TITLE: RANDOMIZED CONTROLLED TRIALS (RCT)

PRESENTER: Julie E. Buring, ScD

Slide 1: This is the last of our three lectures on the design and the conduct of studies in epidemiology. And in particular we are going to talk about randomized control trials.

Slide 2: Reviewing again, we said that we want to know where we are in our knowledge on a particular question at particular time, we look at the totality of evidence.

The totality of evidence is going to be made up by basic research studies which give us an idea of why, the mechanism, by which an exposure could affect an outcome.

But epidemiologic studies will give us direct evidence in human. And we characterize them as the descriptive studies, which allow us to look at who is getting the disease, what disease or outcome are they getting, where are they getting it, when are they getting it?

And by doing that raise hypothesis that will be tested by the analytic studies of why. The analytics studies are broken into two groups, observational studies, case control and cohort and intervention studies, the randomized clinical trial.

The observational studies simply record whether people have the exposure or don't have the exposure, develop the outcome or don't develop the outcome, but the exposure is self selected by the participants and the investigator in no way allocate them to the exposures, it's just we are observing what's happening.

In an intervention study randomized clinical trial, again, you haven't exposed non-exposed group, but the investigator allocates the exposure making it the closest to a basic research study that we have in epidemiology.

Slide 3: So again, observational studies, the exposures are self selected. Intervention studies, the exposures are allocated by the investigators, they are not self selected.

Slide 4: So, diagrammatically again, we have the exposed group and the non-exposed group, we follow them and forward it to the development of the outcome, but in an intervention study the investigators

right there at the beginning of the study, allocating the exposure to one group and allocating the other group to have a comparison of some kind.

Slide 5: Let's go back now and think about the observational studies that we just discussed, the case control and the cohort studies and see how they differ inherently from a randomized trial in ways that will greatly affect the validity and the findings and the interpretation of the findings.

Let's say we just wanted to do a study, to look at whether postmenopausal hormone therapy affects risk of cardiovascular disease. In an observational study we would take women free of disease at baseline, classify them as users or non-users of hormone therapy, follow them over time and compare who does and does not develop cardiovascular disease among the exposed group and the non-exposed groups.

Slide 6: And if we do a meta-analysis and look at these findings, there have actually been 40 observational studies of hormone therapy and coronary heart disease. And if you look at the data that are up on the slide, those who ever use hormone therapy, had a 36% lower risk of coronary heart disease and those who are current users of hormone therapy had a 50% lower risk of coronary heart disease.

This data is extremely consistent as you can see from the diagram that's on the right hand side of the slide. So, why isn't that enough, why isn't that enough for us to recommend that if you take hormone therapy you will see the same benefit on coronary heart disease that these women experienced?

Why is it that we are hesitant to say that it is the hormone itself that is responsible for the reduced risk of coronary heart disease and not the lifestyle or the characteristics of the women who were taking the hormone therapy?

Slide 7: Well, the problem or the concern is that women who take hormones for an extended period of time, so women who self select to be exposed would differ from those who don't, in many ways it could be related to the outcome of interest.

Why are they taking the hormones? In part, certainly it could be for the symptoms of menopause.

But it also could be that they rather we're told that hormone therapy will reduce the risk of heart attack, might keep them young, keep their cognitive function high. And because they are taking the hormones for those reasons to prevent the diseases of aging, they might say, and let me do everything else that I can to prevent the diseases of aging. Let me also eat a better diet; let me have a lower percentage calorie as fat. Then you will have a lower body mass index.

Then they will say, while I am eating better maybe I should be exercising more. And maybe, you know I was trying to stop smoking all this time, but haven't been able to, let me stop smoking now also.

And in fact that is what we found in the observational studies, when we compared estrogen users to non-estrogen users, the users were leaner, less likely to smoke, more physically active, more likely to see their doctors and more educated.

And any of those factors because they are related to coronary heart disease could explain the lower rates of coronary heart disease that we have seen among the hormone users, rather than the nonusers.

It may not be that the hormones reduced the risk of coronary heart disease, but the other lifestyle characteristics that are associated with hormone use, were the ones that were responsible for the decreased risk.

Slide 8: In a clinical trial then we're going to try to get the two groups the same with respect to every other characteristics that's related to the outcome understudy, except for the exposure that we are looking at, the intervention that we are looking at.

So in a randomized placebo-controlled clinical trial, participants who are eligible would be randomly assigned to the hormone use or to not use hormones to take a placebo.

Followed over time to see how many developed the disease, one group compared to the other group. And basically, because we're doing this by the luck of the draw, by a random number table every characteristic of the people in our populations are going to be evenly divided between the hormone used group and the placebo group if the sample size is big enough.

So randomization with a large sample size will ensure that the hormone group will be similar to the placebo group in lifestyle factors, medical and family history, and other factors.

Slide 9: When we talk about randomized clinical trials we are going to have a lot of words that people do, characteristics of trials, such as an emphasis on large sample size, the use the randomization of blinding and double-blinding, a placebo-control.

Focus on high compliance; focus on low loss to-follow-up and analysis, which is a particular kind of analysis called Intention-To-Treat. And all of these trial features are designed to minimize the alternative explanations of chance, bias and confounding.

And it's important to realize, now we need to think about what the specific niche is that a randomized trial plays in all of the designs that we have available for us to use.

It wouldn't be necessary to be spending so much time carefully designing a trial if we were talking about effect sizes that were large. If we are talking about metformin, if we are talking about penicillin, where mortality rates went from 85% to 15% when the drug was introduced.

We actually could do this with an observational study. We could actually do it without a comparison group, the difference is so extreme that's not what we are usually doing. We would be very pleased if a drug made a 10%, 20%, 30% difference between the mortality versus usual care plus as new drug.

And 10% differences, 20% differences, 30% differences are very statistically difficult to assess and is very, very meaningful to us on a clinical or public health standpoint.

So the unique niche of randomized trials is that they are optimal to detect statistically small to moderate, but clinically worthwhile treatment effects, because of everything we do to reduce chance, bias or confounding, the noise that is in the study that would mask the magnitude of the effect that we're trying to show.

Slide 10: The bottom line, I will repeat this at the end, but it's important to realize at upfront as we begin to talk about trials. They are more logistically difficult, more expensive and have more issues related to ethical considerations than other epidemiologic design strategy.

But if it's ethically appropriate to do a trial and if it's well designed and conducted using the characteristics we just mentioned before, they will actually provide a level of assurance about the effect of the intervention itself on the outcome that just cannot be achieved by any other epidemiologic design strategy.

Slide 11: We use some terms for trials, if you're doing a treatment trial, meaning that the person already has clinical evidence of the disease and you are trying to reduce risk or the recurrence of the disease, dying from the disease, sequela from the disease, that's called a treatment trial or a secondary prevention trial.

If on the other hand you have healthy people with no clinical evidence of the disease, they can certainly have risk factors, they can be at high risk, but they haven't been diagnosed with the disease yet. Then those are called prevention trials and we are doing primary prevention.

You can also randomize individuals or you can randomize groups. The majority of the trials will always be randomizing in the individual, to one group or the other. But there will be circumstances like you are trying to make a difference in a school, where you can't randomize individual children in the fourth grade classroom, you really have to take that fourth-grade classroom and another fourth-grade classroom and randomize those two classrooms to get the intervention, because the children will talk to each other and it will be very hard to keep one exposed and one non-exposed.

Slide 12: There are some special issues and trials that we always must keep in mind. There is no question that it will cost more and has more feasibility issues, than an observational study, because you are intervening in someone's life, you are actually giving them the intervention, whether it be a drug, whether it be a program, whether it will be a lifestyle change, even something as simple as changing diet. You have to work with the person, get them to understand what to do, give them booster sessions to keep it over time and that's going to be much more costly than just simply observing what people do.

The feasibility aspect is you need people who are willing to be randomized in the trial, to getting drug A or drug B, to getting a radical mastectomy versus lumpectomy for their breast cancer.

For getting usual protease inhibitor cocktail, versus one that has an additional drug in it. They have to be willing to do that, they have to be willing to be randomized to the two groups and in many cases they won't know whether they are taking the active drug or not taking active drug.

So that's very different from an observational study where the people self select it what they were going to take.

The third thing that we must always keep in mind when doing trials or issues of ethics and the issue of ethics is something called Equipoise. Equipoise means the following:

If you're doing a trial of aspirin to prevent heart disease, very simple question, I take aspirin on a regular basis, well; I reduce my risk of developing my first heart attack.

At the moment that that trial begins, there have to be enough belief that aspirin will be beneficial in the prevention of heart disease to justify giving it to one half of the participants in the study.

And at the exact same time there have to be enough doubt that the risks to benefit ratio for aspirin in the prevention of heart disease is going to come out in a beneficial way to justify withholding it from half of the participant.

So Equipoise says that the time that you start the study and you are deciding whether you will put a patient into a trial or the patient is deciding whether they will go into the trial, we have to be able to say, I really don't know what the best thing for you is. There are advantages and disadvantages of this; there are advantages and disadvantages of that. I do not know which one is best; the trial will answer that question.

We all know we can't randomize to demonstrate harm, you can never randomize smoking to participants to show the adverse effects of smoking, but you can randomize the stopping or cessation of smoking, two different and intensive ways to get people to stop smoking, versus the usual educational campaign that is done.

So you just have to think again at all times of the ethics of doing the study, and the questions will be, is doing a trial ethical? That's what we just discussed.

But then the alternative question which is just as important is not doing a trial ethical. When you know that people are already using a drug or an intervention or a vitamin or a supplement with the idea that it will make a difference on their subsequent diseases and outcomes, but we don't know it and we will never know it, unless we do a true trial and definitively evaluate it.

And the key issue of timing of trials is really timing, that window of opportunity, when people are going to be willing to be randomized, before it just becomes so common in the general public community or considered to be standard care or usual care in the physician community even though there is no evidence to demonstrate efficacy.

Slide 13: I am going to talk now about the design of trials and I am just going to anchor it to two studies as examples. Physicians health study, randomized trial of low-dose aspirin, 325 mg every other day and beta-carotene, a vegetable form of vitamin A, 50 mg every other day to the equivalent of five servings of fruits and vegetables.

In the primary prevention of cardiovascular disease and cancer among 22,071 US male physicians between the ages of 40 and 84, trial was funded by the NIH, drugs were provided by industry.

Slide 14: We also did a complimentary trial in women. Women's health study, randomized trial, lower-dose aspirin, 100mg every other day and vitamin E 600 international units every other day, in the primary prevention of cardiovascular disease and cancer among approximately 40,000 US female health professionals over the age of 45.

Slide 15: Now before you ever can start a trial, preliminary work piloting if so key. In the example that I just gave you of let's say aspirin, let's focus on that for now. We had to do a bioavailability study first, to confirm that the proposed doses of the aspirin that we were going to use, 325 every other day and a 100

mg every other day are adequate to give us the mechanism that we postulated will be responsible for the reduction in risk of cardiovascular disease.

So, adequate irreversibly inhibits platelets, in both men at 325 every other day, and women at 100 mg every other day. We also needed a pilot study just to show that coming up with these sample sizes, 22,000 in men, approximately 40,000 in women is feasible.

So rather than just claiming in a grant application or to anybody that we will be able to deliver that kind sample size, we have to show the feasibility of doing so, so we get a pilot study in thousand physicians and we figured out that we could identify them, we could mail to them, we now know what their response rates are, we know what percentage are going to be eligible, what percentage said they were willing to be in the study, the percentage that was willing to give a blood specimen.

So, by doing that, we have reassured ourselves, as well as funding organizations that we need that sample size to do the study and we can deliver on the sample size. We also needed to consider regulatory issues, and in particular, aspirin is an established drug, but not for this indication, so we needed an investigational new drug application to be submitted to the Food & Drug Administration in the United States, focus to be able to start a trial of aspirin in the primary prevention of cardiovascular disease and cancer.

Slide 16: All trials have a population hierarchy that underlines all of the different groups that we are going to see. The first group is the reference population. The population to whom the results are going to be generalizable. In this particular case of aspirin in cardiovascular disease, we wanted to do a primary prevention trial. So the reference population is everybody in the world who has not been diagnosed with cardiovascular disease. But we can't do a study in everybody in the world who has not been diagnosed with cardiovascular disease; we can’t control that kind of study.

We can’t identify everybody, we can’t be giving them their agents, we can't be encouraging them to stay compliant, we can't be getting information from them on a reliable basis, it’s just too much spread, too many places.

So, we are going to actually have to pick a population in which the study is going to be done. A population in which we believe we can get a valid result; that we can identify them, they will be compliant, will be able to follow them up over time.

And that is related now to getting a population in whom the results will be valid. So the reference population is to whom do we generalize, the experimental population is the one in which we get a valid result, and always remember that we can only generalize a valid result. If we don't get a valid result, then talking about generalizability makes no sense.

So we always keep in mind the reference population when we are choosing our experimental population, but the validity of that population will be more important than generalizing to the reference population.

Slide 17: So the next thing that we have to do is get the participants into the study, in particular we take our experimental population and we invite them to be part of the study, we asses whether they are eligible to be part of the study and then whether they are willing.

We give them an informed consent which describes the state of knowledge on this question at this point in time. What we are trying to find out in our study, what will be required of them if they are part of the study; that they are free to not be part of study and their medical care will not be compromised in any way. And then if any new knowledge comes up during the study that changes equipoise; that changes the current state of our knowledge; we will inform them of that.

Slide 18: And the important thing to know is really how difficult it is to get participants in to the study. I have heard so many people say, I have absolutely no problem getting my participants into the study, I see maybe a hundred good people in my clinic on a monthly basis, I am sure 95 of them would be willing and eligible to be in the study.

It just doesn't work like that, it's orders of magnitude down to get people who are eligible, because the criteria for eligibility are more strict than your usual treatment criteria and willing to be in the study.

And just to give you an example of that in the physicians' health study, we actually mailed the letter of invitation to over 250,000 US male physicians who are registered with the American Medical Association, and just look at these numbers by orders of magnitude, about half or 112,000 returned their questionnaire, about half or 60,000 were willing to participate in the study, and a little more than half or 33,000 were eligible and enrolled in a 12 week run in period like a practice period to see if could be compliant with what we were asking them to do in the study.

Slide 19: So now we have our participants and we need to allocate them.

Slide 20: And the optimal way of allocation of the study regimen is randomization, where the treatment group is allocated at random, by computerized random number generator and that random number generator can translate into a phone call, you can translate into opaque envelopes that are put in surgery, so you just open one after you assess the patient's condition. It can be done however it is needed to do to get you that information.

Slide 21: The strength of randomization is its unpredictability, so you cannot anticipate what the next patient that you see is going to be part of this study, what group that they would be in.

Therefore, neither the participants nor you are selecting in any way the type of patients who is getting treatment A versus treatment B, because you have no way to predict what they would be getting.

The very special thing about randomization is what it does to confounding, that all confounders known and unknown are in average distributed equally among the study groups. And remember the word 'on average' and that's why we keep on saying, if you are going to do a trial, do it big.

Because you cannot say that if you do a trial of 24 people; 12 in treatment A, 12 in treatment B and all the other characteristics of the participants in the study will be evenly divided between the two groups of 12, we know that's not true.

But if the sample size is bigger, and the bigger it is, the more that the underlying distribution will be such that everything is evenly divided between the two groups. And not only the confounders we know, but the confounders that are unmeasured or even unmeasurable. So randomization minimizes selection bias and confounding.

Slide 22: And to just show you, this is usually the first table of the results of a randomized trial where we take the two treatment groups aspirin, placebo in this case, and we just compare the known confounders, the known variables that could be related to the outcome of cardiovascular disease.

And if you just look down these two columns of numbers while I read the variable, you can see that age is evenly divided, history of hypertension, systolic blood pressure, diastolic, history of high cholesterol, cholesterol level, history of diabetes history of angina, history of post myocardial **infarction.**

Slide 23: Current smoking, passive smoking, daily alcohol, exercise more than once a week, body mass index and multivitamin use. Now, that's wonderful that those known confounders are evenly divided between the two groups. If they weren't, then we would have to control for them ourselves in the analysis. But the fact that they are evenly divided for the known confounders, gives us very reassuring evidence that the variables that we didn't get information on, the unknown, unmeasured, unmeasurable confounding factors, are also evenly distributed between these two groups.

And that's because of the size of the sample and the fact that the known confounders are evenly distributed.

Slide 24: There is one more kind of allocation and that is something called a Crossover Trial. It's a before-after trial where individuals actually serve as their own historical control.

The advantage is you never have to worry about confounding with that, because the person in their own control; there is perfect matching.

But on the other hand there are some special characteristics. It actually requires that the individual's condition can change over time, because this is being done sequentially. So, you do one drug and then you do another drug, or one kind of cream and another kind of cream and it can't be that the condition itself is changing during the period of time.

And you also have to be able to have an adequate washout period. So you have to make sure that the first drug that you give doesn't have irreversible outcomes or long-lasting effects, which means that when the second drug is given, you are seeing the results of the drug plus the long-lasting effects of the first drug.

Slide 25: Once you get the people who are willing and eligible to be in the study and you are going to randomize them, one design that you can use is something called a Factorial Design. In our case, a two by two factorial design, because we are randomizing to two drugs.

We take the 22,071 US male physicians who are to be randomized into the trial; we allocate them first to active aspirin and aspirin placebo and then we randomize them one more time to beta-carotene and beta-carotene placebo.

So we actually have four groups in the study. One group has both activated aspirin and beta-carotene, one group only active aspirin, one group only active beta-carotene and one group both placebo.

And this design is very, very efficient if the two agents do not interact with each other, aspirin and beta-carotene were not hypothesized to be synergistic in their effect.

But if in fact you do have two agents, which might be synergistic in terms of either benefit or harm; this is an opportunity we begin to look at someone taking the two agents together, versus only taking one at a time.

And begin to raise hypotheses about whether there is a difference in the two agents, versus one alone.

Slide 26: The nature of the comparison group had to be carefully brought through in every single trial. It cannot be less than standard of care; nobody can suffer, because they are part of a randomized trial and be denied something that people getting usual care who are not in the trial would have access to.

So often the comparison group is actual usual care. But it can also be usual care plus other things. So it can be looking at other doses of the same treatment. It can be other treatments or it could be that you're giving a treatment and a placebo to that treatment, where a placebo is an inert agent that looks indistinguishable from the active agent.

Slide 27: And the use of the placebo is designed to minimize observation bias. And the people who are part of the study, if they know that they are getting a new drug and they know that new drug could have particular side effects, every symptom that they feel during the trial is going to be interpreted by them to the lens of knowing what drug that they are on.

And so, if we could get everybody to just take a drug and half of them are getting an active agent and half a placebo, but they don't know what they're getting, then they are just going to report everything to us and we can compare it in the two groups and minimize bias in terms of the reporting to us as investigators.

The need for a placebo depends very much on the subjectivity of the outcome. If you're taking only a hard outcome of mortality, you are not going to need a placebo, but even if you're doing cause of death; that could be influenced by your knowledge of whether the person had surgery or medical therapy or drug A versus drug B.

And if you are studying any outcome where the measurement of it is subjective, quality-of-life, grip strength, if you're doing an arthritis drug, feeling better, having more energy, getting out more, those are really need to be thought of how do we minimize bias in those ascertainment.

So if we can do a placebo that would be one thing to do. But a placebo or blinding, so that the person doesn't know whether they're getting activated into placebo, may not be practicable or cannot be done in some situations.

Obviously, if you are doing medical versus surgical, you can't do a placebo or drugs with characteristic effects, you can't come up with a placebo that has that effect.

Slide 28: But you keep on trying to build in the minimizing bias everywhere you can. So if you can blind the participants, then single blind, you need to do it. If you can do the participant and the investigator double-blind, then you need to try to do that.

By the way, blinding is also referred to as masked or double-masked in some trials. But you can also blind in the assessment of the outcome. So if your outcome is a blood pressure measurement for example, there is no reason why the person who is taking the blood pressure medicine is going to the blood-pressure protocol. We would need to know whether the person was on treatment A or treatment B.

So even if for some reason you can't blind the participant or the investigator, you might be able to blind the person who is assessing the outcome. It can add credibility to the study, but also complexity and cost.

And no matter what, try to use objective criteria determining the outcome. And as we said before, this will be especially important if you cannot use blinding or placebo.

Slide 29: After you allocate the participants to the active intervention group or the comparison group, now we need to see whether the people are actually doing what they're supposed to be doing. Are they complying or non-complying with the intervention and with the comparison group?

Slide 30: Compliance is absolutely crucial to the ability of the trial to demonstrate a true effect. Noncompliance will do one thing; it will bias the relative risk towards the null value.

It will make it that you could actually miss the effect of the drug that is truly there, because not enough people were taking the drug.

So we have to ascertain compliance, we have to figure out whether the people did comply with your intervention or not, and during the conduct of the trial, the methods to maintain high compliance are absolutely critical. What we need to do is make it so that allocating people to the regimen is equal to their taking the regimen. And whatever we need to do to that will be worth the cost.

Slide 31: And to just show you as an example, we have healthy people in our study; we are trying to prevent their first heart attack. Healthy people are not great at taking medications; even sick people are not good about complying with the taking of medications, but it’s even worse for healthy people.

So we came up with a calendar pack, a little bit based on the oral contraceptive pack, where each pill was in a little blister packages, you just press it right out aluminum foil on the back, this is a monthly supply. The date of the month is next to each of the pills, so you can look at the date, look at your pill pack, figure whether you have taken it or not.

And pilot studies have shown that this helped in people complying with the taking of a regular pill, especially healthy people in primary prevention. But these aren’t cheap; these were a $1 to $1.50 per calendar pack. We had 20,000 people in one of our studies, 40,000 in another study, the study went on five years in the physician’s health study for aspirin.

It went on for 10 years on women's health study for aspirin. So you can see how much of the budget was actually involved with packaging the pills to give to people.

But on the other hand, this is the part where people do not comply with the taking of the intervention. You can have the most beautiful idea, scientifically worthwhile, perfectly designed, but if the intervention does not have good compliance, you will not be able to show the effect of the intervention.

Slide 32: And finally then we need to figure out the outcomes in all of the groups.

Slide 33: And how we ascertain those outcomes, really depends on how the study is being done. We happen to do it as self report, the participants self reported the trial in point. Then we got their permission, we got the relevant medical records, we had those reviewed by the endpoints committee of physicians who were blinded to the randomized treatment assignments, so there is a place we can put in that triple blind for the outcome assessors.

We used pre-specified objective criteria and only confirmed endpoints are included in our final analysis. But for other people it will be bringing people in to the clinic, it will be going to their home and collecting information, whatever it is, it has to be done the same, between those who are allocated to the active intervention and those allocated to the comparison group.

Slide 34: Now, tying back again to affects of doing the study. If the participants don't know what they're taking, the investigators don't know what they're taking, those assessing, the outcomes don't know what they're taking, somebody absolutely has got to be following the participants with respect to safety and making sure that nothing has changed in the equipoise we believed within place when they signed their informed consent.

So this is done by a group called the Data Safety Monitoring Board. It's an independent group with expertise in various disciplines, but the key word is independent. They have to be scientifically and financially independent from the trial that they are monitoring.

It doesn't mean that they don't care; it doesn't mean that they aren't interested and hope that aspirin would reduce the risk of cardiovascular disease. But their scientific careers cannot be based on whether aspirin does or does not reduce risk of cardiovascular disease, their financial lives will be better if it does or does not work, and so they can just wear the hats of protecting the participants.

Their charge is to safeguard the participants, protecting them from something unexpected; from unexpected harm were unexpected benefits that had not been communicated to them at the beginning of the study or change in the equipoise and basically, ensuring the integrity of the trial.

Making sure it is along at the speed and the rate it is supposed to and the data that are coming in are of high quality. They will review the progress of the trial and they will review the unblinded data on the outcomes.

They will consider early stopping roles, either because there is unexpected benefit, unexpected harm or futility, either because the data are not able to be obtained the way we hoped they were, there were not enough participants or the results were showing that there are just no differences between the groups and it is unlikely that by the end of the study that will change.

They can recommend modification or termination of the trial, based on the information from the trial, from other trials that are ongoing or new basic science information. The Data and Safety Monitoring Board is required of all phase three trials funded by the National Institutes of Health. And a data and safety monitoring plan is required for all trial Phase 1, Phase 2.

Slide 35: The analysis of a trial is very basic, very similar to that of a cohort study; we compare the rate of the outcome of the treated group, versus the comparison group, to the exposed group to the non-exposed group.

But the first table is going to be ascertain, if randomization worked. In other words, are the treatment groups comparable with respect to the baseline characteristics, these potential confounders?

If they are not, then we need to control known confounders in the analysis. But we cannot use the strength of randomization to say that now the unknown confounders are evenly distributed.

Slide 36: The analysis is a little different in a trial, than in a cohort study. In that it is called Intention To Treat. We have our study participants; we have randomized them to the active treatment groups versus the comparison group.

And now we have sub-groups, we have compliers versus non-compliers, then those who get the outcomes versus, who don't get the outcome, in both the active group and the comparison group.

The primary analysis is an Intention To Treat Analysis, which means once randomized, always analyzed. It doesn't matter whether the participant complied with the taking of the intervention or not, they are always going to be included in the group to which they were randomized.

So the active group will be compared to the comparison group for all outcomes that happened to those randomized to the active group, versus all outcomes to those randomized to the comparison group, regardless of whether that participant complied or didn't comply with the intervention.

And the reason behind this is when we randomized to the active group versus the comparison group; that is when the confounders are evenly divided between the two groups.

Anytime you go down and take a sub group, like compliers versus non-compliers, those who comply with the intervention are going to be systematically different, than those who don't comply, in ways that may affect the outcome understudy.

So it is only at the very first level of active group versus comparison group, where we can invoke randomization to have evenly distributed the known and unknown confounding factors between the groups.

The intention to treat analysis with the one with the power of randomization, and it is an analysis required by regulatory bodies such as the Food & Drug Administration of the United States.

Slide 37: On the other hand almost everybody will then go on and look at a secondary analysis which is compliers only, where we take those in the active group who complied with the intervention and compared those to the comparison group who complied with the comparison group.

Just be aware, this is not a randomized comparison, you have to control confounders in the analysis yourself. And the problem is we don't always understand whether the characteristics of those who comply versus those who do not comply.

Slide 38: Just to give you an example of this, this is an older study from 1980, this is used as an example ever more of understanding that subgroups are different than the overall population.

This is a coronary drug project in which a lipid lowering drug clofibrate was examined in terms of reduction and mortality post myocardial infarction; sort of the beginning of looking at lipid lowering, well before statins.

And they first compared the two groups, the clofibrate group and the placebo group with respect to five-year mortality. And the mortality was basically the same in the two groups, 18% in clofibrate 19.5% placebo.

With the conclusion then, the clofibrate did not reduce mortality, post myocardial infarction.

Well, the investigators or the clinicians who used that drug in practice said, well, that's not fair to include everybody in the group, even those who didn't take it.

Because clofibrate was not an easy drug to take, that's why a lot of people became noncompliant over time. They said, please let's just look in the clofibrate group, those who did or did not take their clofibrate and see what happened to them.

So, that's the first column under clofibrate, and what they found, pleased those investigators tremendously. Because they found that if you complied with the taking of your clofibrate more than 80% of the time, you had 15% mortality, and if your compliance was bad, if it was less than 80%, then in fact you had 25% mortality.

So if you took your drug you had a lower mortality rate. But then what would be wrong with just looking at clofibrate group, you have to look at what happened to the placebo group also.

And in the placebo group, if you took your placebo more than 80% of the time, you had a 15% mortality rate. And if you weren't good about taking your placebo compliance less than 80%, you had a 28.2% mortality rate.

So you had a benefit if you are a good complier on mortality, regardless of whether you are compliant with your clofibrate or compliant with your placebo. It wasn't being compliant with the right agent; it was just being compliant with all the characteristics of people who stick with a regimen.

And those other characteristics might in fact be related or were in fact related to the outcome understudy.

Slide 39: Another point in the analysis that we need to always remember is that an effective sample size of a trial for actually any study is not the number of participants, but the number of endpoints those participants experience.

So you can have a massively big sample size in your study, but if you pick the population where they don't have the outcomes, then you are not going to be able to answer your study.

But if you have a smaller number of participants that are very high risk of getting the outcome, you would be better able to answer your questions.

Slide 40: So just to give you an example from the physician's health study, we looked at myocardial infarction. We found 139 myocardial infarctions in the aspirin group and 239 in the placebo group when the trial was ended by the Data Safety and Monitoring Board.

At the relative risk of .56 a 44% reduction in risk, statistically significant. So with our sample size, we had an adequate number of events of myocardial infarction, to find a statistically significant difference between the groups.

But there was one other endpoint that we really cared about which was the main risk of the taking of aspirin. One of the main risks for the taking of aspirin is an increased risk of bleeding, and the most severe form of that would be the hemorrhagic stroke.

So if we're going to do a risk to benefit analysis, we would certainly want to understand not just the benefit of myocardial infarction, but the possible adverse effect on hemorrhagic stroke.

And for the number of participants that we had in the study and the length of time we followed them up, hemorrhagic strokes are simply not as frequent as myocardial infarctions are.

So for that period of time, whereas, we would have almost 400 myocardial infarctions, we only had 35 hemorrhagic strokes, which gave us a relative risk estimate, which is over two folds, 2.14, that was not statistically significant, because we didn't have enough events to definitively answer that question.

Slide 41: And in fact for us to be able to answer the question in men versus women, we actually needed a massively different sample size in the two groups, because women at any given age have a lower risk of the outcome of cardiovascular disease than a man does, but they have the same risk of developing the side effects from the taking of aspirin.

So we needed enough myocardial infarctions and strokes in women, for us to be able to answer the benefit question. And what we could accomplish in 22,000 men required 40,000 women simply that did have a lower rate of the outcome. And even though the cost per participant was the same in men and women, without any question, we had a more expensive trial in women, only because the sample size was bigger.

So all of this has to be looked at when you think, I'd really like to answer this question overall, but then I want a particular group, I want to look at people of a certain age, I want to look at a certain ethnic group, I want to look in a certain gender.

Every time you do that to definitively answer the question, you are going to have to pick, you are going to have to evaluate to figure out what your power must be to have an adequate sample size and number of the events in each of the groups that you want to compare.

Slide 42: So in summary, on the intervention study, it's a type of cohort study in which the exposure is allocated by the investigator. The strength of the intervention studies of the trial is that they are optimal for detecting small to moderate side effects.

You have the greatest degree of control over the exposure. If you would use randomization, then you're minimizing your selection bias and confounding, and the confounding that's being minimized is both known and unknown factors.

If you use placebo or blinding or objective outcome definition then you will minimize your observation bias. And the limitations are ethics that is more expensive and harder to do, that you have to keep people compliant with the intervention and you have to minimize losses to follow up.

And like a cohort study, there was a fundamental trade-off of internal validity with external validity or generalizability. Without any question needing internal validity first, but always being aware, did I pick a population that's really going to limit me in terms of generalizing beyond my particular study?

Slide 43: So back to my bottom line slide, the trials are more logistically difficult, more expensive and have more issues related to ethical considerations, than any other epidemiological design strategy.

Recruitment can be harder than anticipated and the event rates are often lower than expected. But if ethically appropriate and well designed and conducted, trials provide a degree of assurance about the effect of the intervention itself on the outcome; that just cannot be achieved by any other epidemiological design strategy.

Slide 44: And randomized trials really serve as a theoretical goal standard when we design our observational studies.

Slide 45: Thank you!