

AZORE: Hello everyone. We will be starting now.

ZANDNIAPOUR: Hello everyone. And welcome. My name is Lily Zandniapour and I work at the Office of Research and Evaluation at AmeriCorps. I want to thank you all for joining us today. We thrilled to see the interest in this topic and webinar, especially given the many preoccupations that you all have at our homes and at work due to the pandemic. Of course we hope that you, your families, colleagues, and friends remain healthy and stay safe.

The research and evidence webinars series is one of the venues that we use to share information, new resources, and research and evaluation results with our audiences. They provide learning opportunities for us, with our stakeholders and colleagues inside the agency and outside. Before we begin, we'd like to cover a few housekeeping items for Adobe connect. So I will turn it over to Jenelle Azore from ICF Next [ph.] our communications partner that supports this webinar. Jenelle?

AZORE: Thank you. This webinar will be recorded and posted online following the presentation. There is no dial-in phone line. All audio is broadcast over the internet using your computer speakers. All participants will be in listen only mode until the question and answer session following the presentation at which time you can ask a question using your computer's microphone by selecting the raise hand feature from the menu bars. You can also ask questions at any time during the presentation by using the Q&A or chat boxes below.

As mentioned earlier, this webinar is being recorded and if you have any questions or experience technical difficulties, please let us know using the Q&A or chat boxes below. Lily, I believe that takes care of our housekeeping items.

ZANDNIAPOUR: Thank you to Jenelle. Let me just start by giving you a little bit about the format for today's webinar. We're going to start with a brief introduction and I will then pass the mic to Dr.

Carrie Markovitz from NORC to lead the webinar and introduce our main presenter, Dr. Eric Hedberg.

The NORC team's presentation will likely take about 15 minutes or so and after that, we will open it up for Q&A and any comments or discussion, and we will have about half an hour for that. And my colleague, Dr. Andrea Robles will facilitate that portion of the webinar.

I'm delighted that we are able to offer today's webinar on power analysis to you. This webinar has been a long time coming and is the first in a three part or three levels series of trainings that we have planned on this topic. Level one provides an introduction to power analysis, level two, we'll delve deeper into the basic mechanics of power analysis, and the focus of level three training will be on applied power analysis.

This webinar, and really the totality of the three-part training series will be delivered by our evaluation training and technical assistance

provider NORC at the University of Chicago. We have been fortunate and have been working with NORC for some years now to strengthen our grantees evaluation studies so they are better positioned to produce credible and quality evidence about the programs and interventions they implement.

This also provides us with good information about the program models supported by the agency and their effectiveness. I want to take a moment to tell you why we chose to focus on this topic and I know that my colleague from NORC will also expand on this some more.

Following the passage of the Serve America Act in 2009, AmeriCorps as an agency began to more intentionally organize and develop the evidence base for national service programs and increase it's focus on evaluation, evidence building, and use. We've been conducting systematic evaluation plan and evidence reviews as part of our grantmaking and grants management process for nearly a decade now.

Even though AmeriCorps continues to support program models at different points along the evidence continuum, there is a focus on rigorous evaluation, on identifying and supporting solutions that work and on measuring program impacts as individuals, organization, and community levels.

We have learned a lot during time. Our experience has underlined the importance of well-designed and well implemented evaluations to produce high quality and credible evidence, particularly when it comes to impact studies that aim to measure effects and demonstrate that the program actually caused outcomes that were achieved. If you embark on designing an impact evaluation, when selecting your study sample, power analysis becomes relevant and key to making sure you're able to capture and measure impact successfully.

It is a critical element of a well-designed and rigorous impact study and it's important to conduct it during the planning phase and not so much at a later date. Doing the power analysis is really just

part of doing good research and it is a way of making sure that you've thought through every aspect of the study and the statistical analysis before you start collecting data. So sharing more about this topic, we think is really helpful.

This webinar will give you more grounding on this topic, and we specifically wanted it to be introductory and make it relevant for a broad set of audience groups. You may not be conducting a power analysis, but it is helpful to know what it's about and why it's needed. The second and third trainings as I described earlier, will have more technical content for application purposes.

So please stay tuned on what your come your way on this topic from us in the near future. With that, I will now turn it over to Dr. Carrie Markovitz, the principal research scientist at NORC and project director for our evaluation training and technical assistance work for AmeriCorps state & national program to lead us in this webinar and introduce our main presenter. Carrie?

MARKOVITZ: Thank you. NORC at the University of Chicago has been collaborating with AmeriCorps Office of Research and Evaluation for almost a decade to strengthen the existing evaluation guidance and tools for AmeriCorps state national applicants and grantees. Our activities include reviewing and providing feedback to grantees on their evaluation plans and reports. We also provide intensive individualized evaluation technical assistance to select grantees.

And we continuously discuss challenges, identify grantee needs, and make recommendations to improve the evaluation capacity of ASN grantees. Over the years, our TA team has assisted and supported numerous AmeriCorps grantees with their evaluation plans, design and implementation challenges, instrument development, and reporting. And through this work, we have identified common areas of need and then worked with the Office of Research and Evaluation on developing tools, webinars, classes, and presentations to fill these gaps.

So this class on power is one of several classes on evaluation topics that were developed over the past 10 years by NORC and are currently available online on AmeriCorps' website.

These classes include information on topics from logic model development, to drafting research questions and addressing other design topics, to budgeting and managing and evaluation. So, we encourage everyone to check out these other classes based on their own evaluation needs.

As Dr. Zandiapour mentioned, this presentation is the first in a three-level series of trainings on statistical power. The idea for these power classes grew from a presentation that Dr. Hedberg conducted up to 2019 AmeriCorps State and National Grantees Symposium. He conducted the presentation because our TA providers were fielding a lot of questions on power analysis, especially from grantees designing impact evaluations of their programs.

The session was so well attended that AmeriCorps decided to support an ORC in developing a full series of classes on statistical power. So, we would like to thank Lily Zandniapour, Andrea Robles, and Mary Hide [ph.] at AmeriCorps Office of Research and Evaluation for their vision and their support in developing the series.

Today, you will be provided an introduction to the concept of power and why power is an important consideration in evaluation design. NORC also developed two additional classes on power that are designed for more advanced audiences. Level two provides more detail on the elements that affect power levels and how they factor into power analysis.

Level three is even more advanced and it focuses almost primarily on how to conduct power analysis for several different design options. The level three presentation includes multiple live demonstrations of power calculation. We hope to have

these additional classes available soon on AmeriCorps' website.

So finally, I would just like to introduce Dr. Eric Hedberg. For Dr. Hedberg power is a major methodological topic interest. Most known for his work on evaluation design, Dr. Hedberg is a deeply interdisciplinary quantitative methodologist. His research interests include several areas of methodology related to evaluation and analysis. And he recently authored a Sage Little Green Book on statistical power analysis.

Dr. Hedberg is an accredited professional statistician by the American Statistical Association, and he is a sociologist. His current areas of research include investigating the design of evaluations in education and criminology, in addition to measuring social capital through social network contextual effects. He is best known for estimating and publishing important experimental design parameters.

Dr. Hedberg has authored or coauthored over 30 methodology focused papers and books that have appeared in education, medical, and criminological journals while also contributing to numerous reports and presentations at major research conferences.

Dr. Hedberg earned his PhD in sociology from the University of Chicago. So now I just want to pass this over to Dr. Eric Hedberg.

HEDBERG: Thank you, Carrie. I, you know, when I want to hear about that, I'm glad I went to college. Okay. So, hi everyone. I'm here to discuss power and I'm really glad everybody's here because I think this is one of, a topic in modern research that just can't get enough attention right now. And this first level is really designed to sell you beyond just signing up for a webinar, really sell you on the importance of paying attention to power analysis and sort of the very, very broad mechanics of it. So you can talk to your neighbor about it if you really want to.

Level two, we'll go into a little bit more of sort of the basic mechanics and level three, like we said, is where the equations and the demonstrations are really going to come in. And so let's dive in, shall we?

So, like I said before, level one is this is the broadest message that I want to give. I want to tell everyone about, I've even talked to my dog about it. And we're going to introduce the concept of power analysis and I'm going to really sort of touch on conceptually why it's so important for evaluation planning. The sort of main reason is it serves sort of two of the primary audiences. First is it really helps lend more support, lend more intellectual validity to any one evaluation.

But then it also, if everybody has high power in their analysis or in their designs and data, it also really supports the broader social science or medical science, or any science that you are doing your studies in.

So after this, the next recorded webinar that should be available pretty soon is going to really kind of get into some of the mechanics of type I and type II errors that many of you may remember from your introductory to statistics course that you enjoyed and or suffered through. And then level three will really sort of extended to beyond, and really sort of give you the tools, the buzzwords, the necessary concepts you need to conduct a power analysis or work with other statisticians in your projects to conduct their power analysis.

And then level three will also have a little bit of information about how to really write up a power analysis, or for those of you judging other write-ups what to look for when you read a power analysis.

Okay, so why is power important? So, you know most people who are tuning into this webinar are evaluating programs or know people who are evaluating programs, and you want to sort of see, okay, does this particular intervention work? And sometimes you know, it costs money to collect data

and sometimes, you know, there's more data that could be collected than you have money for. And so you're going to want to collect some kind of sample. And that means you need to sort of say, all right, I have so much money. How many participants do I need to measure?

You know, it could be an online survey, or it could be a very expensive psychometric tool. So, as you plan how large your sample needs to be, you want to sort of check and look into the crystal ball a little bit, to get a sense as to whether or not you have a chance to detect the impact of your program. Because it's not always guaranteed. And so this whole broad topic of power analysis, it's really a family of many other smaller types of analysis that really helped, you know, sort of play what if scenarios and say, if the effect is likely this, and I have a sample of this, what's my chance of getting a statistically significant result?

Or if you sort of say, god, if I only have this many, a number of data points, what's the smallest

effect I could credibly have a chance of detecting. And so power analysis is really a family of you know, what if scenarios that help you plan as you go along. And as Lily mentioned, it's also a very helpful process while you're planning, because it really kind of makes you think about every aspect of your study, the whole life course of your study, you have to sort of think critically about.

And so it's kind of a, it's really an important process in any study, I think. So again, to sort of bring you back a little bit to stats 101, this is a chart that I would make for my undergraduates that really sort of, kind of explicitly lays out what statistics are all about. The idea is that, you know, you have a population of interest, the sort of solid blue and these populations have parameters. And these parameters are sort of these platonic noble quantities that we will never actually know because it's impossible to actually know everything about the population.

So, what we're often left to do is to sort of draw this ignoble sample, a subset from this population, drawn in hopefully some systematic way from which we will estimate a statistic and estimate, you know, if we compare treatment to a comparison group, what's sort of the difference on average. And the idea is that this statistic drawn from this sample is supposed to help us infer something about the parameter, to which we'll never know.

And, you know, the whole a game with statistics is, of course, we actually don't know the parameter. We don't know what the average has income is. We don't know the average effectiveness of any particular intervention. All we know is our statistic and whether it's statistically significant and what is our sort of estimated size of the impact. And so power analysis is really a process through which we're trying to plan our study so that our estimated statistic from imperfect samples actually give us some information about these parameters.

And one of the things that I really want to emphasize in a lot of this is, you know, we all love our datasets and we all love our Excel sheets of information that, you know, hundreds of thousands of dollars were spent collecting. But we have to realize that this is one sample out of an infinite possibility of samples you could have collected. And so we have to keep thinking, not in terms of all one dataset, but that this one data set is sort of one jelly bean out of this jelly bean jar. And so, you know, will our study be effective in showing the impact should it truly exist? And so we have to sort of keep thinking in terms of not one data set, but this is one data set out of a jar of many possible datasets that we could have done.

And so, you know, this lends us to hypothesis testing and, you know, typically in classic statistics, we walk in and we say the null hypothesis that the program is not effective. We're going to compare it to this alternative hypothesis that these treatment and comparison groups are different and, you know, the basically probabilities

that we'll talk about in the later classes you know, curves and sampling distributions, and all sorts of things that basically we draw for each of these hypotheses. And power analysis, numerically is ultimately cutting up those mounds of dirt or those areas under the curve to talk about what are our chances of finding a statistically significant effect or not.

So, with that in mind, you know, if somebody asks you, what is power? Power is ultimately a probability of rejecting the null hypothesis that you know, that there's no impact, that you can best expect from your study. So it's the chances, based on everything you know, that your study will actually deal the statistically significant result.

And so obviously this is looking into the future so power analysis is about planning and the power analysis is sections and paragraphs in reports or proposals are really, you know, estimates about the future. And it should force you to start thinking

about how you're using your evaluation resources wisely.

And this is something I'm going to hammer home a lot, but power analysis is really only useful when you're talking about planning a study. Once the study is done, a post-hoc power analysis are sort of saying, well, we got this effect but the power analysis is less helpful because it's actually only one of the jelly beans out of this whole jar.

So, you know, if everybody did have highly powered studies, if everybody did recycle, it has benefits for individuals and for the larger ecosystem of science and trying to make life better for people. For your grantees, it helps increase the chance of detecting the effects. For the funders, people who fund webinars like this and some studies it helps ensure that you're resourcing evaluations that are likely to produce reliable evidence. And then for researchers and science writ large, it helps increase the reliability and trust in study findings.

Because if everybody has a poorly powered study, there's a chance some people are going to win the lotto, find an effect that doesn't really represent reality and go off and publish it. And so, you know, this sort of low level of power in science is what's led to a lot of replication crises in many sub disciplines.

So just to sort of, again, kind of hammer home this sort of idea is, you know, going back to this idea that there's this true parameter that we'll never actually know, this is what happens when we have a poorly powered study. Let's say there actually is an impact, this this program does, you know, increase outcomes for certain individuals. But we try to detect this true impact with a poorly powered study. It has a low probability of having a statistically significant effect.

So we can't say we detect the impact, and we have to tell the world that there's no evidence that the program was effective. So, this is sort of

ultimately a little cartoony example of what could happen if you have a poorly powered study. Whereas if you have a well-polished study, you are able to detect the actual true impact, and you get a little ribbon and a gold star, and you can report to the world that at least from your one sample, you have evidence of an effective program.

Okay, so hopefully by this point, I've sold you on the idea that it's an important thing to worry about and to worry about early. The probability that you have a statistically significant result is based on a number of things and they all kind of work together. One is the design of the study. And by that, I mean, how do you create a treatment group? How do you create a control group?

The amount of data and the structure of the data. Are you just going to randomly pick people at the mall or do these people go to class at particular schools, first pick the schools and then pick the people in the school. So, all of those sorts of things, you know go into power, but, you know,

generally more data was more power, but not always and we'll get into some of those nuances in the later courses.

Power is also impacted by how you analyze the data. Do you just do a simple, independent samples T test, or do you do a regression with a bunch of covariates? You know, those will have an impact on power. And then power is also impacted by the actual impact that's out there in the world. You know, if that impact is rather large, then a good random sample is likely to pick that up. If the program is actually not effective, well, then if the sample is good, then you probably won't get a very big impact and you won't have a statistically significant result.

Power is also impacted by your significance level or sort of how many stars you're going for, for those of you who you know, write reports and put little key value stars. If you think 0.05 is great, you have more power than your colleagues who are being a little more strict and say, nope, we're only going

to say a P value of 0.01 is important. Same test statistic has different levels of power, depending on those scenarios you walk into.

So, you know, the idea here is we're really kind of getting into this ecosystem of power analysis. And, you know, if you walk into a study saying, you need power of 0.9, you need a really high probability. Well, that's going to mean you're going to need a bigger sample and, you know, you'll have the ability to pick a certain effect size. If on the other hand you say, I know the effect size is this. I can tell you how much sample you need for a given level of power. You know, there's sort of a, it's like an IKEA shelf. It's sort of, you know, you bang one shelf and the other one pops out and so it's sort of a system that works together.

But broadly, you know, if you want to go home and tell your significant other what you learned in today's webinar, you can tell them broadly two things that the bigger the more power any given sample is likely to have. And I have an analogy to

kind of bring this home in a second. The other thing is generally the bigger sample will have more power. You know, more data is, you know more information typically. It's not always the case, but you know, this is sort of a broad thing that you can say over dinner today. You have my permission. Whereas the smaller samples tend to have less power.

One sort of analogy that I sort of like to use often is this idea of a microphone. I went to sort of an arts high school and I had a lot of friends who had bands and so I spent a lot of time sitting in basements while they recorded on 8-tracks for those of you who remember life before computers. And you know I often got into these deep conversations that the people who had the bigger microphones were able to pick up very quieter much more nuanced sound. Whereas, you know, the, you know, kids who had these little, teeny microphones, like the ones on your cell phone, they really had the rock out and play really large sounds in order for it to be picked up.

Well, this is kind of a similar thing with power and your sample sizes. If your effect of the program, if your program is really effective, you don't need a very big sample, you don't need a very big sample to get a statistically significant result. If however, your program is effective and is meaningfully effective, but it's more nuanced, it's kind of a smaller effect. Then you have to need more data in order to get that statistically significant result. And so, you know, we often want to think about our samples as sort of these tools of detection, much like a microphone is for sound.

And so you know, in thinking through all these aspects that I mentioned, how you perform your groups, how you analyze the groups, how you sample. Doing this power analysis early on is really going to force you to sort of do your homework, it's really going to force you to look at all aspects of this study. And so I think power analysis, even if you are sort of, you know, have large samples that you can expect lots and lots of money, I think power analysis is still a good thing to do, because it's

going to force you to think these things through and sort of walk in with a plan. So, power analyses are very important for you.

So, what exactly does a power analysis tell you? Well, again, like I said, if you think of that triangle that are sort of different aspects that go into what you can broadly say are power analysis. But it typically boils down to two questions. One example would be you know, sort of how many students are needed to detect an effect size of X. So, you know, let's say I have a particular intervention that's supposed to increase number of students who take the AP exam or get a job or go to college. And many other people have studied this, so I can sort of say I think my program is going to have this kind of an effect.

Well, then I could use that information and say, okay, how many kids in schools do I need to put into my sample? Another way you could ask a similar question is to sort of say, you know, I'm doing this program in sort of rural Iowa and, you know, all the

schools that are being impacted by the program, or it could be a comparison school to sign up for this study. So my sample size is six. Well, the question then is, okay, well, what's the smallest effect I'll likely be able to detect?

And what you can do is then answer that question and say, all right, what's the sensitivity of the data that I do have, and compare that against other studies and say, you know what, this is actually in line with other studies, so I think I have a pretty good chance.

And so again, just to sort of really kind of drive this home, this is the one thing you're going to tell somebody over dinner or take out today is when to conduct your power analysis. It really needs to be before data collection. Because that's when you're working with these hypothetical jars of jelly beans and endless possibilities and you're really trying to say, okay, what's the probability that this might happen and what's the probability of that this other thing might happen.

Once you've collected your data, you're not talking about the jar of jelly beans anymore. You just have your one study. And so, the power analysis based on just that one sample that you actually did achieve, actually isn't very informative about sort of the broader range. And that's kind of what that paper below there is really sort of talking about.

And so if you are in a situation where you have done a study and your chances were not very high, well, the problem is, is I think a lot of you know, time and possibly money has possibly been wasted, right? Because you don't know if there really isn't an impact of the intervention, or maybe there was an impact. But you didn't have enough data to actually detect it and get that small p value that you need to tell the world that you have in effect. You just don't know and it's hard to tell even after you've collected your data. And so this is why power analysis prior to data collection is really, really, really key.

So this is sort of a whirlwind and, you know, not a lot of equations, just sort of, you know mentioned stuff that's coming up in the other sessions that'll be recorded, but I just sort of want to sort of summarize what I hope the big takeaways from this session are. First that power is a chance. It's the chance of detecting your impact and for it to be statistically significant.

Poorly powered studies, you know, may not be able to detect the effect if the effect is a little smaller or more nuanced. And power really is the result of everything going into the study. It's not just a statistical analysis, but it's the design and everything. You know, if you're thinking of it in terms of your undergrad courses, it's both statistics and research methods. Both of those go into a power analysis.

Larger samples with larger impacts tend to have larger power. And then again, like I just sort of mentioned, our analysis must occur prior to collecting the data in a study.

As Carrie mentioned I wrote a book. So since these are my slides, I made mine bigger. But there's lots of really good resources at many different levels of equations and Greek and sort of advanced-ness. The little brown book, Jacob Cohen is sort of the classic in this area. It's kind of a phone book, you know, there's some text, but then it's just tons of tables because this was written prior to the software being out there.

Applied Power Analysis by Aberson is a really good on-ramp into the subject. How Many Subjects is another good book that kind of goes through a range of situations. And you know, the other book, the yellow book is really sort of a great phone book for all of, most of the designs that I see out there, so.

And then my book, of course to me is sort of an introduction, it basically takes your treatment versus control or treatment versus comparison paradigm, and really runs its paces. Whether it's a

simple sample, cluster sample and all the rest of that stuff. So, and there's lots of blogs and lots of resources that will be discussed in the other webinars. But I hope this has been helpful and interesting, and I'm happy to take some questions for larger discussion. So, thank you for tuning in this was fun. I love talking about this stuff, so thank you.

MARKOVITZ: Thank you so much, Eric, for this informative introduction to the concept of power and its importance in evaluation planning, and design. I hope everyone attending now understands both the role of power and its usefulness for informing evaluation design decisions.

So, as I mentioned at the beginning of this presentation, this class is the first in a series of three classes on statistical power. So we hope you will take the time to view these additional two classes when they are available because they will provide more context and instruction on the process of estimating statistical power.

However, still we would very much like to answer any questions you may have some going to pass this on to Dr. Andrea Robles who will facilitate our question and answer.

ROBLES: Well, thank you, Eric and Carrie, and as a qualitative sociologist, I want to say that this was great. I learned a lot and we do have some questions that have come in. So just to mention for our online audience, you can ask a question by typing it in the Q&A box or the chat box. You can also ask a question and we can hear from you using the microphone on your computer by selecting the raise hand feature from above, from the menu above and we will grant you microphone rights in the order the questions come in.

So the I'm just going to pick up on some of the questions that were asked. So, Eric, are power analysis only necessary when you were using a sample? If I decide to include the entire population

of respondents in my study, then do I still need to conduct a power analysis?

HEDBERG: This is a great question. This is one that has been discussed with rigor and a little bit of [inaudible] among the statisticians over the years. So essentially just like, you know, James Coleman, a famous sociologist says that social capital is defined by its function, whether or not you need a power analysis for your study is really defined by what's the function of the study?

As a general rule. I would say typically, yes, you want to do a power analysis for your study because typically you want to have a statistical test. Even if it's a census of the population. And I say this for a couple of reasons, one is if you are using treatment versus control or other covariates, you aren't imposing on the data. You're not going to explain everything in the data and thus you want to, you know, acknowledge the fact that you haven't explained every last bit of variance and so you're going to need to do a statistical task.

Another situation, even though you have census data in which you want to do a power calculation is if you want to take these results and speak to the future or speak to other data, you know, unless you are only discussing the data sets that you have in terms of that state or that set of schools, or, you know, if you're just sort of discussing that data, if your analysis is really data reduction, just to give people averages, and you're not making any claims beyond the data, that's like the one situation in which census data may not need a power analysis.

But more often I would say you probably want to make sure that you can detect your effect. That being said, I just want to, you know, for folks doing, AmeriCorps grantees or for folks in other statistical agencies in education and out there, everybody has an opinion on this and all funders and scientific boards have slightly different opinions on that.

So, with what I've just said, I just want to very clearly state that there's a big old "it depends" that I should preface any conversation about this way. So, I hope that's helpful.

ROBLES: Okay. Thank you. Let's see just reading through these, how is power analysis different from confidence intervals? Don't they tell you how sure you are and your hypothesis is correct?

HEDBERG: Ah, yes. So a confidence interval is taking information from any one sample and just you know, taking that information from the one sample and saying, you know, if I were to sort of repeat this sample many, many, many times, you know, what's sort of 95% of, you know, the range of what the answers could be. And so a confidence interval for any one statistic is taking data from your one sample and basically trying to say, okay, what do I think the balance are with repeated samples like this?

Whereas a power analysis is, you know, again, sort of thinking in terms of many, many possible samples,

but it's a little bit more future looking. It's a little bit more forward thinking. And the idea is you think, I think the impact that exists out there is on average this, which would be to assess a test statistic on average of this, and you're basically drawing a little curve around that and you're sort of saying, okay, how many of those possible future samples are going to be past that critical value, which would give me the small p value that I'm looking for. So they are very much related, but I think one is more forward looking and one is sort of more summarizing the results of any one study.

ROBLES: Okay, great. Thank you. So onto the next one, if there is a small sample size and a smaller realized impact, is it possible to extrapolate and assume that a larger sample size would have realized a larger impact?

HEDBERG: It's a good question. It's hard to tell. And oftentimes what I find is the smaller samples that do produce a statistically significant impact are the ones that got lucky. Whereas, and this is part

of what led to a lot of the discussions of the replication crisis in say, psychology. For those of you who don't know, I'm sure if you Google it the papers will come up, but essentially what they did is they took about a hundred psychology experiments and they replicated them. They did everything they could to replicate the conditions and got new data, had new participants. And I think only about a third of them that were, one third of the 100 that were significant before wound up to be significant again. And so this sort of was, caused a little bit of a panic within psychology.

And one of the reasons was a lot of those psychology studies had smaller samples, which means that, you know, the ones that were small samples that did get effects got a larger than typical effect. And so when you go in, you kind of draw again you didn't get the effect. I had a similar experience with blackjack you know, I was younger and I played blackjack and I won and I'm like, oh, I'm a great blackjack player. Well, I went the next day and I lost and I went a third time and I lost again.

So that first time I went to the blackjack table thinking I was some great blackjack player, I actually was just getting lucky. And on replication of my little study of whether Hedberg is a Vegas Maestro, I found out that no, I really wasn't.

ROBLES: Thanks. So this one has a few parts to it. Let's see, improving the power of an evaluation is important. Are there methods to address social power and equity? Who defines the inquiry, who benefits engaging evaluations subjects, participants in evaluation design, avoiding extractive evaluation practices? So, does that make sense? Small phrases together yeah.

HEDBERG: Yeah. When I hear that question, I see power being used in many different ways, not just a statistical power, but actually the social power that researchers hold and all the rest of it. You know, as a trained sociologist, not a ton I want to say about this, but I'll stick to sort of the statistical power. And mainly just to sort of say

that a lot of times when people are planning their study and their sample sizes, they tend to be pretty monolithic and what they're expecting. And obviously the world is far more complicated than that and so if you are going to do a study, and you do want to say see how effective a program is for a particular subgroup, not just everybody.

Well, a good rule of thumb that I like to tell people is every analysis has a corresponding power analysis you want to do. So your power analysis for the overall effect of everybody does not provide information about the power for the effect for the Latino participants in your study or the males or females or any other subgroup that you would want to create. So, if you are planning to do those studies and you have reasonable expectations as to how the impact may vary across those groups, you can do it. There there's a power analysis for that. Kind of like there's an app for that.

And so I think it's really important again, that power analysis is a chance to really think through

all aspects of your study and plan on them. Because a lot of times what people do is they'll plan a study for the overall effect and they may not get it and then they'll just start doing all the interactions, no subgroup stuff, even though they were planning on it the first place and they may get lucky a couple of times and all of a sudden, oh, you know, it doesn't work for everybody, but it works for this one group.

Well, I can go back to my Vegas story again, where, you know, you could've gotten lucky with that one particular sub analysis, so.

ROBLES: This next question is you might, it kind of feeds off of what you mentioned a little bit, but how does the movement against statistical significance impact the need for or use of power analysis?

HEDBERG: Okay. So, to summarize sort of for everybody, there is a large movement in many research discussions and in the statistical agencies and the American Statistical Association that are really

trying to rail against p values and significance testing writ large.

I can say that from my point of view, not speaking for any organization on the phone, that a lot of this is being driven by what's ultimately, it's sort of a misuse of p values and a feeling among a lot of the statisticians that nobody's using p values correctly so let's just stop and do something else. And so in that context, yeah, I think the sort of connotation of the question is right in that, you know, power analysis is a little less useful.

For a lot of the remedies that people put forth to, you know, moving away from statistical significance and p values, whereas it's a discussion of effect sizes, a discussion with confidence intervals.

Power analyses can be very informative towards those things as well. Another thing, and I won't, I'll just throw the buzzword out there for people who want to Google it, but a lot of people want to sort of move towards more Bayesian analysis as sort of the standard and move away from this sort of school

of significant. And I think there's a lot of use in that too.

But I also think that, you know, looking up what you think your effect is going to be, how precise, what's the variance going to be around that effect, which are all the activities that go into a power analysis are useful, whether or not you're chasing stars or not.

ROBLES: Okay, thanks. So here, I see a couple of questions on small sample sizes, so I'll start with one, and then I'll look at the other ones. Please someone who already asked a question and they say, thanks so much just to follow up on my prior question about smaller sample sizes and realized impacts what constitutes a small sample size? Is it a percentage of respondents or beneficiaries?

HEDBERG: So this is something that I hear come up a lot and unfortunately while the sampling literature in statistics has you know, a lot of things as far as, you know, finite population corrections and stuff,

if you are working with a limited group of individuals that could ever possibly be impacted by the study, the math is less forgiving as far as statistical significance goes.

You know, if there's only 10 people that are being impacted by your study, and you can only have 10 people in your study, you know, you're going to have a lot of statistical uncertainty with regards to your estimates. That being said I would lean personally, you know, and again, every institution has different approaches to this, but you know, it may not be statistically significant, but highlight the effect size. And we'll talk about that in later videos.

And, you know, sort of say, okay, it's not statistically significant, but boy, these people did a heck of a lot better than the control. And that's useful, if, in other situations, if you know, my cartoonish example of 10 people possibly being impacted by the program, well, then I think we need to stop and say, okay, why are we evaluating a

program that could only possibly impact ten people?  
And so on and so forth, and so that's less power and  
more sort of broader issues. But you know, I think  
you know, evidence takes many forms and I think it's  
worth talking to your stakeholders about it.

ROBLES: So this is asking a similar question, but I'm going  
to ask it and see if there's anything more. What if  
you're constrained by your sample size or the  
population size?

HEDBERG: Let me think about this. I mean, again, if you're  
so constrained that it's, you know, a very, very  
small number of individuals that could even possibly  
be in the study, then I think you know, it's worth  
having a discussion with your stakeholders and sort  
of really talking about is this the right avenue for  
some statistical inference. You know, again, effect  
sizes, other sort of measures you know, might  
provide useful information.

But, you know, at the end of the day you know, the  
SPSS or R or state or stats or whatever you're using

is going to be ultimately not as forgiving as you would hope it would be if you know, you're ultimately working with a small population and you have to infer beyond that small group of individuals.

ROBLES: Okay, thanks. Have questions in a few places, so I'm looking to make sure we have all of them. Do I need to conduct a power analysis if I'm only collecting information on the people in my program? In other words, I don't have a comparison group.

HEDBERG: If you are only looking at people who were impacted by your program and, you know, you want to do some kind of statistical tests, a lot of people will want to do a pre/post analysis, sort of a pair sample tests and my messages yup, there's a power analysis for that too. It's a different power analysis and this is one of the things we're going to really hammer home in the later studies is, you know, every button you have at SPSS there's actually an associated power analysis you can do for it.

Sometimes it has to be really complicated where you make up all the data yourself, but, you know, the general rule is every little type of analysis you have, there's a way to do a power analysis for it. So even if it's a pre/post with a single group, there's a pair of T tests. In fact I think I have a paper on that that I'd be happy to share with folks.

ROBLES: Thanks, let's see., this follows up the question on consent. What level of consent is required for power analysis? Should the program design a generic consent form, use an existing consent form from the school site, or is a more rigorous process used like IRB?

HEDBERG: This is a great opportunity for me to remind everyone that power analysis is best done prior to actually collecting any data. And so the power analysis is not so much an analysis of your data, but it's an analysis in that you're using formulas and software to plan for what the chances are that you're going to detect the effect with the data that at that point doesn't exist.

So you know, the data you need to collect for a power analysis is typically on academic reports and papers and library work and less actually collecting data from participants at that point.

ROBLES: Okay, thanks. And I'm again, reading some of the comments. So we have a Rebecca Shaftner [ph.], I'm sorry if I mispronounced your name, who is on and she said, thanks for bringing up the issue of reproducibility. This is actually my study, for folks who are interested in the reproducibility project, you can read the article and she has the article here. I don't know if you want to say anything to Rebecca.

HEDBERG: I love that paper. I use it all the time and you're doing a higher power's work for sure. And I think it's an issue that affects, I can't think of an issue that is analyzed statistically that the reproducibility crisis hasn't impacted. So it's great, it's a great piece. I don't think it's behind

the paywall. I think it's easily available from that link and I encourage everyone to pick it up.

And then there's other papers where people are arguing about p values and other things and The Economist magazine a few years back did an article about reproducibility in science and I don't know if they still have it, but they have a cute little animation that kind of goes through the logic of what I was lightly touching on as far as some studies getting lucky and so forth. So, thank you, Rebecca. I appreciate you being here. You know, please let everyone know in the chat, if I said something or misrepresented something.

ROBLES: And also people could raise their hand and speak up. So just in case for everyone who wants to say something rather than write it in the chat box. Okay, so here's another question. What if I don't know the effect size of my program? How can I calculate the power level?

HEDBERG: Another good question. So this is where some solutions is really helpful. I go into this a little bit in the book and I think there's actually a couple of slides in the later classes about this, but, you know, given our discussion with Rebecca about the reproducibility crisis, then, you know, we can infer that not a lot of reproduction has actually been happening over the past few decades of research. And so unless you are evaluating a program that's been evaluated you're probably not going to find good resources as far as what you can expect for an effect and so on and so forth. So what do you do?

My broadly unscientific, but sort of practical, approach that I talk about in my book and others is you want to kind of go at it from two angles. First is look at studies about your dependent variable, whether your dependent variable is increasing test scores and academics or getting jobs or not dying or whatever. Look at the outcome and look at how other interventions that are even remotely in the same

neighborhood as yours, how they have moved that needle.

That'll give you a sense as to sort of how much can this needle be moved on this outcome. And flip it, and then you do another set of a set of studies and literature reviews where you focus not so much on the outcome, but you focus on interventions like yours. Is it a coaching, is it a tutoring is it youth, is it a curriculum? And see how that's altered and impacted other outcomes. And so by looking at how much does your outcome change and how much does your intervention change other outcomes, between the two, you typically start to get a sense as to sort of what you can expect.

Like I said, this isn't a scientific approach, but this is the best thing I can think of to sort of, you know, get reasonable expectations.

Other things you can do with some academic outcomes are people have looked at how much students typically grow from say here to here on academics

and this sort of natural growth benchmarking and it's something I've been playing around with and doing for criminology too, this sort of natural growth benchmarking. It kind of gives you sort of, this is the upper limit.

For example, kids, young kids grow one and a half standard deviations in reading from kindergarten to first grade. And so somebody comes to me saying, I powered my study to detect, you know, a 10th of a standard deviation, okay, I'll bow my head and say, I think he got a good shot at that. If however, the same study says, I'm going to pick up a 10th of a standard deviation on high school reading in which natural growth is, you know, a 20th of a standard deviation, I'd say, are you going to double the amount of reading they're going to do in high school? And so I'd say that's less plausible. And so those three avenues are sort of where you can start to put many hands on the elephant to kind of get a sense as to what you can expect.

ROBLES: Thank you. Okay, I have another question. I just want to ask everyone to please go ahead and if you have any questions, just to put it in the chat box. But are there power analysis with statistical matching techniques for non-randomized studies? For example, not just looking at the size of your sample, but how well your comparison group matches to your treatment group?

HEDBERG: Ah, yes, it's a great question. I actually was able to get IAS [ph.] to give me a grant to study this and come up with some advice. So I'll give you a preview. If your sample is exactly matched, then you know, like with courses exact matching or you've actually, you know, exactly matched your treatment and comparison groups, the buttons you ultimately push on your software to analyze it are the same buttons you push as if it was a randomized design.

So really causal inference comes more from the design and from the statistical analysis. If you do other ways of matching, say it's not quite perfectly matched and there's a little imbalance, you know,

your treatment group's a little lower on the pre-test than the control group that will hurt your power a little bit. If you do, and I'm just kind of throwing things out here, but if you're doing things like propensity weighting, well, then you're now introducing weights into your data set and the same kind of design effects that you know, my friends at NORC worry about when you have, you know, wildly weighted to data also will influence your outcome as well.

I'm also looking into other situations for quasi-experimental designs, like time series, you know, if you're sort of following, you know, a bunch of schools for five years, something like say COVID happens, and that's not really an intervention, but you know, something happens and you want to see how that's changed. Well, there's a certain auto correlation there. If you have enough time points, power is going to be impacted by that as well.

So ultimately I really appreciate your question, Catherine. I used it as a moment to pitch my own

work, but ultimately I hope I was able to give you some answers and that if you've matched your groups, you pretty much, you know, assume the power can be what you think it can be.

My one small caveat with that is a lot of times when you do match the effectiveness of your pretests or your covariates changes a little bit. And so there's a lot of other sort of buttons and dials that go into power analysis for some of these things. And so when you start matching groups and you're using information from the population to plan your studies, you want to take a quick look to see if that information still applies.

ROBLES: Okay, thanks, just a couple of questions to all of you. I mean, you Carrie and Lily, I know you have seen our grantees evaluation plans and evaluation studies and obviously power analysis probably come up as an issue. So what advice can you give to people who are doing this kind of evaluation work for some of the lessons learned that you've seen by looking at these studies?

ZANDNIAPOUR: This is Lily. I just want to just point out as Eric said, and everyone, Carrie also emphasized, you know, it's really important to do this pre-work in the planning phase. So, I think one of the things that is of concern to us is it's really a shame if you sort of invest resources in an impact evaluation, which as all of you know, it's not necessarily kind of cheap to really go down the path of doing and investing resources in it, if you really will not be able to really draw conclusions or inferences about the results at the end of it. So my takeaway is definitely try to do this ahead of time. And I don't know if Carrie, you have other points you want to make.

MARKOVITZ: Well, you know, that is definitely an important point. And when we work with grantees on their design, especially when they're doing impact evaluations, because that's such a large investment of resources and time. We always work with grantees in a way that we want to set up their design in such a way that if their program is having an effect

that, you know, we maximize the ability of the study to be able to detect that difference or detect that effect.

And so the power analysis, like Lily said is a very important component of that, and it should be done during the design stage. It helps programs understand if they have the adequate sample. Also if, you know, somebody asked the question about, you know, what do you do if you don't know the effect size? And there's a lot of work that can be done prior to doing an impact study.

You know, at this point I'd like to mention, you know that AmeriCorps has this evidence continuum. And so, we oftentimes say when programs haven't done much evaluation work on their program, we start with a process evaluation to confirm that a program processes, that the program model is being implemented as intended. And then we may do a number of pre/post evaluations. A lot of times programs will do an annual evaluation of their program,

looking at a pre and post measures to see how participants or beneficiaries have changed.

And if they see an improvement, and then based on that information, you can estimate an effect size. And then after you've gathered that type of evidence on your program, then you may be at a point where you're ready to conduct an impact evaluation. So I think a lot of times I talk with programs that are in a hurry to do the impact evaluation, to do the more rigorous work, which is terrific that they have that attitude, but there's just a number of steps I think and people need to think of evaluation as a continuous tool that can be utilized by programs to help with program improvement and documentation, and to help expand opportunities and funding.

And so as your program grows and changes, so too can your evaluation activities. So, I guess that's just what I would kind of emphasize in general, and then the effect or the power analysis is a component of the design process in designing these evaluations.

ROBLES: Thank you. So back to Eric, I have another question. Would you say the pandemic has affected what a typical year's growth would be? How should comparisons be viewed in light of how challenging this year has been?

HEDBERG: I have a lot of thoughts that I'm not going to explicitly share because they're sort of thoughts and not based on rigorous analyses that I've done, or that others have done that I probably haven't read yet. I will say likely going to have an impact. I think in my own experience I have a twelve-year-old that the sort of socioeconomic status that my wife and I enjoy has impacted a lot of the decisions that we've made and so on and so forth.

And so I think all I can say at this point is just rejecting the null hypothesis and say, I'm pretty sure it's going to be different. I think it's going to be a lot lower for some, I think it might actually be higher for others. I just don't think we know yet. And I think all of the resources that are out there, including a lot of my own that I've

published, that help people sort of you know, power their studies and plan their studies. I think after a couple of years of the post COVID world, we're going to have to recompute all those numbers and just sort of see what the, how the world is today.

Prior to the pandemic, I could say a lot of the sort of impacts that I, well, not impacts, but sort of parameters that I examined set tended to be pretty stable, you know, from year to year. But obviously the pandemic has been a major shift in all the social aspects of our existence. And so, you know, there's no reason to think that you know, a typical year's growth is going to be the same after the pandemic as it was before. Which way it's going to go, I don't know. So that's why the folks in the field now should definitely publish their effect sizes and provide as much information to others as they can. And then my last, yeah, so anyways, I'll end there and stop babbling about the COVID stuff, but yeah.

ROBLES: Okay, so from listening to the conversation, Eric, and the different questions are there any other things you want to add that, you know, or something you want to touch on?

HEDBERG: Ultimately, I think you know, the best power analyses are ultimately arguments and the best arguments are based on some real information. And as you plan your studies, as you power your studies, you know, you're going to be asked for lots of parameters, whether it's what you expect the effect is going to be, or how much a pre-test will help you out.

You know, there's always an intuition to sort of move the dial on the power analysis software until it gives you the answer that's within your budget and then say after the fact, I really think I'm going to have a big effect. And so that's a temptation that everyone faces. And so really just every argument that you make in your power analysis you know, just try to back it up with something you've seen somewhere or pilot data that you have.

ROBLES: Thanks. I'm getting a note here that let's see Gwendolyn and Julia rose their hands. I don't know how to be able to listen to what your question is. So, if you don't mind writing it in the chat box, that would be great. Okay. Carrie and Lily any other comments from hearing the discussion?

MARKOVITZ: I would just add one more comment for the AmeriCorps state and national grantees that are out there. I just wanted to point out that in the notes at the bottom of the screen, there's information on how to contact us. We have an email address and there's also a TA portal where you can put in requests for technical assistance on the topic of evaluation. We help grantees with all types of needs and sometimes we can even just be a good bridge between grantees and the program staff, if there's a specific issue around evaluation, we can help facilitate an answer to that.

So, you know, we just encourage everybody to use the TA portal. We're happy to talk with people and help

them. We provide a lot of one-on-one hands-on TA and we provide it to all different types of state and national grantees. So I hope everyone will take advantage of that.

ROBLES: Thanks. Lily?

ZANDNIAPOUR: I don't have any other comments.

ROBLES: Okay. and if I don't hear from Gwendolen and Julia, I don't think I haven't seen their questions. Eric, any last closing questions as I look at the chat box for any last-minute questions?

HEDBERG: Just keep watching your emails for the next video if you are further interested in it and thank you for coming.

ROBLES: Okay, thank you. So yeah, I just want to remind people again, for those of you who tuned in later, that this is part of a three-part series on power analysis, and we will be sending out an email to all of you when we have level two and three and this one

on our website. We also have a handout of key word definitions that will, that's a companion piece of the presentation, and that will also be posted.

I'm just going to look one more time to see any last questions. Okay, so I'd just like to thank all of you again, this was really informative and from our response, I mean, we had 500 people sign up within a few days and so this is definitely a needed discussion and I thank you for the time that you've put into this.

And I also want to thank Jenelle from ICF Next International and her colleagues for helping us coordinate this. We will be sending out a post-webinar survey. So please let us know your thoughts on this webinar or ideas for the future. And again, we will have this posted within the month. So thank you and have a wonderful rest of your day.

ZANDNIAPOUR: Thanks everyone, thank you Eric and Carrie,  
bye.

**END OF FILE**