the study at any time without consequence

## Types of Intervention Studies

### Therapeutic vs. Preventive

Clinical trials in individuals can be classified as either therapeutic or preventive, as in these examples:

**Therapeutic Trials:** New treatments are tested for the effectiveness in treating disease, e.g.,

- Does the drug herceptin improve survival in women already diagnosed with breast cancer?
- Does treatment with Tamiflu shorten the duration and improve survival in patients with bird flu?

**Preventive Trials:** Healthy or high-risk individuals are tested to determine whether a treatment prevents disease, e.g.,

- Does the drug tamoxifen prevent development of breast cancer in women who have a high risk of developing breast cancer?
- How effective is this year's influenza vaccine in preventing the flu?
- Does a Mediterranean diet reduce the incidence of cardiovascular disease?

### Individual Trials vs. Community Trials

Preventive measures can also be allocated on a community level – so-called community trials. A classic example is the Newburgh-Kingston Caries Fluoride Study which began in 1947. Fluoride was added to the water supply of Newburgh, NY, and the incidence of

dental caries in Newburgh was then compared to the incidence in Kingston, NY, which did not receive fluoride. The trial demonstrated that addition of tiny amounts of fluoride to the water supply reduced dental caries by two thirds in children who began drinking fluoridated water within their first two years.

The key difference is that in community trials the treatments being studied are allocated not to individuals, but to entire communities.

#### Individual Trial:

Treatment allocated to *individuals*, e.g., HIV treated with 2 drugs vs. 3 drug regimen



## Phases of Individual Therapeutic Trials

When most people hear reference to a clinical trial, they think of phase 3 trials in which large numbers of subjects are enrolled and randomly assigned to one of the treatment groups. However, phase 3 trials of new drugs with potentially harmful side effects are preceded by extensive studies in lab animals and by phase 1 and phase 2 trials in human volunteers.

### Phase 1 Clinical Trials

If studies in animals suggest efficacy and safety, a phase 1 trial can be conducted in a small group (10-30) of human volunteers over 2-12 months, primarily to test for safety and to identify side effects, but also to get some information on effective dose.

### Phase 2 Clinical Trials

Phase 2 clinical trials involve more volunteers than phase 1, and they typically last about two years. They usually involve two or more groups receiving different doses of the new drug in order to establish its therapeutic range of the drug, i.e., doses at which it is effective and has an acceptable level of side effects. If results suggest efficacy and safety, a phase 3 trial will be conducted.

### Phase 3 Clinical Trials

Phase 3 trials are similar to prospective cohort studies in their design, except that the exposure of interest is a drug or some other intervention that is randomly assigned to the participants by the investigators. To facilitate this presentation of phase 3 trials we will

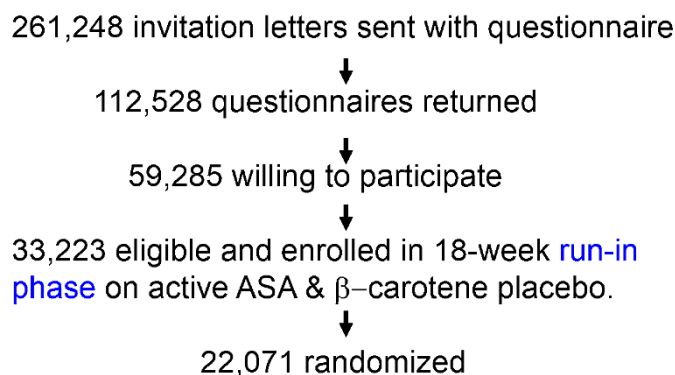
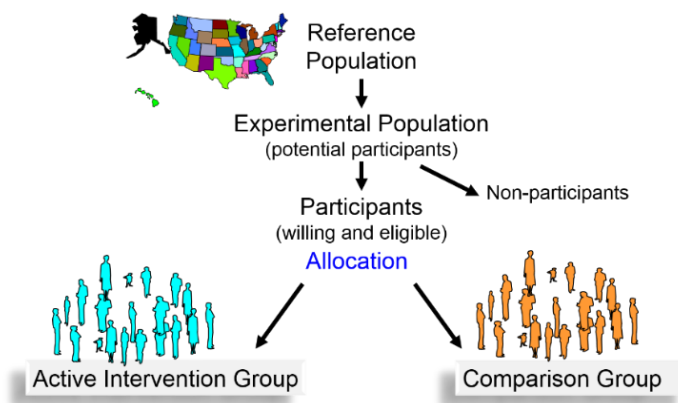
focus on the first [Physicians' Health Study](#), which began in 1981 in order to test the efficacy of aspirin in primary prevention of myocardial infarction. A second goal of the study was to evaluate the efficacy of beta-carotene in preventing cancer, but this discussion will focus on the aspirin component.

## The Physicians' Health Study on Aspirin

As early as the 1950s there were case series and small clinical trials suggesting that aspirin might be beneficial in preventing myocardial infarction (heart attack). However, the reduction in risk appeared to be modest, and the studies were too small to demonstrate a statistically significant benefit. Therefore, investigators at Harvard Medical School sought funding for a large phase 3 clinical trial.

### Enrollment

In 1981, after receiving approval from the Institutional Review Board at Harvard Medical School, the investigators mailed invitation letters, consent forms, and enrollment questionnaires to all 261,248 registered *male* physicians in the US between 40 and 84 years old. (Phase 1 and phase 2 trials were unnecessary, because aspirin was a commonly used drug with known dosage range and known side effects.)

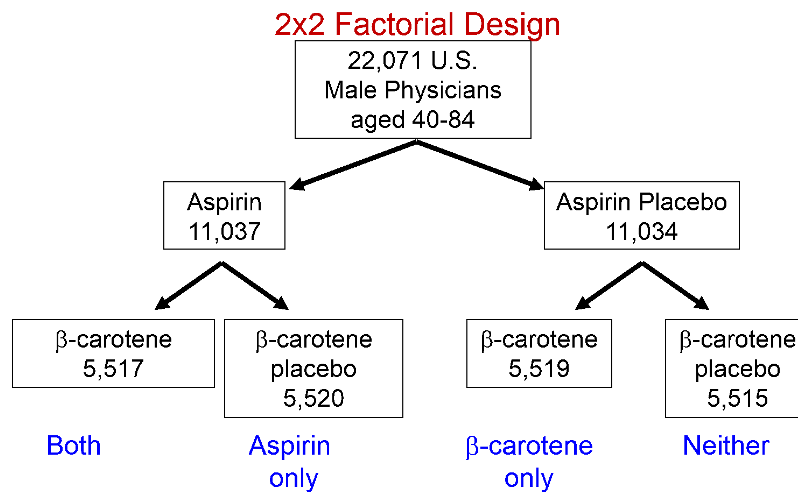


Questionnaires were returned by 112,528 physicians, but only 59,285 of those were willing to participate in the trial. Of those, 26,062 could not participate because they had one or more of the exclusion criteria:

- past myocardial infarction, stroke, or transient ischemic attack;
- cancer (except non-melanoma skin cancer);
- current renal or liver disease;
- peptic ulcer;

- gout;
- contraindication to or current use of either aspirin or beta-carotene.

Informed consent was obtained from the 33,223 who were willing and eligible to participate. Since regular aspirin use has the potential to cause gastritis and bleeding problems, these physicians were enrolled in an 18-week run-in phase, during which all received active aspirin and placebo beta-carotene for 18 weeks. Some had unpleasant side effects, others decided not to participate, and some were excused because they didn't take the medications reliably. The remaining 22,071 men were then randomly assigned to one of four treatment groups.



## Randomization and Blinding (Masking)

### Randomization



Randomization is a method of allocating subjects in a clinical trial to treatment groups such that every subject has an equal chance of receiving any one of the treatments or interventions. This can be achieved by any fair method that assigns subjects in a completely unpredictable fashion. One could use the flip of a coin if there are only two treatment options, but more commonly a table of random numbers or computer-generated random numbers are used. Other methods, such as assigning subjects based on odd or even calendar date, can be "gamed" in a way that biases assignment.

If assignment is truly unpredictable, then there is no bias in assignment, and neither the subjects nor the investigators can influence assignment. In addition, randomization of a large number of subjects tends to result in groups that differ only in treatment and are comparable with respect to all other factors and characteristics that might influence the outcome. As a result, randomization is the best method for eliminating confounding.

### Blinding

**Blinded** (or "**masked**") studies are those in which the subjects, and possibly the investigators as well, are unaware of which treatment the subject is receiving, e.g.,

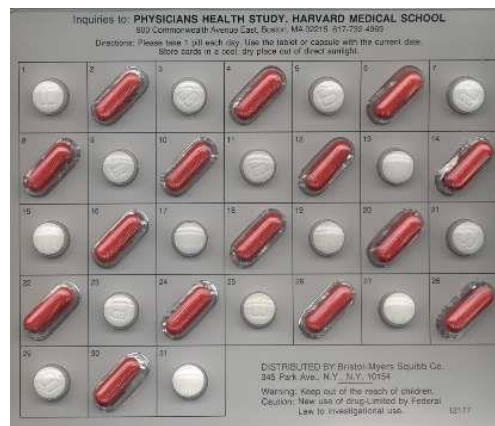


active drug or placebo. Blinding is particularly important in drug trials when the study is assessing **subjective outcomes**, such as relief of pain or anxiety.

It isn't always possible to mask the treatments. For example, subjects randomly assigned to follow either a specific exercise regimen or continue their usual level of activity cannot be blinded.

- **Single-blinded:** the subjects are unaware of which group they have been assigned to.
- **Double-blinded:** Neither the subjects nor the investigators are aware of the treatment assignment until the end of the trial.

A **placebo** is an inert substance identical in appearance to the active treatment. Its purpose is to facilitate blinding by making the groups as similar as possible in the perception of treatment and to promote compliance. In the Physicians' Health Study participants were given a blister pack for each month (shown in the image below) that contained white tablets and red capsules that were taken on alternate days. The white tablets contained either 325 mg. of aspirin or an identical-looking inert substance; the red capsules contained either beta-carotene or an inert substance. The use of monthly blister packs also made it easier for participants to keep track of whether they had taken the correct pill each day.



It is not always ethical to use a placebo. If there is already a standard treatment or method of care, it would be unethical to withhold it. A new treatment should be compared to the standard therapy rather than to a placebo.

**Example of Placebo Use to Achieve Blinding:**

Glucosamine and chondroitin are naturally occurring substances that are structural components of the cartilage that lines our joints. Health food stores began selling supplements to people as a prevention (or treatment) for osteoarthritis despite a lack of evidence of their benefit in humans. Clegg and colleagues conducted a double-blind, randomized clinical trial in 1583 subjects with symptomatic osteoarthritis of the knee. Participants were randomly assigned to one of five treatment arms in order to test the efficacy of glucosamine and chondroitin. The primary outcome was greater than 20% decrease in total score on the WOMAC pain scale from baseline to week 24. Some of their results are shown in the table below.

	<b>Pain relief &gt;20%</b>	<b>Minimal Effect</b>	<b>Total # Subjects</b>
Placebo	188	125	313
Anti-inflammatory drug	223	95	318

Glucosamine	203	114	317
Chondroitin	208	110	318
Glucosamine + Chondroitin	211	106	317

Data from Clegg DO, et al.: *Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. N Engl J Med 354:795, 2006.*

Perhaps the most remarkable observation is the response in the group treated with the placebo which had a cumulative incidence of >20% pain relief of 60% (188/313 = 0.60 = 60%)! This is an example of **the "placebo effect"** in which patients who perceive they are being treated often report subjective improvement, even if the treatment has no effect. Placebos make the *perception of treatment similar* among groups and provide a reference group that takes into account the placebo effect. Note also that the group treated with glucosamine and chondroitin had only a slightly greater response rate of 67%.

## Analysis of Clinical Trial Data

The analysis of clinical trial data is very similar to the previously described analysis of data from a cohort study. The first step is to generate simple descriptive statistics on each of the groups being compared in order to characterize the study population and alert you and your readers to any differences between the groups with respect to other exposures that might cause confounding. If large numbers of subjects have been randomly assigned to the treatment arms, the groups should be comparable. If there are more than minor discrepancies, the investigators need to review the randomization procedures and consider adjusting for confounding by other methods.

The table below shows just a portion of the data from the table of descriptive statistics from the Physicians' Health Study on aspirin.

	Aspirin (n=11,037)	Placebo (n=11,034)
Age (years)	53.2 ± 9.5	53.2 ± 9.5
Systolic BP (mm Hg)	126.1 ± 11.3	126.1 ± 11.1
Diastolic BP (mm Hg)	78.8 ± 7.4	78.8 ± 7.4
History of hypertension (%)	13.5	13.6
History of high cholesterol (%)	17.5	17.3
Cholesterol level	212.1 ± 44.2	212.0 ± 45.1
History of diabetes (%)	2.3	2.2

Note that the two groups were remarkably similar on these and other characteristics, indicating that randomization had been successful.

After generating the descriptive statistics, the next step is to generate crude estimates for the magnitude of association between the primary exposure and the outcomes of interest.

## Test Yourself



After 5 years of follow up In the Physicians' Health Study, an interim analysis found that among the 11,034 men assigned to the placebo group there had been 213 non-fatal myocardial infarctions. Among the 11,037 men assigned take 325 mg. of aspirin every other day, there had been 126 non-fatal myocardial infarctions.

Summarize these finding in a contingency table and compute the cumulative incidence in each group, the risk ratio, and the risk difference. Then interpret the risk ratio and the risk difference. Complete all of these tasks before comparing your answers to the ones provided in the link below.

Answer

## Strengths and Limitations of Clinical Trials

Large randomized clinical trials can provide strong evidence of the true effect of a treatment or intervention, because they provide excellent control of confounding, but they also have some limitations:

### Strengths of Intervention Studies (Clinical Trials)

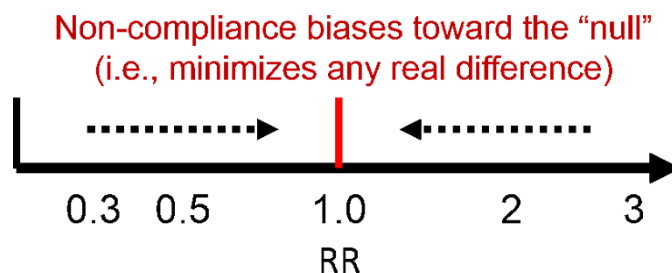
- They provide the best means of minimizing the effect of confounding
- They avoid bias in allocation to exposure groups
- Large randomized clinical trials are the best design for detecting small to moderate effects that may be important

### Limitations

- Ethical issues need to be considered
  - o Risks to subjects versus potential benefits
  - o Does equipoise exist? Some questions cannot be answered ethically with a clinical trial.
- They are usually time consuming and costly
- Lengthy trials run the risk of loss to follow up (LTF), and if LTF is different for one of the exposure-outcome categories, the measure of association will be biased, just as with prospective cohort studies(see the m Bias).
- Invariably, some subjects will fail to adhere to the protocol, and non-adherence will cause an underestimation of the measure of association (see below).

## Non-adherence

Ideally, the investigators want to compare exposed subjects to non-exposed in groups that are similar with respect to confounding factors. The true benefit of a new drug will be underestimated if subjects given the active medication fail to take it, causing subjects who were actually not exposed to be mixed in with the exposed subjects who were actually taking the medication. This mixing of the exposure groups dilutes the apparent benefit causing underestimates of association. The same thing occurs if people in the placebo group begin taking the active medication. This occurred in the Physicians' Health Study in which follow up questionnaires estimated that about 15% of the subjects assigned to the aspirin group did not take it, and a similar proportion of subjects in the placebo group used aspirin fairly regularly. This would cause an underestimate of the true benefit. In this case, in which the exposure was preventive with an observed risk ratio = 0.59, the true risk ratio would have been even smaller. In other words, non-adherence caused a "bias toward the null," an underestimate of the true benefit.



Non-compliance can occur due to side effects of the treatment, illness, or loss of interest in the study.

### How to Promote Adherence in a Clinical Trial

1. Begin with an interested group of participants
2. Make it easy to participate
  - o Present a realistic picture of the protocol during informed consent
  - o Exclude participants with pre-existing conditions that make compliance difficult
  - o Simplify the protocol as much as possible
  - o Conduct a run-in period if necessary
3. Use blinding and placebos
4. Keep in touch
  - o Maintain frequent contact with subjects WITHOUT interfering with treatment
  - o Provide incentives (free check-ups, transportation, t-shirts, birthday cards)

## Data Safety and Monitoring Board

All clinical trials that involve more than minimal risk are required to have a Data Safety and Monitoring Board (DSMB), which is an independent board of experts not involved in the study who periodically review the data in a trial to evaluate safety, study conduct, and interim results. They can recommend that the study be continued, modified, or terminated. The DSMB for the Physicians' Health Study recommended that the study be terminated after five years because the benefits of taking low-dose aspirin were so clear that continuing to withhold aspirin from the placebo group was not ethically justified. The DSMB felt that equipoise no longer existed.

### Intention-to-Treat Analysis versus Efficacy Analysis

The greatest advantage of large randomized clinical trials is that they provide control of confounding. However, as already noted there can be problems due to loss to follow up

and lack of adherence to the protocol. It might be tempting to limit the analysis to subjects who completed the study and who adhered to the study protocol, but this efficacy analysis may not provide strong control of confounding, because subjects have, in essence, self-selected whether they would remain in the study and adhere to the protocol. For this reason, well-done clinical trials will conduct and report the results of an intention-to-treat analysis in which subjects are included in the analysis in the groups to which they were randomly assigned regardless of whether they adhered to the protocol. We already noted that non-adherence will bias the results toward the null, i.e., underestimate the association if there is one. However, the intention-to-treat analysis provides the best opportunity to examine the association in the absence of confounding. Many reports will provide the results of the intention-to-treat analysis and the efficacy analysis as well, and they may also analyze sub-groups of subjects, but these analyses need to use other methods to minimize the effects of confounding.