Defense Health Agency

Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury Webinar Series

"Technological Updates in the Treatment of Mental Health Conditions"

July 28, 2016 1-2:30 p.m. (ET)

Operator: Welcome, and thank you for standing by. At this time, all participants will be in a listen-only mode for the duration of today's call. Today's call is being recorded. If you have any objections, you may disconnect at this time. I will turn the meeting over to Major Demetrius Pittman. You may go ahead.

Maj. Pittman: Thank you. Today's webinar is Technological Update the Treatment of Mental Health Conditions.

Good afternoon, and thank you for joining us today for the DCoE psychological health July Webinar. My name is Major Demietrice Pittman, and I'm a clinical psychologist and psychological health subject matter expert at the Deployment Health Clinical Center. I will be your moderator for today's webinar.

Today's presentation and resource list are available for download from the files pod.

Before we begin, let us review some webinar details. Live closed captioning is available through the Federal Relay Conference Captioning. [inaudible] presentation slides. If you experience technical difficulties, please visit dcoe.mil/webinars and click on the troubleshooting link on the monthly webinar heading. There may be an audio delay as we advance the slide through this presentation. Please be patient as the connection catches up with the speakers.

This continuing education activity is provided through the collaboration between DCoE and professional education service group. All who wish to obtain continuing education credit or certificate of attendance and who meet eligibility requirements, must complete the CE evaluation. After the webinar, please visit dcoe.cds.pesgce.com to complete the online CE evaluation and download or print your CE certificate or certificate of attendance. The evaluation will be open through Thursday, August 11, 2016.

Throughout the webinar, you are welcome to submit technical or content-related questions via the Q&A pod located on the screen. All questions will be anonymous. Please do not submit technical or content-related questions via the chat pod. Participants are also encouraged to chat

amongst each other during the webinar using the chat pod. Please refrain from marketing or promoting your organization or product in the chat pod.

I will now move on to today's webinar, Technological Updates in the Treatment of Mental Health Conditions.

Evidence-based treatments are a requirement in clinical practice, and technological innovations in the delivery of psychotherapy are no exception to this rule. Clinical trial that use a null or waitlist control group are appropriate in the absence of an established standard of care. Technical adaptation of an evidence-based practice, however, negates the use of a null or waitlist control group. In this case, direct comparison of active treatments is required. One approach that is becoming more prominent in the literature is the non-inferiority trial. The goal is to demonstrate that an experimental alternative is no less efficacious when compared to the standard of care. Unfortunately, the reports of their application demonstrate confusion or a lack of awareness of the technical and philosophical nuances of this trial design. In this webinar, we will develop the tools needed to evaluate the quality of the evidence base to inform clinical practice. We will use trials of technological methods of administering psychotherapy compared to the in-office standard of care as a practical case study.

At the conclusion of this webinar, participants will be able to identify the key design elements of a non-inferiority study, interpret the results of a non-inferiority study, and evaluate the credibility of the evidence base for a treatment approach based on non-inferiority designs.

I would now like the introduce our speaker, Dr. Derek Smolenski. Dr. Smolenski is an epidemiologist with interests in applied multivariate analysis and behavioral epidemiology. He earned his advanced degrees at the University of Texas Health Science Center at Houston and completed postdoctoral work at the University of Minnesota. He has been a lead quantitative asset with the National Center for Telehealth and Technology since 2012. Dr. Smolenski has worked in several content areas including sexual health, alcohol abuse, behavioral assessment, depression, and suicide. He currently serves as a lead analyst for the Department of Defense Suicide Event Report and has been the lead methodologist on several clinical trials at the National Center for Telehealth and Technology including a non-inferiority trial [inaudible] serves as the impetus for this webinar.

Welcome, Dr. Smolenski.

Dr. Smolenski : All right. Thank you, Major Pittman, for that kind introduction. My name is Derek. I'm going to be giving you a talk today about how to best evaluate what is available of the literature in regards to the various technological applications that are out there and whether or not they should be incorporated into clinical practice based on what we know in the evidence base.

One of the biggest issues that you will have in evaluating this literature is that the design of the study that is used and what is being used more often is a little bit different from the classic trial design that you would have discussed in graduate school or an undergrad statistics course.

I am going to put out there right now that I am not a clinician by background, and so I'm not able to speak about some of the nuances of what the actual treatments necessarily look like, how they may affect specific factors such as the therapeutic alliance, whether or not people with certain types of conditions or certain levels of severity and symptomatology are appropriate or suited for using a technological analog to whatever it is that you would normally do in an in-office treatment environment.

What I am able to talk to you about today is what are the statistical factors in play? What is that people are really trying to do and trying to say when they do a study that compares an existing treatment to a new technological adaptation of it. When are people potentially cooking the books? Is the design solid enough for you to think that the evidence they presented is reliable and actionable? These are the things that we'll discuss. I hope that you're able to take away some key points from this and that you find it relevant to the work that you're doing in your everyday clinical practice.

The structure of today's talk is going to start with a pretty lengthy discussion around the non-inferiority trial in and of itself. What does it look like? How is it different from standard clinical trials? Why is it being used more now? What are the pitfalls that people can fall into? How does the design of this make it easier or more difficult to come away with strong conclusions.

Then we're going to actually look at a case study of recent papers that used this design in evaluating technological approaches to applying psychotherapy for primarily depression, PTSD, and obsessive compulsive disorder, and we're going to see how the claims made in the evidence stack up to what we learn about what we should be seeing in a non-inferiority trial.

Then, finally, we'll talk about some recommendations for what you can do in your clinical practice in terms of evaluating the evidence, thinking through what pieces seem like that they could be promising and that you want to try bringing in, also how to evaluate future literature that comes out. How to design your own research if you're going to engage in a research-type environment in addition to your clinical practice. Without too much further delay, let's go ahead and begin.

Let's see here. Just a second. Oh, I hit a right button. It's always good to know about technology when you're talking about it.

First thing, when you get into a clinical trial, you probably have experience with the notion of whether or not is there a difference between X and Y. You have two conditions. You're assigning people to drink a generic soda, and you're assigning people to drink a name brand soda, and you're trying to see is there any difference between these two groups on whether they like the taste of the soda. One way to do it, and what we would call this design is actually an equality study. Either the groups are equal, or they're not equal, and so what we're trying to look at in one way, and this is the two-tailed test, one could be better or worse than the other one. We don't know which direction it's going to go, so we're going to have two groups try this, take their satisfaction scores and compare them, and we expect the groups to not have a zero difference, but I can't tell you upfront if it's going to be positive or negative.

The other side is you may have some knowledge going into the trial that says, "I'm pretty sure people will like the name brand better than the generic," and so we're going to assume that there will be a difference. It's not going to be zero. There will be a difference, and it's going to be positive that the name brand people have a higher satisfaction score than the people who drank the generic, and this is where you do the one-tailed test. These are all variations on the same thing. The underlying null is that there's no difference that's going to exist between your groups, and you're trying to see if you can find one, either positive or negative or definitively in one direction or the other.

When you move away from this design, and the big reason that you would do so is you're normally comparing things against nothing, so we're going to give people an injection of gamma globulin versus an injection of saline. We're going to give people cognitive behavioral therapy for depression versus we're going to put them on a wait list for it, and that's where you want to see if there's a difference between essentially doing nothing at all, and in those cases, you're able to use this question of, "Is there a difference versus is there no difference?" Where you start to run into and ethical problem or as a science within any particular area advances, you start to have the emergence of accepted standards of care.

Standard in-office psychotherapy, there are accepted protocols now for a variety of conditions that you should be using as part of your clinical practice to then decide, "Hey, I'm going to create a mobile app," or "I'm going to decide to deliver this to patients via a video conference instead of right here in my office with me." It becomes very questionable whether or not it's ethical to say that you'll compare this video conference or this app to doing nothing for them at all. It would now be bound by ethics to at least provide standard of care as your point of comparison.

Well, this raises a little bit of a problem. If I'm changing the way I'm delivering therapy, do I really think that, by going from sitting in my office to talking to somebody through Skype, I'm going to make the therapy better? My protocol hasn't changed. The theoretical aspect of the

treatment that I'm delivering, they haven't changed. I'm just changing the modality, so do I think it's better? I don't necessarily think that's the case. Do I think it could be worse? Maybe, but I don't think it's going to be that much worse.

Well, this sets us up for a situation where you're expecting you would like to see a difference of actually zero, that it is no different. You're doing the same thing. The problem is, you can't prove the negative. You can't prove that it's zero. Mathematically, it's undefined, and so this is where the non-inferiority equivalence design really come into play, and this is where you say, "I'm pretty sure they're not different from each other, but I'm going to go ahead and establish upfront an amount of difference where this new approach could be a little bit worse than what's the current standard of care, but it has all these advantages for people who are too far away to come to the clinic. Specialists can provide care to more patients across a broader geographic range. We have these benefits for the practical application, and there may be a little bit of a decrement in efficacy, but it's compensated for by its benefits. I'm going to define upfront how much worse it can be but still be meaningfully no worse than the standard of care, so you're allowing for an arithmetic difference in efficacy but that is theoretically not all that important, and that's where you define a margin.

This is the Greek letter, d, that [inaudible], so the null is, if I take my new CBT approach through Skype and compare it to the in-office, the difference on the outcome score is going to meet or exceed that margin. It's going to hit the red line, and I'm going to say, okay it's actually worse enough that I can't in good conscience recommend this for my patient population versus to alternative which is whatever difference we found didn't hit the red line. It's trivial enough that we could say that it would be reasonable to do ahead and use this. If you get into the situation for equivalence, you're now saying, not only is it no worse than, it's also no better than. This is something you're going to see a lot in the pharmacological literature where they're trying a new formulation for a generic alternative to a name brand drug. You want to see that it is functioning within equivalent ranges for the processes that it's supposed to enacting in the body.

For our purposes, we're going to focus mostly on non-inferiority. For those of you that are more graphically focused, I provide a visual here just to capture exactly what we're talking about. For the superiority, you can be testing up and down the range from negative infinity to infinity in terms of a positive or a negative difference. For non-inferiority, you're focusing on this issue of, "Well, let's see. It could be a little worse, but it even could be better." We could look at both of them, and we'll talk about that in a little bit. The equivalence trial is the strictest one. This is where you're actually get to as good as. Note that in the literature, some will market a non-inferiority study is as good as, and I think that's misleading because I would argue that as good as means no better and no worse, which is what you get with equivalence. You're bounding it above and below. With non-inferiority, it's no worse or better. Well, if something is better, it's not

as good as. It's better. I would be careful in describing a non-inferiority study is as good as if you're trying to put it into lay language.

The goal that we're focused on here is to reject the a priori margin. That's that planned difference that we were discussing, and we're going to spend a lot of time talking about the margin because this is probably one of the most important parts of the design. If somebody cannot tell you upfront how much of a difference they are willing to tolerate in their study, it cannot be a non-inferiority trial, no matter how much they want to argue that it is. How do you come up with this magic number? This is a question where the methodologists don't have a uniform consensus yet. There are different approaches to take. They all have positives and negatives. There is no one definitive answer. We're going to talk about a few of the common ones and look at what the implications are for our study.

In some cases, people will focus mostly on clinical. They say, if we're using a BDI, a difference of 5 points is considered the threshold to say that there's been a clinical change in symptomatology, so I'm going to use 5 as my margin. If I can show that it's no worse than 5 points, well that means we haven't demonstrated that it's a difference that's clinically relevant, so we're going to say 5 is the red line. Others will argue that you need to take into account statistical considerations. This becomes a little bit more complex to fully understand, but I think we're going to be able to get there today.

The big issue is you don't have a null control. You have an active control. You're making an assumption, and in some cases a bit of a leap of faith, to believe that you're active control is behaving the way the literature says your active control should behave. If you go into a trial saying, "I know that CBT causes this amount of benefit for a patient population," you have to believe that that amount of benefit is actually happening in your population, in your study, for your active control, and now you have to come up with a margin that says how much of that benefit do we need to preserve? How much of the advantage from the standard of care do we need to maintain to say that we really believe that it's okay to do this other alternative that was worse but not really that much worse.

At the end of the day, folks will argue that, in either clinical or statistical approaches, you need to come up with a way to define the minimal expected effect, that minimal change that you're going say becomes important, and that's where you draw the line.

We already talked a bit about the clinically meaningful difference. Problems associated with this is we are ignoring statistical issues. What if you're going against a standard of care that only benefits you by 5 points, in general. Well, you've now set a margin that's 5 points which means you're allowing your treatment to be 5 points worse than the standard of care that's only 5 points better than doing nothing at all. That means you're allowing your alternative treatment to be no better than doing nothing at all but still be no meaningfully worse than your standard of

care. This is where the statistical issues have a little bit more weight, at least in my opinion, for designing the study and determining what your margins should be. You need to be aware of how much the standard of care is supposed to benefit the people before you determine how much of that effect you need to maintain. I can't believe this is the margin.

The way you would do this from a statistical perspective is to go to the literature and find a meta-analysis, and if it doesn't already exist, this is an excellent opportunity for you to do a meta-analysis and get an extra publication out of it, especially for those of you on the tenure track. The focus here is, okay, I'm going to get a uniform estimate that tells me this is my expected difference between doing nothing at all and doing the standard of care. I'm going to say use the lower bound of the confidence interval to say that's your safe bet on the expected difference. It's conservative, but it can be quite useful because that is still well within the range of an expected difference, so take the lower bound of the confidence interval.

Now take a portion of that confidence interval. The proportion that you chose will be inversely related to the amount of the effect that you preserve. The smaller proportion of that confidence interval that you retain, the more the effect you're preserving, and this is useful because let's say a standard that's in the literature. We're going to preserve half the effect of the confidence interval. Well, what does that really mean? It's half as good. Sure, I'll take a headache medication that's half as good as the one that's generally available, or I'll provide treatment to patients that's only half as good, but I can reach more of you. From a population standpoint, there could be arguments made in favor of that, but from a clinical standpoint, I think patients would take issue with a notion of you're going to provide something that's only half as good, but it passed the statistical design. You can do it that way, but be aware of what it means.

What about 25%? In this case, we're preserving 75% of the effect. If the study is successful, this means that it is at least 75% as good as the standard of care. That's a bit better. Down side? It becomes a much stricter test to pass which is going to make the design and the execution of the study harder to do, so of course there are going to be trade offs. The focus though is be aware of what's going on. Look at what somebody says is their margin and look at how they defend their selection of it. There should be a thorough assessment of either the existing literature, the clinical literature. There should be a discussion of the statistical factors. If they just say we used a clinically meaningful difference, they did not consider how much effect to preserve at all, so beware.

Again, to try and give you a bit of a visual, here's what we're talking about with this. In this case, we set the margin as the observed difference from a meta-analysis. We could find a worsening of effect all across the red-bracketed range, but at the end of the day, we've now ruled out preserving any effect of treatment, so even if our finding is favorable, we get what we wanted to see, we're effectively saying is that, while it's no

worse than this margin, it's probably no better than the null. It's no better than nothing. That's a distinct possibility. You are opening yourself up to that. The benefit of doing something like this is you're probably going to have a wide margin which is not going to burden your sample size very much.

Let's go to the lower bound of the confidence interval. Again, you're probably not preserving any effect against the null. You are preserving a little more than you would if you used the point estimate from a meta-analysis, but you're going to have a bigger sample size here because it's going to be a smaller margin between zero and the red line, and at the end of the day, since there's variability inherent in that meta-analysis, you're not going to have much of a leg to stand on in arguing that, even though you passed a non-inferiority test, it's really unclear how much better than doing nothing it really is.

You go to preserving half of it, your margin is getting tighter, so this is going to be increasing your sample size. This could be increasing your study complexity, but look, you have ruled out half of the effect that you would expect against a null, so you at least can come away with this study saying that you're pretty sure there's some benefit versus doing nothing. This is a big payout, a big advantage.

If you go to a smaller one, you're now preserving 75% of the effect if you use 25% of that confidence interval the half width there. Again, a benefit that you're now going to get to probably a very large study because of the small difference. Trade-offs, implications, just be thoughtful of it.

The final words I want to use on the margin, you will see a mixture in the literature of people defining their margins using [inaudible] as a standard metric, so it's standardized difference [inaudible] 0.5. The problem with this is that dependent on sample-specific standard deviation. If there is a difference of three between two groups on a study with a standard deviation of 6 with a d of 0.5, but if the standard deviation went up to 10, well now it's a d of 0.3, the same raw difference, but you're seeing two different values in the standardized metric, and if you've defined you margin using the standardized metric, your sample standard deviation is going to have a huge effect of whether or not you reject the null hypothesis in your study. I would caution away from relying on a standardized difference as your defining point for your margin.

It has to be a priori as we said you have done the homework before going into the study to be able to say that you have a red line and you can defend it, and remember preservation of effect, be mindful of what it is your saying. How much of the difference that you would expect if you did the standard of care do you think your preserving so that you can protect against the idea that what you found at the end of the day may be no better than nothing at all? So just take caution.

The way that the results tend to be interpreted, the contemporary guidance is that we focus with a 95% confidence interval we're looking at the lower bound if the margin is below zero, so in this case a lower score is better on the outcome measure. If your study group actually has ... I'm sorry. A lower score in this case is worse. Even I can still get confused with this sometimes. It is a bit of a backwards design from what you're used to, so it's an ongoing process. Bear with me here for a second. If you have an outcome situation where a higher score is a worse outcome, then comparing the experimental group to the standard group, your difference would be positive, so you're going to have a positive margin, and you want to make sure your upper bound of your 95% confidence interval is below that margin, so this is where you're bringing variability back into play.

Conversely, if lower scores are worse, then if your experimental group had a lower score, and it was in the negative territory then your standard of care group, then you've got a lower bound margin that you're bouncing things against.

One of the benefits of using the 95% confidence interval versus the traditional approach which was a 90% because you're focusing just on the one tail looking at where you set the red line and only that test of whether or not your confidence level crosses the red line, you can actually look at superiority. It's possible that you've come up with an experimental approach, some new technological advance to your treatment that actually improves your clinical outcome. In a non-inferiority design, you can test both for no worse than or better, and so these are some scenarios that you could come up with at the end of day with your analysis, and this is one way to go about interpreting them.

In the first one, the blue dot, we see that the estimate as well as the confidence interval are completely in the red area. This is where we favor the standard treatment, so the in-office approach to psychotherapy. Worse off, the confidence interval is entirely above the red dashed line, which is our margin that we've established. This is a case where non-inferiority can't be rejected. The null hypothesis that you achieve can't be rejected, and you clearly have a case for inferiority, so this would be a fail in terms of the treatment that you want to provide via a technological approach being no worse than, very much a fail. In the two cases next to that, the purple dot, the point estimates in both of these are favoring the in-office approach, but the amount of difference between them isn't too much.

In the first case, your point estimate comes up right at the margin, so you have failed to reject the null in a non-inferiority test. Your upper bound confidence interval that [inaudible] is well above the red line. Same for the third dot here. The confidence interval is well above the red line, so you're not able to come away with that conclusion of non-inferiority. You could argue for the third dot that it's a sample size problem since the point estimate is in between the black line and the red line. Well, if we just had

more people, we'd have a smaller confidence interval, and we could get there, which is the situation you have with the first green square. You have a very, very tight confidence interval.

This is the fun mouthful of we reject the non-inferiority null hypothesis, so we can say it's non-inferior. The interval is below the red line, but the interval also excludes zero, which is the range of inferior, so this treatment is at the same time inferior and non-inferior, which sounds like it can't work. Something can't be a cat and not be a cat at the same time, but because we cannot directly test to prove that the difference is zero, we have to use the range in that margin.

When we get to the two final green squares, this is where the point estimates are favoring this new treatment, so we're moving in that direction. In the first case, our confidence interval does not hit the red line, so we can reject the null hypothesis, but we can not argue for superiority because we're still covering zero. In the final case, the confidence interval excludes zero as well as the red line so you've now got your two tests of non-inferiority as well as superiority, and this would be a shout from the rooftops finding for the treatment that you're testing.

Sample size? I don't think we're going to beat this one too much. You have your typical alpha and beta. You have your margin. You have the actual difference between your point estimate. This is a fun one that a lot of people don't pay attention to. It's unrealistic to think that at the end of the day, when you take to arithmetic difference between two groups on an average score, that you're exactly going have zero. Here's going to be some amount of difference between them. Consider a little bit of difference that favors the standard in your calculation, and that can help you have a more robust study design at the end of the day. As we talked about conventionally they use the one-tailed test and a conventional power of 0.80, so these are standard across trials and literature.

Here's the problem with that though because we find problems with everything now. Your errors are in reverse, so in the superiority-inferiority equality study design, you were concerned for a type I error that you are going reject your null when there actually isn't a difference between the two groups, and then your type II is you're going to fail to reject the null of no difference when there actually is a difference. Well, it's completely reversed here, so your type I error, you're going to end up rejecting the null of inferiority for something that is inferior. That's bad. You're going to end up saying that something's no different when it's actually worse in your type I.

Then your type II is you're going to fail to reject the null for a treatment that is not inferior, so in this case, it's not different, but you end up in a situation where your data show that it actually is. The type I error is worse, and it could have really bad implications for clinical practice.

Why is that important? This is sort of the last big conceptual piece to think of before we get to the real heart and meat of the content discussion. The traditional thinking and research, non-differential error bias is to the null. You're going to have mixed effects in study groups. Some people are going to do everything you tell them to do and follow the timeline precisely. Other people are going to start the study and then drop out because they don't like it. They're only going to show up for a couple of treatment sessions and keep extending the timeline because life happens, and they can't make it into the clinical environment on a regularly scheduled basis. People are going to move.

All these things happen, and they will dilute your ability to detect an effect, and as long as that's happening in a roughly balanced way between your two study groups, what this ends up doing in diluting any difference that you would detect between them, and so it pulls that detectable difference towards zero, and so in a traditional equality study, your null is centered around zero in which case that non-differential error is biasing you toward the null, biasing you towards zero. You're going to end up more likely committing a type I error because of that non-differential error, and that's okay, and that's what makes it conservative.

It produces the opposite problem for a non-inferiority because what we've been looking at is our null is a zero, but it's a line somewhere away from zero, and we're trying not to touch the line with our confidence interval. If you have all that error in your study, you're going to need bias in your estimate towards zero, and since that becomes the center point for your confidence interval, you could be moving the end point of the confidence interval artificially away from the red line that you've drawn, and so you're now going to conclude non-inferiority when inferiority may be the truth. I hope that makes it. That's what's going on. This is way it's an important consideration.

Here's a graphical example, so in truth we have a situation where the experimental treatment is worse than the standard of care, and if we do a study where there's low error, we can get an approximate point estimate and a confidence interval that would allow of to make that conclusion, but we introduce measurement error. We introduce people having idiosyncratic issues in both our study groups, and we see that the point estimate is getting pulled toward zero, and you'll notice the confidence interval is moving there as well. The saving grace here is that we had enough error that it also made our precision pretty poor, so we can't reject the null of non-inferiority, so we would still come away with the right overall conclusion but with a potentially rosier picture of what it could have been in a study with a little less error, so bear that in mind.

The final word on analysis. Per protocol and intent-to-treat, both of these are used. Both should be used. Both should be reported. If you're reading an article where they're not both reported, start thinking about bias issues, in particular the per protocol analysis should be there.

Intent-to-treat is the standard. This is where you get the benefit of randomization. You have not biased your study by selecting which people are going to get which treatment. If you have a large enough sample size, you will get the asymptotic benefit of randomization which balances your study groups. You can make cleaner causal inference[inaudible] to be conservative, and as we talked about conservative means moving towards zero, which in our case is not good when we're testing for a red line that isn't zero.

On the other hand, you have the per protocol approach. The benefit here is you're actually comparing two groups as you would have implemented the treatment, though in per protocol, you're taking out treatment non-completers. You're taking care of people who weren't compliant with all of the aspects of the study. You're maybe dealing with people who provided full beta on measures.

What this tends to do is it tends to strengthen the differences you can detect between your treatment groups because you now have an idealized group on the experimental side and an idealized group on the control side and so what ends up happening is it takes away a lot of that error. It enhances the difference you can detect, and in that case, it pushes your estimate away from zero, away from where it probably would have been in the intent-to-treat model, which gives you a better test of non-inferiority because, if there's not much difference between the per protocol and intent-to-treat, and it's close to zero and you're interval is reasonable, you're probably going to reject non-inferiority in both.

If you see a big difference between the intent-to-treat estimate and the per protocol estimate once you've cleaned up some of that error that's affecting your outcome measures, then you have a situation where you may have some serious bias that's affecting your ability to draw conclusions. The down side though is of per protocol is you've lost the benefit of randomization because you're now cherry picking the people who have done really well, the people who are doing what you're telling them to do. They could be different from who you randomized, and the way that they fell out in your study groups, the real issue, the way they fell out in your study groups may be different, and so you no longer can say you're doing a pure apples-to-apples comparison, so you have an issue where there could be some selection bias contributing to that point estimate. Safeguard for this would be consider using a regression model where you throw in a covariance to re-balance the group, ideally, maybe consider an observational study approach where you using something like a selection model or a propensity score to re-balance your study group on covariance, so that that's not driving differences between them, once you've taken away the benefit of randomization because you've dropped both out.

Big things for sample size then, and again this should be very explicit in anything that you're reading in the literature. You should be using a two-tailed alpha. You should set your power at 90%, so your beta equals 0.1,

and 1- beta should be 0.9 in this case, but the big thing to focus on here is, so your numerator's going to get a little bigger because you're going to use a standard alpha, and it should be two-tailed. You're also going to have an issue in the denominator because it's both how much of that difference you think you're going to find arithmetically as well as the margin that determines how big your sample size is going to be, and the smaller the margin, the bigger your sample size requirement.

Technology and psychological help, the real reason you all are here. What we're going to talk about here is what are some of examples that we see in the literature, and what is it telling us about where we're headed in terms of technology and what we can use or not use in clinical practice.

As we've discussed, you know have the tool kit going for the non-inferiority trial. We have that ethical concern about whether or not we can use a null or waitlist given the plethora of options that are already standard of care. We have the robust evidence base for that. We're going to use an non-inferiority trial so it's standard of care versus an experimental option. I found 11 studies that were recent that used Telehealth in the context of psychological treatment for behavioral health conditions.

Here is a listing of the studies that we looked at, mostly depression, a few with post-traumatic stress, a few with obsessive compulsive disorder, and it looks like insomnia and social anxiety. In each of these, they used an established treatment protocol, but they modified it either by doing it over video conferencing or using a web app or some form of web-based component in the treatment.

Three of the studies used the video conferencing, and in this case, it's individual therapy through video conference versus individual therapy in the office setting, so a direct head-to-head comparison of the same protocol, just modified by the modality. Two studies that were on the telephone using the CBT for obsessive compulsive disorder and then the self-paced, internet-based one. This is where they've created the protocol content in an internet environment. Folks are accessing it, doing it on their own versus the in-office . The comparisons as we were hinting at a bit ago for video conferencing and telephone looked pretty good. In terms of you're isolating the technology components of the study so that you can really study that piece of it versus mixing the exposures which is what tended to happen with the self-help or internet studies. You had an issue with the content may not have been exactly the same because it's hard to convert an in-person therapy protocol to a web platform that's self-paced.

The other thing that happened though, and this was fascinating, was you have self-paced internet resources being used for this treatment study. Again, in-person group therapy, you've now completely mixed your exposures. What would drive any difference that you would detect. Is it the internet versus the non-internet, or is it the individual therapy versus group therapy. You've mixed it, and we did we talk about that happens

when you start mixing effects, you start mixing what's going on, you're adding more noise to your study which biases things towards zero, which in the non-inferiority trial, makes it more likely that you'll come away with the conclusion that you wanted, the [inaudible].

Let's take a look at some of the methods. Are these articles reporting what they should be reporting for you to make an informed judgment about the quality of the evidence. I'm not going to worry too much about the number of tails that are used. Again, they are still valid references that say one tail is fine. Contemporary ones say two tails is better. Remember, when studies are getting a report into the literature, this is after years of work on it, so they designed it under one [inaudible] framework years ago that may have been updated by the time of their publishing, so that's just the nature of science.

The biggest one, the real crime against science, is that there are two studies that [inaudible] themselves as non-inferiority but did not mention their margin. In this case, they found no difference, and because they found no difference, well, that means it's not inferior. That's absolutely not what that means. It just means you found no difference. If you didn't specific a margin, you're not testing against that margin, so you can't do this [inaudible] and still come away or rename it, rebrand it, and think that you're going to get a meaningful finding. You're cooking the books. This becomes a situation where that evidence is not very strong, and I'm not going to consider those two studies moving forward.

Reported power, only one study, the Acierno from 2016, and it's not entirely surprising that they would, they have the power of 0.9 which is what we said we wanted it to be, but I'm assuming this study being one of the most recent also had the best chance of having exposure to the updated methods literature.

The findings at post-treatment, so this is where we're going to spend a little bit of time to understand what's going on and what we think about the interpretations and the study quality. For the nine studies that have an established margin, we can see what the margins were, and these are to the extent possible in the raw difference metric, so this is point on the measure that they define.

One exception is the Egede study. They took change on the BDI, converted into groups, those who improved versus not improved, and then used a percentage difference as their test statistic and established that as the margin, so it's a 15 percentage point margin. The other is the Yuen study at the bottom where I was only able to find ... They explicitly defined their margin using the standardized d of 0.42.

We look at the baseline standard deviation for the control group. This is how you can get to an approximate d statistic. How much of a difference is there between the groups relative to the baseline standard deviation in the control group? That gives you an approximate d. Now remember our

discussion from earlier, if you're using your margins in the raw metric, they become fairly easy to replicate for other studies because they can use that same raw measure irrespective of sample standard deviation, but if you look at these studies, you now have ... d is all over the place.

Once we take into account the variability in the study data, we have differences from 0.32 to all the way up to past one. Now it can be difficult to appropriately interpret a [inaudible] statistic in terms of magnitude, but when you start thinking about a whole standard deviation movement from one point to another point, that's a large change. It really is. It's a large change, and if you're saying that that's your margin, that's really, really generous. You're now saying to people that they can be a whole standard deviation or more worse off than the standard of care.

Most clinical trials will use, if they're using a standardized metric, just to do a general power calculation for a traditional equality study, they'll use a d of 0.3 or 0.4 as a desired effect size to detect to say that it's meaningful but to then turn around and say you have a margin that's one and that that much difference is not meaningful, I doesn't really compute conceptually. So be mindful of that.

I would say one of the better ones ... so there are two that use the BDI, so the one percentage one with 15%, you will see a lot of percentage if you're in more of the biomedical literature. Fifteen percent was used for a while. I think they're moving toward ten percent as a general rule, but there are variations in it based on what you expect the base rate of the outcome's occurrence to be overall, and the BDI one ... studies will use a d of 0.5, which is a five in the raw difference metric, as the red line. The Ly study used 2.5, so they're saying in their defense that it's going to be half of the expected clinical benefit. They didn't take into account what the actual benefit of the standard of care was, but they at least were trying to preserve some of the clinical benefits because if you say BDI and use a margin of 5, you're saying, we'll rule out a clinically important difference as not being important in terms of decrement of effect.

If you stop and think about some of the approaches to non-inferiority that are in the literature, sometimes you can walk away with a bit of a headache in terms of what is it really implying at the end of the day? No surprise, looking at the estimates across the board, we find that eight out of the nine studies here were able to conclude non-inferiority was achieved. The confidence intervals were either above or below the margin depending on the direction of the estimate.

The one that failed actually was the Ly study, the BDI, a margin of 2.5, so they cannot reject the null hypothesis of non-inferior, and if you notice, her point estimate was a 2.42, so that was right on the money in terms of the point estimate hit the margin of no difference, so this isn't necessarily even a question of power or sample size because if it doesn't effect that point estimate, you would have to have a confidence interval of effectively

zero, so reject the null hypothesis is non-inferiority because this is one that says I hold [inaudible].

Other observations from the research, as we've already addressed in brief, the absence of a statistically significant difference does not mean that there's no true difference, and it does not mean equivalence or non-inferiority. If those are not designed a priori then you cannot really address it with any credibility.

Equivalence and whether or not it's the same as non-inferior. Sometimes we'll try to say that a successful non-inferiority test means that the two treatments are equal. No, it just means that one is not appreciably worse than the other one, but they're not necessarily equal and especially if you're not testing for that, don't talk about it. Try to align what your reporting and what you're able to report with what it is you did in your study design, and it's you as a reviewer, as a clinician, as somebody who's judging the quality of the evidence be mindful of they're describing what they're finding given what it is that they did and what it is they reported. Again, I can't stress this enough. If a margin was not defined upfront and explicitly stated and ideally defended, then they're going to have a really hard time selling as evidence of non-inferiority.

Let's see here. Other things. Heterogeneity and margin width. This [inaudible]. We know there's no one great way to define a margin. We know that there are things that need to be considered, but as a reviewer of the literature the consumer of the scientific literature, pay attention to what they set as a goal post, and you can make any judgment for yourself. Is it reasonable? What does it mean? Is it taking into account what we expect the standard of care to be doing? Is it preserving any effect, or is it really ruling out, or is it allowing for a difference that could wipe away any clinical benefit once you just looked at the standard [inaudible] You've got this notion of the sample size, so a lot of the studies that we looked at ... One of the urban legends around the non-inferiority is that it increases [inaudible].

This is possible but not always true. When we looked at the slide with the sample size formula, that denominator, is how much difference do you think is actually going to occur, which a lot of people assume is just zero, and that just leaves you with the margin squared. This is the same as in a traditional trial where it's just the difference squared, so you can use the same sample size as long as you're willing to believe that you will have no arithmetic difference at the end of the day when you have conducted your study and analyzed your data.

I would say play with it a little bit, allow for that beta to punish you just a little bit so that you're protecting against the fact that there may be an arithmetic difference that says you're a little bit worse, but you have made your confidence interval a little bit tighter as function of your sample size so that you could still come away with an appropriate conclusion if it is deserved.

Another problem though to consider is, with the sample size and with trials, you are always taking into account this notion that randomization will yield unbiased comparisons. Again it is an asymptotic property of randomization that would be randomizing really, really large groups then, yeah, they probably on average will be quite comparable to each other. If you're doing a small study, even if it's what your sample size calculation calls for, based on what it is you're looking for, you may still have imbalance because you've got a smaller sample size, so you're not hitting that asymptotic moment of randomization, so have people pay attention to their study group differences. It's okay to go ahead and consider incorporating observation study method to improve group comparison even under the assumption that randomization should have handled everything. Be mindful of it. Otherwise, you may be getting some bias, and it won't necessarily be possible to tell which direction that bias is affecting what it is you're dealing with.

Most studies relied on the clinically meaningful difference, so again we've discussed this. It's useful from a clinical perspective, but it doesn't take into account the statistic issues in play, so we don't know how this new treatment would really fair against a null waitlist control because we don't know how good the standard of care is against the null waitlist control.

Two studies did use the lower bound of the 95% confidence interval, but in that case it meant they preserved zero effect, so they used statistical consideration, but they still walk away with potentially having a treatment that is non-inferior but up to and including the point at which it's not longer any different from the control if you had a null control.

Other things that we need to be mindful of moving forward with non-inferiority, so the first is the assumption. We have to believe that whatever our active control group is is a faithful representation of what literature says it should be.

If you're doing a trial, you have followed your protocol, you have established what the theoretical components are for your treatment to be effective, and that it matches what's done in the literature, what is used in the body of evidence to say that doing dialectic behavior therapy for Axis II conditions is an evidence-based treatment, and it should produce this much benefit, on average, per patient population. That only is it as good as the data that goes into it, and your ability to replicate it is only as good as you are following what was established as a protocol that drove the studies that contributed to the evidence base. Any deviations from that and you've kind of entered the wild west in terms of knowing that what you're doing control is actually having a benefit above and beyond doing nothing at all. It's usually reasonable to assume that. If you're using a protocol, even with minor deviations from other established protocol it's probably okay to assume that but know it is, at end the end the day, an assumption and an untestable one because you don't have that null control group in the study to compare against. This is where it flows into that effect of meta-analysis. If you're consulting it, or better if you're doing

it, it's constant, but we know that there's going to be variability, so something to be mindful for.

The final thing on the assumptions, you want to be using the best available standard to really say whether or not something is non-inferior because if you're using something that's mediocre then it's easier to find non-inferiority because you could be comparing mediocre to mediocre. You want to be comparing your experimental approach to the best possible comparator, the one that you believe is going to produce the best effect for your clinical population, to really give it a hard test or whether or not you crossed that red line. Your margin has to be reasonable. It has to makes sense. It has to have meaning. It has to be defended. It has to be well described in the literature, and if you walk away from an article with skepticism about the margin, then I wouldn't necessarily be overly excited about the findings from that research.

Other recommendations on this, best measures available to reduce error. All the studies that we looked at that were addressing behavioral health outcomes, they were using established measures from the literature, so we have good reason to believe that measurement error is not going to be a huge problem in contributing to the differences they reported.

The calculations and sample size, some did a better job than others. This seems to be pretty common in the literature. Consider are they using the best approaches that are recommended now. Are they using something else? Again, whatever it is they're doing, it should be defended.

Mixtures of exposures, it's really unfortunate that we have trials looking at this web-based approach to treatment compared to in-person group therapy because they come away with a finding of non-inferiority in all of them and so that should be exciting and say that this web-based treatment is not appreciably worse than the in-person group approach, and that is the finding, and that is what it says, but we have no idea of knowing whether it was the web-based component or the individual component versus the in-person versus the group. It wasn't a factorial design so you can't separate those effects and worse, because things got mixed, I'm highly dubious as to the strength of an assertion of non-inferiority because we've got that mixture that we're pulling the differences that we could detect towards zero.

That brings us right to the point where, with non-inferiority, a bad study has a stronger chance of giving you the outcome you anticipated than with the traditional clinical trial. If you've done a bad study in a traditional randomized equality study, you're likely to end up under-powered, you're probably going to have a lot bias toward the null because of error, and potentially an imbalance in other things in the design so that is going to make it more difficult for you to come away with a statistically significant finding and say that there's some benefit to this.

Here the opposite is true. If you've done a poorly designed study and you're biasing things toward zero, you're making it much easier for you to avoid the red line and come away saying that it's not appreciably worse, so this is why I have spent so much time harping on these very tedious and esoteric considerations of the non-inferiority design because it is so critical that the design be reasonable, that the decisions that the study team makes are appropriate because that's going to drive their ability to come up with a valid conclusion even more so here than in a traditional clinical trial. The more that journal reviewers, the more that clinician reviewing the evidence base [inaudible] be aware of and have in mind the necessary components of the non-inferiority setting the better our decisions are going to be, and falling from that, the better our treatment outcomes can be.

I am hoping that, for those of you still awake, during this webinar, you learn to identify the key design elements in a non-inferiority study. You will always think of this webinar every time you hear the word margin. You can interpret the results so you can see what is reported in a non-inferiority study. You can look at the margin and you can come away with an understanding of what the margin means and how they drew the conclusion that they did and you can evaluate the credibility of the evidence base. That, at the end of the day, is my hope for all of you, that you can look at an article objectively. You can review what they did. You can see if there's evidence to support this, and you can make a determination based on your clinical expertise. If there's evidence to support a web-based modality or treatment approach or using video conferencing, using telephone approaches to help clients, and it's something that you're comfortable doing. and it is something that your clients would be comfortable participating, then I would love for you to be able to really have the freedom to make those decisions, be able to defend them using the literature and using it in a very knowledgeable way.

I believe that pretty much sums up the material that I was going to cover today. I know we have about a minute thirty on the clock, but I think we're going to be okay, so I won't read you my reference list, but it is there for any of you who want to, and I highly recommend, read the studies in here, especially the ones that focus on the non-inferiority and following the equivalence design if you want to get some more details probably some more eloquent explanations, some of nuances that we discussed, but I thank you for your time and attention, and I look forward to a vivacious discussion here shortly. Thank you.

Maj. Pittman: Thank you for your presentation, Dr. Smolenski. We've had some great discussion already in the chat pod, but it's now time to answer questions from the audience. If you have not already done so, submit questions through the question located on the screen. We will respond to as many questions as time permits.

One of our first questions that we had was, could you explain the central question in clinical trials?

Dr. Smolenski : Sure. The central question in a clinical trial is a question of difference. We make an assumption outright that says okay the safe guess is no difference. We want to see if there is a difference, so we're going to test against a notion of no difference. You randomize folks, and you look to see is there an improvement in an outcome? Is there a worsening of an outcome because that can happen sometimes especially when you're experimenting with new things. For the non-inferiority and the equivalence design is, we know there's going to be some difference. In all likelihood, some non-zero difference will occur. Can we show that it's within an acceptable bound that we're willing to say it's no worse than, or for an equivalence design better than as well, than the established standard of care? The fundamental question is difference. The way we treat the difference and what we're looking for in the difference will vary a bit depending on the study design, but you're trying to come away with the strongest test that you can of whether or not a difference exists in the first place.

Maj. Pittman: Okay, well thank you so much for that answer. We have another question is how can I find out more non-inferiority studies?

Dr. Smolenski : The best way, I think the starting point I'd like to give is in the references, there were a couple of papers that I mentioned. I'd like to draw particular attention to the reference Greene et al 2008. I thought that was a really well-done paper that speaks to a broader audience, not just statistics wonks. For those of you who want a slightly more rigorous treatment, the D'Agostino article from 2003 that's in the reference section, it really lays out a lot of the concepts well and will give you some of the more technical discussion about why the procedures are established the way that they are.

Maj. Pittman: Well, thank you so much. Sorry, we're trying to read the next questions. Can you discuss any ethical issues related to non-inferiority trials?

Dr. Smolenski : Sure. The biggest ethical issue around non-inferiority trial relates to the selection of the trial from the outside. If you going to do a non-inferiority trial, you have already identified and established believable standard of care. That has to be there because that's your point of comparison. You cannot use something that doesn't have its own evidence base as a stand in for an after treatment that will serve as your control against experimental therapy. It's not going to work that well. You have to make sure there's a standard, and it's also a standard that it's really just unacceptable to deny to a patient population for the purpose of testing a new treatment. It's also really appropriate if you're doing just a minor change or a minor improvement to an existing therapy to compare it against what would happen under the old, traditional approach to that therapy.

I guess, finally, an ethical consideration that also bears some weight is the usual ethical consideration associated with trials and that is, going into it, you have [inaudible]. You don't know that whatever it is you're proposing as this alternative is no worse than or no better than or no different or different. You really don't know, and that's why you're doing the trial. If you can come in and you already have strong evidence that says it really should be no different, then you run into the ethical question of whether or not you should be randomizing people to something when you don't necessarily see the potential for a benefit.

Maj. Pittman: Great, great. So much information. Also, could you explain reversal of errors and why that is important?

Dr. Smolenski : Sure. Now the big concept here for the reversal of errors, your big concern in the equality study design with you type I error is you're going to say there is a difference when the reality is that there really isn't much of a difference, so you end up saying generic is better than the standard. It could happen, sure, but you're trying to protect against that, and that's why you set the alpha at 0.05, so you're allowing for a small amount of that error to occur because you can't completely erase it, but you're hopeful that your study results are consistent with an alternative hypothesis that says no there really is a difference.

In the non-inferiority design, it's reversed, and so the concern is that you're going to walk away saying a design is not inferior, it doesn't cross the red line, when in fact, it does. In the traditional approach, it's rare, but it can easily happen still that you will say there is a difference when there really isn't one, but there's going to be a broader literature base where other people are going to be doing trials with that, and if you have a finding that says there is a difference, when a bunch of other people have a finding that says that there isn't a difference, the bodies of literature is going to come back and say that this may not really be much of a difference at all.

Virtually, with no inferiority, we've seen that studies that have measurement problems, studies that have design issues, are going to be in all likelihood biased toward the null, and so you could end up seeing your study come away with a finding in favor of non-inferiority that isn't deserved, that type I error now, you are rejecting when you should be failing to reject the null hypothesis, but since other studies are going to do it, and if they suffer from errors too, you could end up with a more consistent evidence base in favor of non-inferiority that may not be deserved. That's the real problem.

Maj. Pittman: Okay, all right. This next one's a little long, so bear with me while I read it, all right? How does one deal with a situation where one discovers that the statistical data in a study has been falsified to justify the care of the client who entrusted the care providers with their life?

Dr. Smolenski : Okay, to make sure that I've understood. It involves a situation of falsified data?

Maj. Pittman: Yes, that's what the question says.

Dr. Smolenski : Oh, dear. This is something where it would ... Fortunately, I have not personally encountered this situation in my work, and it's something I would never want to encounter. It's truly heinous. I would say that first stop is your local IRB, whoever had ethical oversight for the conduct of that study needs to be informed immediately.

I was say a next step is, if anything had been published with the findings of that study, every single manuscript would have to be retracted. All of the journals that had published would have to be made aware of the situation, and you would have to work with them to identify an appropriate solution which probably would be retraction of the paper, but it doesn't matter what the design of your study was, falsification of data like that is going to scuttle the entire thing, and it should be reported and dealt with by the appropriate authorities which in most cases are going to be your human subjects protection groups.

Maj. Pittman: No, I totally agree. That's a situation none of us want to get caught up into that one.

The next question is, how many supported studies would be necessary for therapy to be considered non-inferior?

Dr. Smolenski : When you get to the point where enough is enough?

Maj. Pittman: Yes.

Dr. Smolenski : If you follow Popper's philosophy of science, enough is never enough because you can always have the studies that come out and say, actually everybody else is wrong. Practically speaking, I believe that the non-inferiority trial evidence base would be evaluated in a similar vein as the standard approach to clinical practice guidelines, so if you have two independent shops that have each done a non-inferiority trial that meet rigorous design specifications that is using a really, really similar approach to the alternative treatment versus the standard of care, and they both find similar conclusions, that become the minimum threshold to say that you a strong evidence base in favor of non-inferiority.

Maj. Pittman: [inaudible] Some people want to know if you could recommend any texts, book, articles that would be good for master level clinicians looking to get a better understand of research.

Dr. Smolenski : Again, I would way the Greene article in my reference list is a fantastic place to start. I would say ... Who have I found to be a wonderful ... There is a book. You know what? This question is better answered offline, so I will send out the name of a book that I think would be perfect for master's

level clinicians to get just a better understanding or perhaps just a re-duct in the ethical and practical consideration in research and also how to interpret it. I will send that title once I get back to my office, up to the moderator, and we can disseminate to the audience.

Maj. Pittman: Sounds good, sounds good. Okay, one more ... let's see. Again, this is long. If I wanted to conduct a study to test an existing therapy with a new one to see if I could the new one or determine if the new one is as good or better than the existing therapy, how would I begin?

Dr. Smolenski : You would begin in a lot of ways that you would begin in any trial study. You look to the literature to see if other people are doing a similar type of test. You look to see what are the kind of recommendations and what is the current evidence base for the way you're actually doing your therapy now and have a good understanding of that, have a good understanding of how much benefit the literature says your therapy should be administering to your patient population, and then you can start to think of, what is it you really think that your therapy is going to do to the quality of the therapy in terms of affecting an outcome, so what is that effect that's going to happen on efficacy? Think that one through.

If you think it really should be zero, then you can consider it a zero, and that would line you up for an non-inferiority equivalence study. Then you really have to start thinking, what is that difference that you're allowed that maybe a little bit worse than the standard of care but not worse enough to say that you would scrap it and not use it at all? Really spend your time thinking through that.

I would say call a friend who's interested in statistics would be able to work with you in establishing what you guys think is a good margin, consider the clinical implications, consider the fact that you won't have a null group so you have to believe that that difference isn't enough of a difference to effectively say it really isn't better than anything at all, so once you can establish that margin, the rest follows the typical study design procedures. You calculate sample size. You recruit your sample. You execute the randomization and move forward.

Maj. Pittman: We have a lot of clinicians on the lines who are doing the everyday clinical work. How would you suggest they go about getting started with doing some of these kind of studies?

Dr. Smolenski : If you're interested in doing these kinds of studies, and I'm excited that I seem to have gotten people interested in doing these studies, the idea would be find out who's doing research right now, either within your clinical practice or if you have affiliations with the university environment where you have access to people who are already in the research game and that would be willing to serve as collaborators, I think that would be the best place to start.

Maj. Pittman: Kind of along the same lines, what kind of time would they need away from their clinical kind of work to be doing this study?

Dr. Smolenski : It's going to depend on your level of involvement. If you wanted to be the principal investigator on a trial like this, that's going to consume a fair amount of time. It will depend on how many people you are bringing in onto the study as collaborators or if you're bringing in post-doc people as providers so that you're not necessarily spending your time on the study doing a lot of the clinical heavy lifting, you're taking more of an administrative role. That can ease up your time for you to still be doing your other clinical duties, but one of the benefits of getting with an established research group or working with people who are in the research game, if that's there primary focus, they're able to dedicate more time to that work. You can be involved to the extent that you want as a clinical subject matter expert but still maintain your patient work and keep up with what it is you're expecting to be doing in that role.

Maj. Pittman: Okay, just trying to find that balance, what your role should be, and how much time you want to invest. Okay. One last question, kind of along those lines, how standard is the approach described in research conducted by the DOD?

Dr. Smolenski : I would say, for the studies that I have seen that use a Department of Defense or Veterans' Administration population, it looks like they are falling into line with what the literature is recommending in terms of methods. That is one benefit right now that I'm seeing from people reading the literature and within DOD or who are in the academic community but work with DOD. They seem to be following the recommendations pretty well, so I would expect that to be a pretty standard requirement for anything coming out of the DOD or VA population that uses non-inferiority at this stage.

Maj. Pittman: Okay, sounds great. Okay, so I think this maybe the last question. Do you recommend avoiding a mixture of exposures? What if the standard and novel treatments have a mixture?

Dr. Smolenski : Absolutely avoid mixture of exposure. What you need to do is identify what it is you want to test. Do you want to say that there is a difference because of X. Identify what is X, and then do the best that you can to make everything that is not X equal between your groups, so that way any difference you find in your outcomes can be attributable to that, and since X is what we're testing, either equality, non-inferiority or equivalence, the cleaner you can make that distinction, the better off you're going to be.

Maj. Pittman: Well, thank you again so much. You given a lot of us regular clinicians a lot to think about, things we don't think about on a daily basis, so it's been great presentation. For those of you out there, after the webinar, please visit dcoe.cds.pesgce.com to complete the online CE evaluation and

download or print your CE certificate attendance. The online CE evaluation will be open through Thursday, August 11, 2016.

Thank you again, Dr. Smolenski. Will be archived in the monthly webinar section of the DCoE website. To help us improve future webinars, we encourage you to complete the feedback tool that will open in a separate browser on your computer.

To access the presentation and [inaudible] webinar, visit the DCoE website at dcoe.mil/webinar. A downloadable audio podcast and edited transcript of the closed caption text will be posted to that link.

The chat function will remain open for an additional ten minutes after the conclusion the webinar to permit attendees to continue to network with each other.

Next, we have our future webinars, the next DCoE TBI webinar of complementary integrative medicine TBI is scheduled for August 11, 2016 from 1-2:30 p.m. The next DCoE psychological health webinar, Combating Compassion Fatigue, is scheduled for August 25, 2016 from 1-2:30 p.m. Eastern. Finally, the 2016 DcoE Summit State of the Science: Advances, Current Diagnostics and Treatments of Psychological Health and Traumatic Brian Injury in Military Health Care is scheduled for September 13-15, 2016.

Thank you again for attending, and have a great day.

Operator: Thank you. That concludes today's conference. You may disconnect at this time. Thank you.