David Murray:
Hello, my name is David Murray, I'm the NIH Associate Director for Prevention, and Director at the Office of Disease Prevention. I want to welcome you to Part Six in our course on Pragmatic and Group-Randomized Trials in Public Health and Medicine. Part Six focuses on a review of recent practices used in Group-Randomized Trials. This is part of a free, seven part, self-paced online course presented from NIH. We provide the slides for each of the modules that you're watching, a complete set of readings, and a set of guided activities for each of the modules.

The target audience for this course includes faculty, post-doctoral fellows and graduate students interested in learning more about the design and analysis of group-randomized trials. We also want to reach program directors, program officers, and scientific review officers here at NIH who need to learn more about these designs. Participants should be familiar with the design and analysis of individually randomized trials, or RCTs. They should also be familiar with the concepts of internal and statistical validity, their threats and defenses, and with linear regression, analysis of variance and co-variance, and logistic progression.

The learning objective is shown here. At the end of the course we want participants to be able to distinguish group-randomized trials, individually-randomized group treatment trials from individually-randomized trials. We want you to be able to discuss the appropriate uses of these designs in Public Health and Medicine. And for group-randomized and individually-randomized group treatment trials to discuss the major threats to internal validity, to statistical validity, the strengths and weaknesses of design alternatives and analytical alternatives, and to perform sample size calculations -- at least for a simple group-randomized trial. You'll also be able to discuss the advantages and disadvantages of alternative designs that can be used to evaluate multi-level interventions.

The organization, of course, is shown here and today we will cover part 6, "A Review of Recent Practices." This work is based on a paper that we published from the Journal of National Cancer Institute in 2008. There have been a series of reviews published on the state of the practice in group-randomized trials beginning in 1990 with the paper that was published by Allan Donner's team. They observed that only 19 percent of the published articles that they reviewed took the intra-class correlation into account in the sample size calculations and only 50 percent of the papers published and review in their article took it into account in the analysis.

A few years later, Allen's team published another paper that showed very little progress. It was still 19 percent taking the intra-class correlation into account in sample size calculations. Things looked a little bit better, in terms of taking it into account in the analysis. We published a paper in 2004 but things seemed to have slipped back. Though, granted, the standards that we were applying than the standards used in previous reviews.

So, that may have accounted for the difference. Only 15 percent of the inter-class correlation taken into account in the sample size calculations and only 54 percent took it account, always, in the analysis. We were certainly interested in whether this situation had improved and, so, we set out to do another review. We looked at papers published from 2002 to 2006, inclusive. We conducted a Med Line and Pub Med search. Studies had, as their primary outcomes, cancer risk factors, cancer morbidity, or cancer mortality. Studies used randomization to assign identifiable groups to study conditions, with observations taken on the members of the groups.

If the article that we were reviewing referred to an earlier design paper, we also read and reviewed that paper. Each reviewer, and we worked in teams of two and three, independently assessed the article on items related to design, sample size estimation and analysis, and then the reviewers met and discussed each paper as a group to work out any disagreements. We found 92 possible trials, in 45 journals. We ended up selecting 75 of those, from 41 journals, because they met the inclusion criteria and there were also 20 background or design papers. 20 percent of the articles were published in preventive medicine, seven percent in the American Journal of Public Health, no more than 4 percent in any other journal. And, as I said, there were 41 journals represented.

We noted that 15 group randomized trials were published each year from 2002 to 2006. Compared to 11. 6 per year in our earlier review and 5. 3 per year in the Simpson et. al. reviews. So, group randomized trials have been becoming more common over the course of time. I want to start with table one from this paper, which summarizes the analytic methods that are frequently used in group randomized trials and identifies the conditions under which they are appropriate for use. So, the first category is "Mixed Model Methods" where we have repeated measures, analysis of variance or co-variance.

These are appropriate if we have outcomes with just one or two time points. We are looking at comparing the variation in the conditioned means -- or, sorry, the group means or rates against the variation at the condition level and the degrees of freedom are based on the number of groups. Another model based approached that is perfectly okay to use is random coefficients. This is appropriate if we have three or more time points and the variation in the slopes and intercepts at the condition level is assessed against the same kind of variation, at the group level. The degrees of freedom are, again, based on the number of groups.

So, those model based methods, mixed model ANOVA/ANCOVA and the random coefficients approach can be used, dependent on the number of time points that are included in the analysis. The generalized estimated equations approach can be used with a correction for limited degrees of freedom. This is called for if there were fewer than 40 degrees of freedom. We need some kind of small sample correction and some of those available in some of the statistical packages. If there's no correction, you can use standard GEE, where you have at least 40 groups being randomized to commissions. But you shouldn't use standard GEE in smaller studies.

Two-stage methods are appropriate in group-randomized trials. Whether large or small, they have to be used appropriately, of course, but they can be used to good advantage. The degrees of freedom are based on the number of groups. Then there are a series of methods that are sometimes used but are not widely recommended. Analyzing the data at an individual level with a post-doc. correction based on external estimates of interclass correlation. It could be valid if the external estimate is appropriate for the data at hand, but there's no way to test that assumption. So, that's not an approach that we recommend. Analysis at an individual level ignoring the group all together, we never recommend. Analysis at a sub-group level, ignoring the group level, is also something that we do not recommend for studies where causal inference is desired.

So, let's look at the results from the 75 articles that we reviewed. 88 percent had two study conditions. So, the most common design, at least during this time period, was an intervention versus control -- or "two arm" study design. It was a smaller number that had three conditions and even a few with more than three. The most common design is a cohort design. 67 percent of the studies that we reviewed used a cohort design. 20 percent use across sectional design and 13 percent used a combination. Combinations are, actually, fairly easy and efficient if you are collecting serial cross-sections and follow up some of the first cross-section as a cohort at the end of the study, you can have both in a reasonably efficient manner.

We and others have been recommending "matching" or stratification in the design for some time. 20 percent of the studies we examined used matching alone. 35 percent used stratification alone. Five percent uses both but 40 percent didn't use either and we think matching or stratification should be used pretty routinely in group randomized trials. A variety of different kind of groups were used. Churches in five percent of the cases, communities, neighborhoods, other kinds of community groups in 19 percent of the cases. Positions are provider groups for commons schools and colleges with common. Work sites were common. So, we see group randomized trials conducted in a variety of contexts with a variety of different kinds of groups.

The number of groups per condition varied considerably. The -- over time the average size of the group randomized trials has been increasing. So, we saw a lot of studies in this review with more than 12 groups per condition. 13 or -- 23 percent had 13 to 25 and 16 percent had more than 25. We consider those to be rather large group randomized trials. There still is a number, though, that has a very small number of groups. Including 5 percent that only had one group per condition, and that's a design that we don't recommend at all.

The number of members per group also varied. Often, we see fairly large studies in group randomized trials. That 23 percent had had more than 100 members per group. It's unfortunate that some of the studies didn't report that information. The number of time points are described. 54 percent analyzed. Post test data, with an adjustment for baseline -- so, we call that "one-time point" being used. 29 percent analyzed pre-post data in a repeated measures fashion. So, that's two time points. And

then we had a number of studies that involved more time points where the appropriate analysis, at least, would be brand and coefficients models. Some studies did not report adequate information.

The focus of most of these studies was on Primary Prevention. Secondary Prevention was also very well represented. Other categories, less common. Individuals with no personal history of the target cancer was the second most common category. Unknown or Mixed, because it wasn't described in the paper, was the most common category. Primary outcome variable ranged considerably. We had studies looking at tobacco, looking at physical activity, looking at dietary variables. Lots of the cancer risk factors. Some studies looked at incidents of cancer and some incidents looked at mortality. But, most of the group randomized trials were looking at either risk factors or use of screening activities.

The analytic methods that are used in the 75 articles are summarized here. We found that only 45 percent of the 75 articles reported exclusively the use of what we consider to be appropriate methods in the analysis of their data. So, that means that for 55 percent either didn't use appropriate methods or used a mix of appropriate and inappropriate methods. Most common type of appropriate method was mixed model ANOVA/ANCOVA. Or mixed model repeated measures with just two time points. We only found one paper that reported a random coefficence model. A few that used GEE with a larger number of groups and a few that used two-stage models. And all of those methods are appropriate for group randomized trials.

We found eight percent of the studies reported using some appropriate methods and some inappropriate method. The appropriate methods that they used were often the same ones that we just talked about, that had been used by other articles exclusively. But then these studies also added some methods that are inappropriate. Analysis at an individual level, ignoring the group all together was the most common one. With five of the six studies doing so.

Then we had 35 percent of the papers that we reviewed, that only reported inappropriate methods. So, they didn't use any of the techniques that we consider to be proper for group randomized trials. The most common poor method was analysis of an individual level, ignoring the group. It's as though these studies weren't aware they were conducting a group randomized trial. These methods have been debunked for many, many years and this stage and certainly when we did our review. And, so, there's no excuse for using analysis of an individual level, ignoring the group all together. But it was still done in an awful lot of the studies.

Now, let me turn to a recent review in practices in an individually randomized group treatment trial. This is a paper published by Sherri Pals of the American Journal of Public Health in 2008. There had been no prior reviews of individually randomized group treatment trials and we decided to focus on six particular journals, rather than reaching across all the journals that are out there. We looked at the same time period, 2002-2006. We focused particularly on American Journal of Public Health and Preventive Medicine because they were well represented in our other review.

We also looked at Health Psychology, Obesity Research, Addictive Behaviors and Aides and Behavior. The procedures that we used were parallel to what we used in the GRT review. The criteria for sample size and analysis methods were quite similar. We found 34 eligible articles, and this describes a little information about those 34 articles. You can see the distribution across the journals and across the years. There was a big jump in 2006. You can tell though from this data if that's a jump that was just a bump, temporary, or if these designs were becomingly more commonly used.

As we found with group randomized trials, two study conditions is most common in individually randomized group treatment trials. 67 percent had intervention versus control. Another 23 percent had three conditions. The number of conditions that included small groups, usually just one but sometimes two, the baseline sample size tended to be large. This is total size, not per group. So, less than 144 percent but often more than 100.

The target populations are shown here. These studies are being used in a variety of study populations and looking at a variety of outcomes and here are the outcomes. Weight, BMI, body fat was common. Smoking or other substance abuse variables, that was common. So, they were -- these designs are individually randomized group treatment trials, are used across a variety of outcomes.

When we looked at the sample size calculations that they were using most made no mention of their sample size calculation and the next highest category was doing sample size calculations there ignored the individually randomized group treatment trial design. 80 percent of these studies reported significant effects but most of them did not pay attention to the kind of design they were using in their sample size calculation.

When we looked at the analysis approaches it was rather dismal. 94 percent performed an analysis at an individual level, ignoring the group all together. We only found one paper, out of the set, that reported exclusively appropriate analytic methods and the rest, it was either entirely inappropriate or there wasn't enough information to judge. So, the situation is far worse for individually randomized group treatment trials published during the same period, than it was for group randomized trials.

Summarizing this information, our results for group randomized trials were pretty similar to what we'd seen in earlier reviews. 45 percent of the articles reported only analysis judged to be appropriate. 35 percent only analysis deemed inappropriate. And eight percent reported a mix. There's a more recent article by Crespi et al. published in 2011 that provides data for some more recent time periods. So, they were focused on cancer screening studies. They reported 92 percent of the studies used appropriate methods when those studies were published, 2003-2006, which is almost exactly the same period that our review -- examined.

Their study was limited to cancer screening where we looked quite broadly across cancer studies. We found 45 percent, they report 92. So, very big difference but probably a great deal of that is because they were focused much more narrowly on a subset of the studies. They also reported better results for the later period, 2007 to 10, but because it was so much lower than 2003 to six, that does not bode well for the broader array of group randomized trials. And here's a reference for the Crespi paper.

Warnings have appeared in the literature for at least 30 years, regarding the development of interclass correlation in individually randomized group treatment trials. Even so, the literature on their design and analysis is quite limited and the use of inappropriate methods in pervasive in the published studies that we saw using individually randomized group treatment trials. The picture for group randomized trials isn't much better than it was in the early 90's, and we hope that situation will improve, over time.

I want to thank you for your attention today as we reviewed the state of the practice for group randomized and individually randomized group treatment trials. This is part of our course on pragmatic and group randomized trials and Public Health and Medicine. We draw your attention to our website, where you can provide feedback on the module that you have viewed today. You can download the slides, you can download all of the references for the course and also suggested activities to go with today's presentation. You can certainly view this module again, and you can view the next module in the course, part seven, focused on alternative designs to evaluate studies with multi-level interventions.

If you have questions, please send them to GRT@mail.nih.gov and we'll respond as soon as we can. Thanks very much for your attention.

[end of transcript]