David Murray:
Hello, my name is David Murray. I'm the NIH Associate Director for Prevention. I also direct the Office of Disease Prevention here at NIH. I want to welcome you to part one of the course on Pragmatic and Group Randomized Trials in Public Health and Medicine. Part one will cover the introduction and provide an overview for the material. This is the first part of a free seven-part, self-paced online course that NIH will provide. We also provide the slides that go with each of the modules, the readings that we'll cite during the course, and we provide guided activities for each of the component parts.

The target audience for this course includes faculty, post-doctoral fellows, and graduate students who are interested in learning more about the design and analysis of group randomized trials. The target audience also includes program directors, program officers, and scientific review staff here at NIH. They often deal with group randomized trials as part of their review, and need to understand the design and analytic issues.

Participants should be familiar with the design and analysis of individually randomized trials, or randomized control trials, often called RCTs. Participants should be familiar with the concepts of internal and statistical validity, their threats and their defenses. And participants should be familiar with linear regression analysis of variants and co-variants, and logistic regression.

The learning objectives are shown here. At the end of the course, we want participants to be able to discuss the distinguishing features of group randomized trials, and the distinguishing features of individually randomized group treatment trials, which are related but different in important ways. We also want participants to be able to distinguish these trials from individually randomized trials.

We expect that participants will be able to discuss the appropriate uses of these different designs in public health and medicine -- and for group randomized and individually randomized group treatment trials, participants should be able to discuss the major threats to internal validity; the major threats to statistical validity; the strengths and weaknesses of alternative designs; and the strengths and weaknesses of analytic alternatives. We also expect that participants will be able to perform sample-size calculations, at least for simple group randomized trials. You would also be able to discuss the advantages and disadvantages of alternative designs that could be used to evaluate multi-level interventions.

The organization of the course is shown here. We'll start with the introduction and overview today, and subsequent modules will focus on the design of the trial, analysis approaches, power and sample size, examples, review of recent practices, and alternative designs and references.

We'll start today with three kinds of randomized trials. The first category is the individually randomized clinical trial, or RCT. This is the one that should be familiar to most of you. Individuals are randomized to study conditions, and have no interaction with one another,

post-randomization. This kind of design is used for most surgical trials and drug trials. Sometimes used for behavioral trials, though it can be more challenging in that context.

Individually randomized group treatment trials are a second category. They start out like individually randomized clinical trials, because individuals are randomized study conditions. The difference is that those individuals interact with one another, or have some kind of connection with one another, post-randomization. That does not happen in individually randomized clinical trials, but it is a distinguishing feature for individually randomized group treatment trials. These kinds of studies are common with many behavioral interventions.

The third category is the group randomized trial. In this instance, we don't randomize individuals at all. Instead, we're randomizing groups' study conditions. The members of those groups generally have had some connection with one another before we came along and began the study -- certainly before randomization. The members of those groups will continue to interact with one another during the study, and we'd expect that they'd continue to interact even after the study is over. So interaction and connection among group participants is standard fare in group randomized trials. These studies are common in trials that are conducted in communities, and work sites, and schools, and so forth.

Distinguishing characteristics are shown here. Group randomized trials, as I mentioned, the unit of assignment is an identifiable group of some kind. We have different groups that are allocated to each condition, so that the groups are nested within the levels of the study conditions, that are the focus of the project. The units of observation are members of those groups, and the number of groups allocated to each condition is often limited. That is not a necessary condition, but it is often the case.

Individually randomized group treatment trials, the unit of assignment is the individual participant, rather than a group. Participants receive some of their treatment, however, in small groups -- sometimes face-to-face, a physical group, sometimes through an online or virtual group. Sometimes through a common change agent, where they may meet one at a time with the same change agent. The number of groups or change agents in individually randomized group treatment trials is also often limited. Not a requirement, but it's often the case.

Alternative labels that are used for these studies -- group randomized trials are commonly called cluster randomized trials. Those two terms are interchangeable. They mean exactly the same thing. These kinds of studies are sometimes called community trials, and all of those labels, as I said, are interchangeable. Individually randomized clinical trials are sometimes called randomized clinical trials, randomized control trials, controlled clinical trials, and those labels are certainly interchangeable.

Pragmatic trials are often used in public health and medicine, in recent years, and group randomized trials are often the design that are used for pragmatic trials. Pragmatic and explanatory trials were first described

by Schwartz and Lellouch in a paper in 1967. Explanatory trials test causal research hypotheses. Pragmatic trials help investigators choose between options for care. These are very similar to efficacy and effectiveness trials. The explanatory trials are very similar, then, to effectiveness trials, and the -- sorry. The explanatory trials are very similar to efficacy trials. The pragmatic trials are very similar to effectiveness trials. I tend to use efficacy and effectiveness, but you can also use explanatory and pragmatic if you wish.

I'll give you some examples, rather than talking about these in an abstract sense. The examples come from the Health Care Systems Collaboratory, which is a common fund project here at the NIH. There are nine pragmatic trials that are funded as part of the Health Care Systems Collaboratory. These are funded as UH2, UH3 trials, and funded by a variety of I.C.'s. Eight of them are group randomized trials. One of them is an individually randomized trial.

The eight group randomized trials focus on different outcomes, and the outcomes are listed here. So you can have a variety of different outcomes in these studies, ranging from mortality, to screening behavior, to management of multiple chronic conditions, and lots of other things could be outcomes in group randomized trials. These are the outcomes in the Health Care Systems Collaboratory projects. If you're interested in learning more about these particular trials, or the Health Care Systems Collaboratory more generally, here are three papers that you could take a look at, that will give you more information on that effort.

I can also give you some examples of individually randomized group treatment trials. These come from the Childhood Obesity Prevention and Treatment Research project, or COPTER. This is funded by NHLBI. They're funded as a series of UO1's. There are four of these studies. Two of them are prevention studies targeting very young children. Two of them are treatment studies targeting youth and adolescents. All involve substantial participant interaction, post-randomization, and that's the characteristic that distinguishes individually randomized group treatment trials from the other categories. There's an overview paper, published just a couple of years ago. Charlotte Pratt, who's the Project Officer for this effort at NHLBI, is the lead author.

I'll introduce a little bit of notation. This comes from my book published in 1998. I wish I could say that everyone uses the same notation across the papers and books that address the material that we're going to over, but that's just not true. In fact, everybody uses their own notation. So I'm no different, I have my own version. I do follow some simple rules. Where possible, I use the same terms that others use. So I use "Y" for the dependent variable. And then for other terms, the subscripts and letters that I use come from the terms that they're describing.

So for "condition," that's represented by a capital "C." We have one to "c" with a lowercase, levels of condition. Usually only two, but we can certainly have more than that. "Time" identifies the measurement occasion. I represent that with a capital "T." "Group" identifies the unit of assignment. These are the groups that are randomized to the

study conditions, and that's represented by capital "G." "Member" -- these are the participants that are members of the groups that are observed, to find out what the effect of the intervention is. They are the units of observation, and are represented by capital "M." "Covariates" are represented by a capital "X." These are covariates that we may include in the analysis, to adjust for imbalance on something like sex or age or employment, or other such characteristics. I distinguish between random effects using bold type. So groups and members in a group randomized trial are always random effects, and so noted in bold. Fixed effects are in plain type, such as condition, time, and covariates.

Let's talk a little bit about the impact on the design of whether we're randomizing individuals with no interaction, or randomizing individuals who then have interaction, or randomizing groups. And we'll start with the randomized clinical trial. Here we've randomized individuals, and those individuals have no interaction with one another post-randomization. There is very good opportunity, in most such randomized clinical trials, for randomization to do its job, and that is to distribute potential confounders evenly across the conditions. And that's because most randomized clinical trials are large, with sample sizes in excess of 100, or often in excess of 1000. If they're well-executed, confounding is not usually a concern in these kinds of studies.

Individually randomized group treatment trials are a little different. We start with randomization of individuals, but then those individuals interact with one another. Because, however, we start with randomization of individuals -- if we have enough of them, we don't have to be terribly concerned about confounding. However, most individually randomized group treatment trials involve a smaller number of participants, often less than 100. In those cases, confounding can be more of a concern.

Group randomized trials stand out even more. The typical number of groups in a group randomized trial is 20 or 25, almost always less than 50. Larger studies are actually rather rare. When we have fewer than 25 certainly -- and even fewer than 50 groups that we're randomizing -- randomization doesn't have the same opportunity to distribute the potential confounders evenly. If we got to do the study 100 times, we wouldn't worry about it on average, but we only get to do it once. We only get to randomize once, and simple randomization may not distribute all the confounders evenly, especially as the number of groups gets smaller. So confounding is more of a concern in a group randomized trial.

Let's talk about the impact on the analysis. Observations on randomized individuals who do not interact with one another can be assumed to be independent, and are analyzed with the standard methods that you all learned in graduate school. The members of the same group, however, in a group randomized trial, share a variety of kinds of connections. They may interact directly with one another, and the members of those groups -- the members of groups created for individually randomized group treatment trials will develop those connections over time. There may not be any connection at the beginning of the study, but as they interact, they will develop those connections.

These kinds of connections create a positive intraclass correlation that reflects extra variation attributable to the group. So the ICC, or intraclass correlation, is something that we're going to be talking about in every single module in this course, and frequently. It's an important concept to understand. One way to define it is it's simply the average by variate correlation among the members in a group. It reflects the degree to which they are associated, or the degree to which their responses are similar. These intraclass correlations tend to be small but positive. The intraclass correlation reduces the variation among the members in the group, so that -- the within-group variation is shown at the bottom of this slide, and it's reduced as a function of the ICC.

It creates a second component of variants, however, called a group component, that is usually a very small fraction of the total variants, but it's still there. And it's shown in the top expression. The total variants is then the sum of the two components, the residual and the group component. Another way to define the intraclass correlation is as a fraction of the total that is attributable to the group. So sigma square G, colon C, divide by the total variation, is another definition of the ICC. This second definition would not seem to provide for negative values in ICC's, where the original definition at the bottom of slide 14 -- or sorry, in the middle of slide 14 -- does allow for negative correlations. Negative correlations can actually happen, even though they may appear to make very little sense, with the formulas shown here. They still can occur in nature, and we will talk about those a little bit later.

The impact on the analysis is shown here. If I create groups by randomly assigning individuals to those groups, and then calculate the variance of the group mean, that variance is simply the usual variant sigma squared Y, divided by the number of observations contributing to the mean, which is M, because there are M observations per group. That's an expression that you may not have seen with this notation, but it's the standard formula for the variance of a group mean. And you've seen some version of that formula in your first-quarter biostatistics course.

If the groups are not established by random assignment, we have a different situation. The residual error, sigma squared E, is divided by the number of observations contributing to the mean, but the group component of variance is not divided by anything. And so the expression is a little bit different than it was at the top. We can re-write that middle expression as shown at the bottom of slide 16.

The variance of the group mean is sigma squared Y, divided by M -- the usual variance of the group mean -- multiplied by a parenthetical expression which includes M, the number of observations per group, and the intraclass correlation, ICC. One plus M minus one times the ICC is sometimes called the variance inflation factor, sometimes called a design effect. If the ICC is a positive value, that parenthetical expression will be positive, and therefore the multiplier will be positive -- and so the variance of the group mean will be larger than it would be, if the groups had been established through random assignment.

There are several implications from this, for the analysis. The first is from a paper by David Zucker, published in 1990. If a factor is nested, it has to be modeled as a random effect in the analysis. So that's an important feature to remember. We have two levels of nesting in any group randomized trial. Groups are nested within conditions, and members are nested within groups. So we always have at least two random effects in a group randomized trial.

Another implication is that the variance of any group-level statistic will be larger. Group-level statistics reflect the extra component of variance, and so they will be larger than if we had constituted those groups through random assignment. A third implication is that the degrees of freedom available to estimate the group-level component of variance will be based on the number of groups -- and if I don't have very many groups, the degrees of freedom will therefore be limited.

As I said, the number of groups in group randomized trials is often 25 or fewer, and so the degrees of freedom -- start by subtracting two from the total number of groups. So if I have 25, there are no more than 23, and they may be smaller than that, less than that if I have group-level covariates. So this is always an issue in a group randomized trial. It can be an issue in individually randomized group treatment trials. The main point that I want to make about this, is that any analysis that ignores the extra inflation or the limited degrees of freedom will have an inflated type one error rate.

And that inflation can be substantial. In a group randomized trial, even with a very small intraclass correlation, the type-one error rate might be 50 percent, rather than the nominal five percent. And in an individually randomized group treatment trial, the type one error rate might be 20 percent, or 25 percent, rather than the nominal five percent.

We don't want to report results based on type one error rates of 25 percent or 50 percent, and so we need to address these issues. It's also important to keep in mind that extra variation and limited degrees of freedom always reduce power. So even if I take them into account, power may be reduced.

Let's talk about the impact on the analysis of these features, and to do that, I'm going to talk about a paper by Scott and Holt published in 1982. They present a formula for the general version of the design effect, which is written as one plus M minus one times two ICCs. ICC sub-Y is the intraclass correlation for the dependent variable. That's the one that I've been talking about so far.

ICC sub-X is the intraclass correlation for the independent variable, and we usually don't talk about it. But it's there in the general form of the equation. The reason that we don't usually talk about it, in a group randomized trial, is that the ICC sub-X is always one in a group randomized trial -- and that's because everybody in the same group gets the same intervention, and so they all have the same value on the independent variable.

For most health-related outcomes, ICC values tend to be small. If we're looking at large aggregates like schools, or work sites, or churches, the intraclass correlations are usually less than 0.05. If we're looking at smaller aggregates like classrooms, or departments within a company, we might see values that range from 0.05 to 0.25. If we have very small aggregates like spouse pairs, or families, we can see very large intraclass correlations, even up into the 70s and 80s.

ICCs tend to be larger for measures of knowledge and attitudes, smaller for measures of behavior, and smaller still for physiological measures. If the groups are crossed with the levels of the exposure of interest -- and that's true for most observational studies -- then we do have two intraclass correlations at play. They tend to have similar magnitude. If the groups are nested within the levels of the exposure of interest -- and that's true in an individually randomized group treatment trial, and a group randomized trial -- then the intraclass correlation for the exposure, or the independent variable, is one -- because all of the members of a group will get the same value for exposure.

We can use that formula to calculate design effects, and they're shown here in this table. If we're conducting a survey project, where we've used cluster sampling for example, intraclass correlations are at work both for the dependent variable and for the exposure. They tend to be small. The number of members in each group tend to be small, but not necessarily small. They can certainly be larger -- but because I have two intraclass correlations, and I'm multiplying them together, the design effects certainly tend to be small. So as shown on the left-hand portion of this table, even if I have 200 people in each of the primary sampling units in my survey design, and the intraclass correlation is fairly large, 0.05, the design effect is only 1.5. So it's not a great number, and if the intraclass correlation is smaller, then the design effect is even smaller -- almost negligible.

On the right-hand side, we have group randomized trials, where the intraclass correlation for exposure is one, because everybody in the same treatment gets the same -- or in the same group gets the same treatment. Here we have group sizes that are similar, perhaps, to what we have in surveys, often ranging even larger. For an intraclass correlation of 0.05, which can happen, and 500 people observed in each group, the design effect might be 25 or 26, and that's a very large number. Even if the intraclass correlation is smaller, the design effect can be pretty large.

The design effect is used to correct an unadjusted F test, by dividing the unadjusted test by the design effect, to get the corrected test. And I'm sure that none of you would want to divide your uncorrected F test by 25 or 26, or even by 6, because it would make it much harder to report a significant effect. So it's very important to take these issues into account, so that you're not left with that kind of situation.

Jerry Cornfield published a paper in 1978, in the American Journal of Epidemiology, and it includes one of my favorite quotes. I like to present this every time I give a talk on group randomized trials: "Randomization by cluster, accompanied by an analysis appropriate to randomization by individual, is an exercise in self-deception, and should

be discouraged." So I love the understatement of that remark. It's certainly true for group randomized trials, and even though his remarks were only addressed to group randomized trials, they also apply to individually randomized group treatment trials.

So let me summarize what we've covered in this first module. The group randomized trial remains the best comparative design available, when an investigator wants to evaluate an intervention that operates at a group level; that manipulates the social or physical environment; that cannot be delivered to individuals without substantial risk of contamination. So it's the best design available if you have an intervention that meets any of those characteristics.

An individually randomized group treatment trial is the best comparative design to use if individual randomization is possible without substantial risk of contamination, but there are good, often practical reasons to deliver the intervention in groups. The challenge for us is to create trials that are rigorous enough to avoid the threats to validity, that are analyzed to avoid the threats to statistical validity, that are powerful enough to answer the question that we're posing, and inexpensive enough to be practical. And we're going to take up those issues over the remainder of the course.

I want to thank you for participating in part one of Pragmatic and Group Randomized Trials in Public Health and Medicine. I draw your attention to our website, shown on this slide, where you can provide us with feedback, give us comments on the series. You can also download the slides that you've just seen. You can download all of the references for the entire course.

You can download suggested activities for each of the modules. You can view this module again, multiple times if you want to. You can view the next module in the series from the same website. So we encourage you to move onto part two, once you've finished digesting the material in this one.

If you have questions about any of the material, or questions about group randomized trials in general, you can send them to GRT@mail.NIH.gov and we'd be happy to try to answer them for you. Thank you very much for your attention, and we look forward to seeing you with part two.

[end of transcript]