Experimental Studies

Two major categories of epidemiologic studies are:

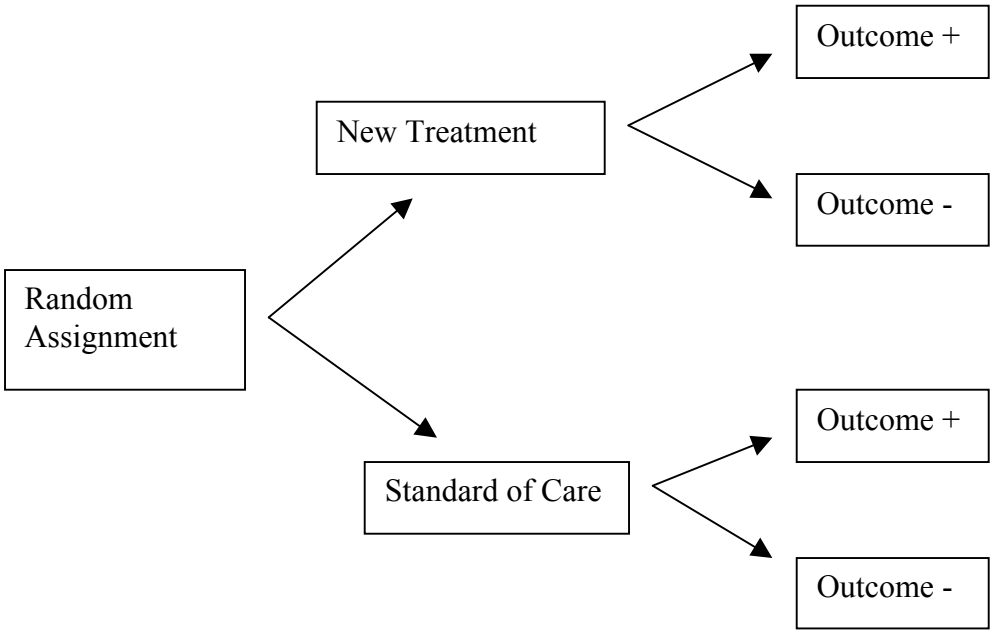
1. Experimental Studies
2. Observational (Non-Experiment Studies

The key distinction between these two types of study designs is the role of the investigator. In an experimental study, the investigator has the active role of assigning subjects to treatment (exposure) groups for the primary purpose of evaluating the effect of that treatment on an outcome. On the other hand, in observational studies exposures (including treatments) are self-selected or determined by someone else (e.g. one's physician) for a primary reason other than evaluating the effect of the treatment. The role of the investigator is passive, to observe outcomes and measure the association between an exposure and an outcome.

For example, suppose and investigator wished to evaluate the effect of daily consumption of a low dose aspirin on the risk of developing a myocardial infarction (heart attack). In an experimental study the investigator might assign some subjects to a regimen of daily low aspirin and assign others to a placebo drug. Usually the assignment is dictated by some type of randomization, whereby each subject has an equal chance of assignment to either group. The use of a **placebo** and **randomization** will be discussed later in these lecture notes.

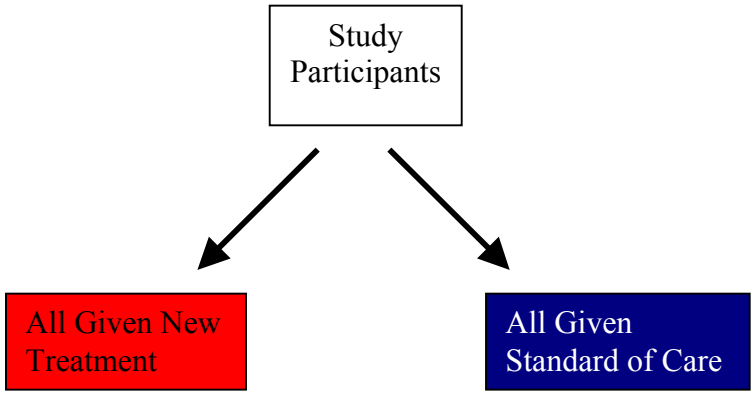
In an observational study the investigator might enroll a series of subjects who report taking daily aspirin and a comparison group of non-aspirin users. The reasons for aspirin use in the first group might because of a personal decision (subjects anticipating some benefit from daily aspirin use) or at the recommendation by one's physician (for disease prevention). One concern with such a study is the group taking aspirin might differ with respect to many indications (confounders) that are risk factors for the outcome. The topic of **confounding** will be discussed in future lectures in this course.

The type of experimental study described above is often referred to a randomized controlled trial or a **randomized clinical trial (RCT)** if it performed in a clinical setting. The following figure describes the basic principles of a Randomized Clinical Trial comparing outcomes for patients who are randomized to receive a new treatment versus receiving standard of care.

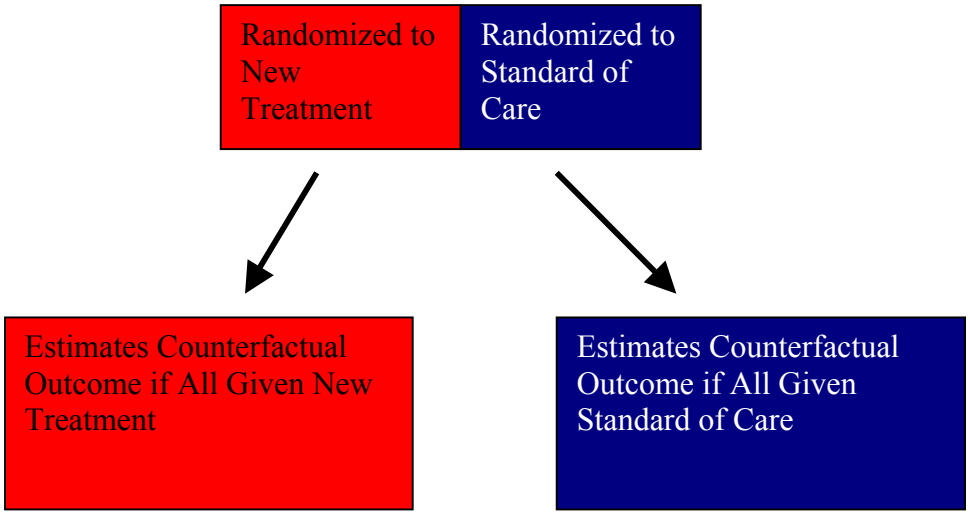


RCT and Causal Inference

Recall in a previous lecture we described the causal effect of a treatment in a group of study participants to be the difference between counterfactual outcome when everyone is given the treatment (Y^1) and when no one is given the treatment (Y^0). This is depicted in the following figure, where the alternative to receiving a treatment is receiving a standard of care rather than no treatment at all.



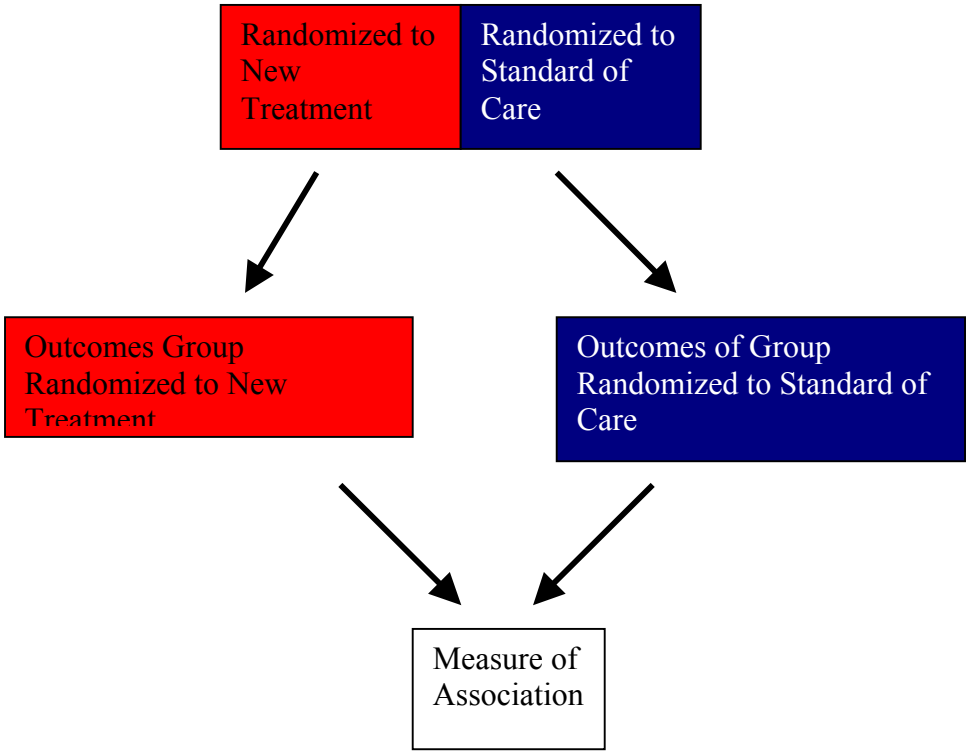
As mentioned previously, the practical problem in measuring the causal effect of a treatment is that only one counterfactual outcome can be observed for any study participant. However, subjects who are randomized to receive the new treatment can be considered a **random sample** of all study subjects. Therefore, their factual outcomes can estimate the counterfactual outcomes for all study subjects, if everyone received the new treatment. Similarly, the subgroup of subjects who are randomized to receive the standard of care can be considered a random sample of all study subjects, and their factual outcomes estimate the counterfactual outcomes of all study participants if none were received the new treatment (i.e. all received the standard of care). This is depicted in the following figure



It follows that a Measure of Association comparing the actual (factual) outcomes in the two comparison groups will estimate the causal effect of the treatment, provided that

1. there is no confounding
2. there are no bias

There should be no confounding for large studies because of randomization, as discussed later in these notes. However, failure to comply with the assigned treatment (**non-compliance**) and biased recording of outcomes are still potential problems that could result in a measure of association not reflecting the causal effect of the treatment.


Randomized to New Treatment
Randomized to Standard of Care
Outcomes Group Randomized to New Treatment
Outcomes of Group Randomized to Standard of Care
Measure of Association

Ethics

Since the investigator assigns treatment, usually at random, in an experimental study, ethical issues are a major concern. If the assignment was based on a random coin flip then approximately half of subjects will receive the new treatment and the other half will receive the standard of care. Therefore, the investigator must have enough confidence in the new treatment to justify giving it to half of the study participants; while at the same time have enough confidence in the new treatment to withhold it from the other half of the patients. This balance opinion about the risk and benefits of a treatment is referred t**o as equipoise.**

Imagine that you have just suffered a myocardial infarction and have been rushed to the emergency room of a nearby hospital. How would you react if you observed a physician reaching into his/her pocket and flipping a coin to make a treatment decision. One would hope that treatment decisions are grounded on solid clinical judgment; however a coin flip is a reasonable action to take when a treating physician is in a state of equipoise.

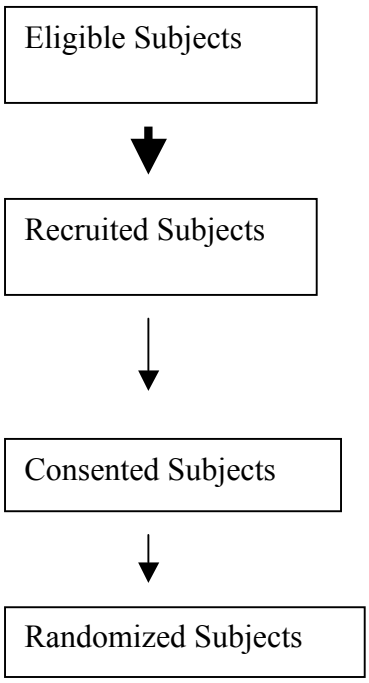
The investigator of an experimental study needs to be in a state of equipoise to justify using a random device to determine treatment assignments. However, others may also need to be in equipoise. For example, if the investigator needs permission from a treating clinician to approach his/her patients to potential enrollment in the study, then the treating clinician must also be in equipoise to grant permission. Often the window of opportunity for performing a RCT is narrow. Once a treatment becomes well accepted as a standard of care, a treating physician may not agree to let his/her patients receive this care.

In addition, Internal Review Boards of a hospital require that an investigator obtain informed consent of each study participant in the study. This involves informing the patient of the benefits and harms of the various treatment options in the study. Hence, the study participants also need to be in a state of equipoise to agree to have equal chances of receiving either treatment option.

Finally, most RCT's are monitored by a Data Safety and Monitoring Board (DSMB). The role of this board is to project the safety of the study participants, Once evidence from the data or external knowledge convince the members of the superiority of one of the treatment options, the members are no longer in a state of equipoise to justify continuing the study.

Efficacy and Effectiveness

Participants in a RCT are often very different for patients who are not eligible for the study. Inclusion and exclusion criteria for recruitment into the study may limit the participants of the study to a small and fraction of eligible subjects. This is described in the following figure



Eligible patients may be all patients diagnosed with a particular disease but only those from a specific set of hospitals might be recruited for the study. Only a portion of the recruited subjects may consent to enroll in the study. Finally, a rule-in phase (discussed later in these notes) may further restrict who might enroll in the study. Each layer of selection may limit the investigator's ability to generalize the results of the study to a wider audience of patients.

In addition the careful monitoring of participants may cause the treatment that patients receive in a clinical trial to be different to what is received outside of the trial, even among patients receiving the standard of care. This means the effect of the treatment in a closely monitored clinical trail (**Efficacy**) might be different from the treatment's effect in actual practice (**Effectiveness**).

Randomization

The main motivation for randomization is to remove the effect of participants and their physician in treatment decision. Random assignment is not influence by patients' characteristics. Therefore, especially in a large randomized trail, the investigator should expect the comparison groups to be comparable with respect to other factors that might influence the risk of developing the outcome. However, for small trials it is possible that some imbalance in the distribution of risk factors may occur and require adjustment in the analysis.

Simple randomization implies that the probability of treatment assignment for one study participant is not influenced by that of another participant. Although randomization is usually determined by a table of random number or a computer generated series of random numbers, simple randomization could be thought of as determined by a coin flip for each participant. The outcome of a coin flip for one participant does not influence the outcome of another participant. An interesting web site for obtaining a series of random assignments under the appearance of coin flips can be found at http://www.random.org/coins/. One limitation of simple randomization, although the statistical expectation if for equal numbers of participants in each comparison group, this might not happen because of the randomness of the coin flip.

An alternative to simple randomization is **block randomization**. Here subjects are assigned to treatment groups in blocks with the condition that equal numbers of participants within each block may be assigned to each treatment group. For example, if the treatments under consideration are labeled A and B, then possible sequences of treatment assignments with a block size of 4 are

AABB ABAB ABBA BBAA BABB BAAB

Treatment assignment begins with a random selection of one of these six possible sequences of treatment. This determines the treatment assignment for the first four participants in the study. Additional treatment assignments are determined by repeating this process. Each time one of the six possible sequence assignments is selected at random to determine the assignments of the next four participants.

One potential drawback of block randomization is the potential to determine the treatment assignment of the next participant who enrolls in a study. For example, knowledge that a block size is four along with knowledge of the treatment assignments of the first three participants in the study completely specifies the treatment assignment of the four subjects in the study. If such knowledge were know to a potential participant, his physician, or the investigator might impact the enrollment of that participant if the treatment assignment was not the one that was desired. This potential problem can be avoided by blinding everyone (except for the person performing the randomization) from the block size or by changing the block size during the enrollment period.

Blocking almost guarantees equal number of participants in each treatment arm in an experimental study. For example, if a block size of four is used, then at the completion of the enrollment process, one of the treatment groups can have at most two more participants than the other treatment group.

Blocking is often used along with stratification to essentially guarantee the similarity of the distribution of a risk factor in the treatment groups. Suppose a study was performed at multiple sites and an investigator wished to have similar distribution of study site within each treatment group to balance the baseline care offered at each group. This can bed accomplished by using a separate block randomization within each site

(stratum). For example the following table demonstrates the use of stratification and blocking in a study performed at three sites.

Study Site	Number of Participants	Selected Assignment Sequence
I	12	ABBA
		ABAB
		BABA
II	8	AABB
		ABBA
III	4	BBAA

The following table displays the distribution of Study Site within each treatment group after the completion of the enrollment period

Study Site	Treatment A	Treatment B
I	6	6
II	4	4
III	2	2
Total	12	12

Blinding

As mentioned above, a measure of association from a RCT estimates the causal effect of a treatment in the absence of confounding and bias. Randomization reduces the potential for confounding, especially in large trials. It also implies that each treatment group is a random sample of the study population, reducing the potential for selection bias at enrollment. However, post randomization selection-bias can still occur due to losses-to-follow-up. In addition, measurement bias can occur in a RCT because of non-compliance with treatment assignment or with outcome detection. **Blinding** is one means for limiting the potential for these biases.

Blinding refers to masking knowledge of the treatment assignment from individuals who might influence the compliance with treatment or the detection with the outcome. Blinding the investigator reduces the potential from having treatment assignment influence the detection and recording of outcomes. The potential for a detection bias is less of a problem for a “hard” outcome like death, but greater for a somewhat subjective outcome like quality of life or performance status. Treatment assignment should not be related to such errors made by the investigator. If any misclassification of outcomes does occur, it is likely to be of the non-differential type.

Blinding the study subjects also eliminates the potential for errors in self-reported outcomes to be related to treatment assignments. In addition, a study subject with a preference for one of the treatment options might be less likely to comply with the assigned treatment is he/she knew that it was not the preferred treatment. This, blinding study subjects may also reduce the potential for non-compliance.

Blinding other physicians who are providing ancillary care to study subjects may also limit the potential prescription of other therapies that might influence compliance of assigned treatment by the study subjects.

In non-clinical trails examining the effect of a preventive agent, the comparison to treatment is often no treatment. In such instances, a **placebo** agent is often used to enhance the blinding of study subjects. A placebo may also be used in a clinical trial, but a standard of care is typically chosen to evaluate the effect of a new treatment because of ethical considerations. A placebo pill resembles the active treatment in appearance but contains no active agents that should influence the outcome. The use of a placebo reduces the potential for knowledge of treatment assignment influencing self-reported outcomes.

Awareness of participating in a RCT may affect the behavior of the study subjects and even influence their outcomes. A series of studies conducted in the 1920s and 1930s at the Hawthorne Plant of the Western Electric Company demonstrated that changes in workplace behavior because of subjects' awareness of participating in a study (Am J Soc 1992;98:451-468). These finding gave rise to the notion of a "Hawthorne Effect", implying that study participants may change behavior during a study and that such changes may influence outcomes that are observed. This is related to the notion of a "placebo effect", which implies that compliance, even with an assigned placebo, may also influence outcomes. Perhaps the strongest evidence to support a placebo effect is found in the results of the Coronary Drug Project (N Engl J Med 1980; 303 -1038-41). This was an RCT comparing the effect of clofibrate (versus a placebo) in the long term treatment of coronary heart disease. Some of the results from this study are shown in the following table.

	5- Year Mortality Risk	
Clofibrate	20.0%	
Compliers		15.0%
Non-Compliers		24.6%
Placebo	20.9%	
Compliers		15.1%
Non-Compliers		28.3%

The first column of data show little overall benefit form Clofibrate (20.0% versus 20.9% mortality risk). On the other hand, the top part of the second column suggests a clear reduction in mortality among compliers of Clofibrate versus non-compliers (15.0% versus 24.6%). However, examining results in the comparison group shows that this reduction is not be attributable to a benefit of Clofibrate but rather reflect the change in the behavior of the participants due to awareness of being in the study and complying

with study medication. The placebo effect is demonstrated by observing a similar reduction in outcome among placebo compliers versus not compliers (15.1% versus 28.3%).

Run-In Phase

Another means to increase assigned treatment compliance is to enroll only participants who are likely to adhere to their assigned treatment. The identification of such subjects might be accomplished during a pre-randomization trial period (**Run-In Phase**), where potential subjects are given a trial period of active or placebo treatment. Non-compliers during this period would be likely to remain non-compliers after randomization, and therefore are not enrolled in the actual trial. The use of a Pun-In Phase will be discussed in more detail later in these notes when referring to the Physicians' Health Study.

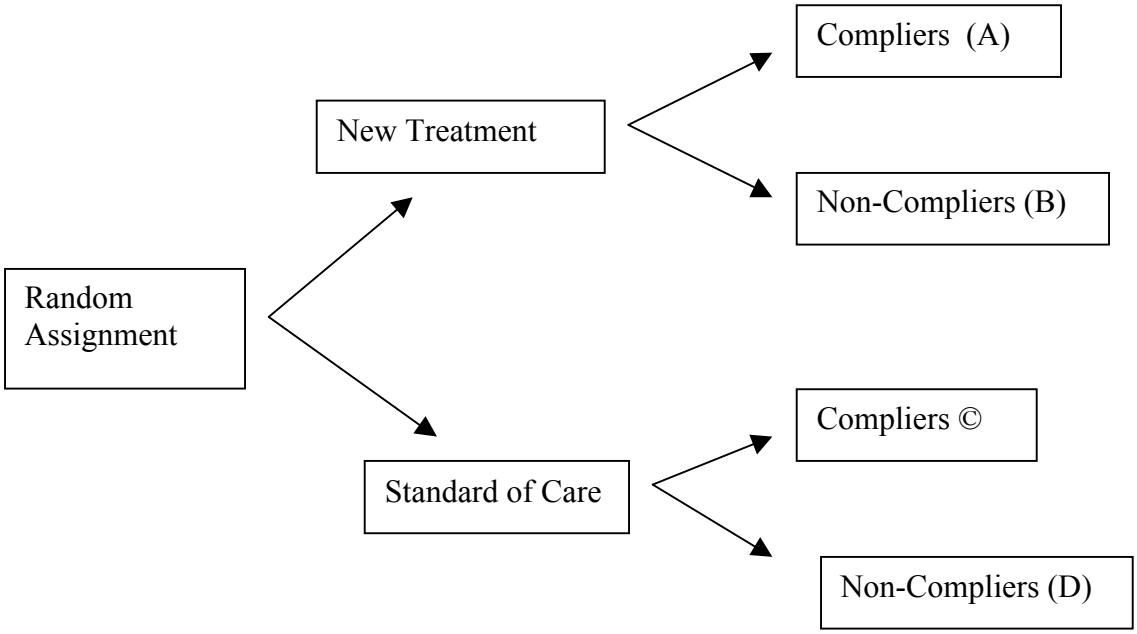
Data Safety Monitoring Board

The role of a Data Safety Monitory Board (DSMB) in a RCT is to monitor and protect the safety of the study participants. This board is usually comprised of outside scientists who are not involved in the study at hand. The DSMB can recommend termination of the study when it is no longer in a state of equipoise and feels that the study participants would be better served with the superior treatment. Possible reasons for terminating the study early include;

1. External evidence of the superiority of one of the treatments
2. Convincing internal information form the data of the superiority of one treatment
3. Severe side effect from one treatment
4. Problems with the operation of the study (including poor recruitment, poor retention of subjects, poor compliance of study subjects, and poor quality of the data
5. Convincing evidence of no likely treatment effect if the study were to continue

Analytic Issues

The following figure displays the potential comparison groups for evaluating the effect of a new treatment in a RCT.



An Intention to Treat (ITT) Analysis compares outcomes for subjects who are randomized to receive a new treatment (A+B) to those who are randomized to receive a standard of care (C+D). This analysis capitalizes on the anticipated effect of randomization to provide two groups that are comparable with respect to other risk factors for the outcome and therefore should be free of confounding (especially in large studies). However, the presence of non-compliers in the treatment group (B), weakens the ability to detect an effect of the new treatment as not all subjects in the group who assigned to the new treatment (A+B) will actually receive the benefit of the new treatment.

An **On-Treatment Analysis** compares outcome only among those who comply with treatment assignment (i.e. compares outcomes in groups A versus C). This analysis might have the advantage over the ITT analysis to detect an effect of the treatment since all subjects in this analysis are compliers. However, this analysis no longer has the advantage of randomization to provide comparison groups that are free of confounding.

Table 1 in a RCT typically displays the distribution of demographic features and potential risk factors for the outcome in the two treatment groups. This table describes the study population and provides a point of reference for the generalizability of the results from the study. More importantly, this table shows the expected comparability of the distribution of risk factors in the two groups from randomization.

Table 2 of a RCT typically displays the overall effect of the treatment on the primary outcome(s). It is usually based on an Intention-to-Treat Analysis. Additional tables may be added to display the effect of treatment on secondary outcomes and the association between treatment and side effects. Finally, results from subgroup analyses,

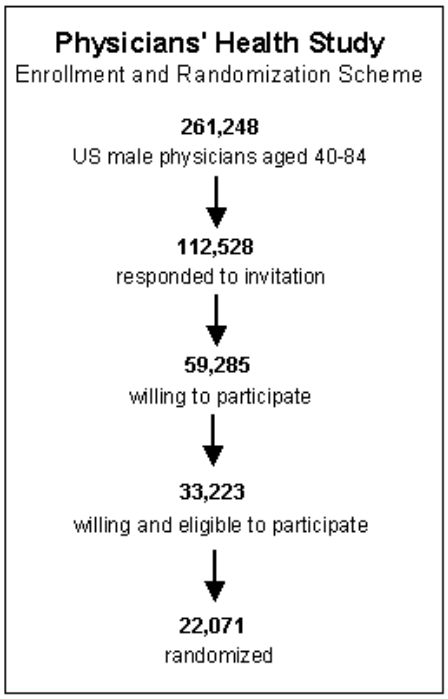
examining the effect of the treatment in subgroups of study subjects are typically performed. One criticism for subgroup analysis is that it may be prone to data dredging (searching exhaustively and reporting subgroups that demonstrate a benefit of the treatment) and problems of interpretation due to multiple testing. Therefore these analyses should be limited to hypotheses that are pre-specified prior to the initiation of the study.

Physicians' Heath Study

The **Physicians' Health Study (PHS)** was a randomized prevention trial with the primary aims to examine if

1. 325mg of aspirin taken every other day reduces cardiovascular disease mortality, and
2. 50 mg of beta carotene taken every other day reduces incidence of cancer

Information about the PHS can be found at http://phs.bwh.harvard.edu . In 1981 the investigators sent invitation letters, consent forms, and enrollment questionnaires to 261,248 male physicians, 40 - 84 years of age, living in the US and registered with the American Medical Association. The following table (provided by Julie Buring) displays the final enrollment numbers in this trial

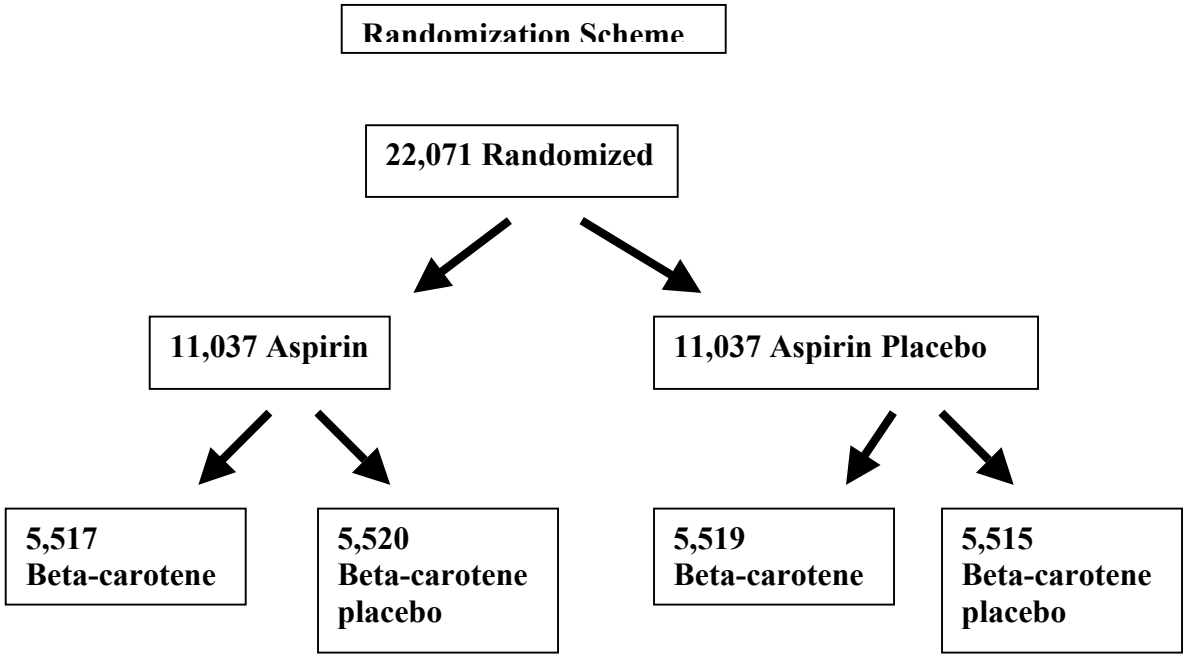


Only 112,528 physicians responded to the invitation, and 59,285 expressed a willingness to participate in the study. However, 26,062 were ineligible because if a variety of

reasons including a history of CVD or cancer; current renal or liver disease; peptic ulcer; gout; or contraindication to or current use of either aspirin or beta-carotene.

The remaining 33,223 were enrolled in a Run-In Phase during which all received active aspirin and placebo beta-carotene. After 18 weeks, participants were sent a questionnaire asking about their health status, side effects, compliance, and willingness to continue in the trial. 11,152 changed their minds, reported a reason for exclusion, or did not reliably take the study pills. This resulted in 22,871 remaining participants for randomization

The PHS uses a 2x2 factorial design whereby each subject underwent two levels of randomization as depicted by the following slide (provided by Julie Buring).



A factorial design allows for the estimation of multiple treatments on the same subjects in a single treatment. The following table displays the comparison groups to examining the two aims of the PHS.

		Treatment Option # 1 :Aspirin		
		Treatment 1	Placebo 1	Total
Treatment Option # 2: Beta-Carotine	Treatement2	A	B	A+B
	Placebo 2	C	D	C+D
	Total	A+C	B+D	Total

The effect of aspirin on CVD mortality can be examined by comparing CVD outcomes among the aspirin users (A+C) versus the aspirin-placebo users (B+D). Similarly the effect of beta-carotine can be examined comparing cancer outcomes among the beta-carotine users (A+B) versus the beta-carotine-placebo users (C+D). A potential problem of factorial design is if the effect of one treatment is modified by the presence or absence of the other treatment (effect modification). For example, if the effect of a treatment is manifested only among subjects given placebo for other treatment then the study may have reduced power for detecting this effect. The topic of Effect Modification will be discussed in a future lecture.

Participants were sent pill packets each month and instructed to take one pill each day. On odd days of the month participants would take either aspirin or aspirin-placebo. On even days of the month, participants would take either beta-carotine or beta-carotine-placebo. The following figure (provided by Julie Buring) displays a picture of the pill packets used in the PHS.

As expected in large RCT, the distribution of demographic features and risk factors were very similar in the comparison groups. This is demonstrated by the series of tables (provided by Julie Buring)

Assignment	**ASA + BC (n = 5,517)**	**ASA + PL (n = 5,520)**	**PL + BC (n = 5,519)**	**PL + PL (n = 5,515)**
Demographics				
Mean age (years)	53.2±9.5	53.3±9.6	53.2±9.5	53.2±9.5
Geographic region				
Northeast	27.8	27.9	27.3	28.8
Midwest	22.4	21.9	22.1	20.3
South or Southwest	26.8	26.1	27.0	27.9
Mountain or Far West	22.0	23.3	22.8	22.1
Other	1.0	0.9	0.7	0.9
Medical history				
Reported hypertension	11.8	11.8	12.4	11.5
Mean systolic BP	126.0±11.9	126.3±12.1	126.1±11.7	126.1±11.7
Mean diastolic BP	78.8±7.6	78.9±7.5	78.9±7.5	78.7±7.5
High cholesterol	7.9	7.5	8.0	8.0
Mean cholesterol (mg/dl)	211.7±43.9	212.3±45.0	212.7±46.5	211.4±44.8
Diabetes mellitus	2.1	2.5	2.2	2.2
Mean BMI	24.9±3.1	24.9±3.0	24.9±3.0	24.9±3.0
Overweight	13.5	14.0	13.6	13.6
History of angina pectoris	1.2	1.4	1.1	1.2
Parental history of Ml	12.7	13.3	12.7	13.6

Health habits				
Cigarette smoking				
Never	49.8	48.8	49.6	50.1
Past only	39.3	40.1	39.1	39.0
Current	10.9	11.1	11.4	10.9
Mean cigarettes/day	21.7±13.0	22.1±12.9	22.4±13.7	22.8±13.5
<15 (of current smokers)	28.8	27.4	27.3	24.6
15-24 (of current smokers)	32.5	33.7	34.7	35.4
25+ (of current smokers)	38.7	38.8	37.9	40.0
Current cigar smoking	4.7	4.2	4.3	4.3
Current pipe smoking	8.8	7.7	8.2	8.0

The DSMB stopped aspirin arm of the PHS in 1989 (ahead of schedule) because of clear effect of aspirin decreasing the risk of Myocardial Infarction (heart attacks). However, too few strokes or deaths occurred upon which to base sound clinical judgment regarding aspirin and stroke or mortality. In 1996 beta-carotene arm stopped showing neither benefit nor harm