**WorkforceGPS**

**Transcript of Webinar**

**What Evaluation Designs Are Right For My State?**

**Wednesday, May 22, 2019**

*Transcript by*

*Noble Transcription Services*

*Murrieta, CA*

GRACE MCCALL: And welcome to "Which Evaluation Designs Are Right for Your State?" So without further ado, I'm going to turn things over to our moderator today, Gloria Salas-Kos, senior program analyst and RESEA evaluation TA coordinator, Office of Policy Development and Research, ETA, U.S. DOL. Gloria?

GLORIA SALAS-KOS: Good afternoon, everyone. And thank you so much for joining us for this session today. Together with Megan Lizik, Larry Burns, as well as the evaluation team at Abt Associates, we are working together to develop and coordinate technical assistance, tools, and resources for the Re-Employment Service and Eligibility Assessment Evaluation – or as we call it, RESEA eval TA.

We have a lot of information to cover in this hour, so without further delay let's get started.

This webinar is a second session about evaluating RESEA programs. Some of you may have joined us for our last session on May the 2nd called "Evaluating RESEA: How Does It Help My State and Where Do We Start?" This earlier event provided a broad overview of the program evaluation and highlighted its benefits for RESEA programs. We also talked about tools that will help you form learning goals and begin thinking about potential evaluation efforts. And we shared key evaluation concepts that related to research questions and evaluation design.

If you did not attend the session on May the 2nd, it is recorded and you can find it on WorkforceGPS under "past events."

There isn't any expectation that you will come out of this webinar ready to decide on a particular evaluation design to pursue. Rather, the intent is to help you to understand what the different evaluation tests do and don't do; to determine whether a particular impact design is feasible; and whether it can provide strong, credible results; and to learn about the basics of some of the impact evaluation approaches that we will show in this session.

Future RESEA eval TA webinars and written products will give you more in-depth guidance into those different approaches.

None of the material shared today is highly technical, but there are some concepts and terms that expand upon impact evaluations that aim to help people with less exposure to impact evaluations. To absorb concepts that may take a bit longer to understand, we will walk through that material carefully because it is an important foundation for anyone involved in designing or contracting impact evaluations.

As we move along through this session, we very much encourage you to post questions in the chat box. At the end of the session, our presenters and other RESEA and WIOA evaluation experts at the department will be available to answer as many questions as possible.

To set context for the objectives of our session today, this slide presents a quick recap from our first webinar about the purpose of evaluation design. As you follow along the pyramid, you will see that evaluation design clarifies the program goals and strategies, identifies relevant and useful research questions, determines the approach type of an evaluation, identifies relevant data sources, and describes data analysis plans that allow for valid conclusions to be made.

Moving forward from evaluation design to types of evaluations, there are three broad or general categories of evaluation, and each category serves a different purpose. The types of exploratory – the first types are exploratory evaluations. And we also think of these as pre-evaluation assessments, which we will not be discussing in detail today, but in essence can help you figure out what services to provide or determine what outcomes to measure and how to measure those outcomes.

During this session we will primarily talk about the last two types of studies often considered for impact evaluations, as more in-depth than the pre-evaluation assessment descriptive evaluations do. Descriptive studies, such as implementation and process studies, of (thorough ?) programs, services, or interventions, primarily examine process loads and may be used to establish logic models, some of the things of which we shared during the first webinar.

Impact evaluations like random assignment and quasi-experimental designs focus on differences between treatment and control or comparison groups. The scopes for experimental impact studies are a bit more involved when it comes to gathering data and conducting analysis.

The approach we're taking for this session follows the list of research questions posed on the right (sic) side of this slide. Keep in mind that as you or your program managers begin to initiate an evaluation, your research questions and data sources may – (inaudible) – from your local American Job Centers, local workforce investment boards, state boards, program staff, or other stakeholders who may have a vested interest in your program.

Most commonly, implementation or promise practice studies are used to answer questions like, what really happens in our RESEA program? Or what gaps are there, if any, in the RESEA services we provide?

Random assignment designs are conducted to answer questions like, does our RESEA program improve outcomes? Or would another alternative strategy yield better outcomes?

While not described on this slide, quasi-experimental studies can answer questions like, can we compare outcomes between claimants with similar characteristics, and who would be the same exact services in two different RESEA programs? Lots to think about.

So our presenters this afternoon will walk you through the different types of evaluation studies that may help you formulate your own usage questions through the methods you select, and will also share information about the evaluation technical assistance that is available from your RESEA state programs to use.

To continue this conversation, joining us from Abt Associates are Andrew Clarkwest, a senior associate and the project director for this program; and Tresa Kappil, a senior analyst and evaluation technical assistance coordinator for the evaluation as well.

And without skipping a beat, I'll turn it over to Tresa Kappil. Tresa?

TRESA KAPPIL: Hi. Thank you, Gloria. Now I'll be talking to you about implementation and process studies, which are description studies. Implementation and process studies can be used to answer the question, what really happens in our RESEA program?

Often the terms "implementation" and "process studies" are used interchangeably, and for this webinar we will treat them as such. However, sometimes process studies are more narrowly defined as the description of the flow within a program. In other words, an implementation study documents how a program is being delivered on the ground, while a process study would process on how claimants flow through your RESEA program and document how claimants engage in services.

But these studies also do more than document; they also assess and explain. What we mean by assessment is that these studies can identify how closely program implementation adheres to plans, including what variations exist across locations. And these studies can help understand what claimants' experiences are and support feedback on areas to improve quality or efficiency.

Implementation and process studies have some further uses as well. Because they document how the interventions and plans implemented, these studies can provide feedback to program managers and accountability to program sponsors and the public. They can highlight best practices and identify program areas that can be strengthened. They can inform future program development or support replication. And lastly, the can help interpret impact study findings.

What do implementation and process studies involve? First, as all evaluations, the RESEA team will need to identify research questions of interest. Our first webinar from May 2nd discusses how to do this. Second, evaluators will have to systematically collect data through interviews with staff, claimants, and stakeholders; focus group with participants; review of program documents; observation of program activities; or administrative or survey data.

While an evaluator does not have to use all of these data collection methods, often strong implementation studies incorporate more than one data collection technique. For implementation studies I've been involved in for Abt, we often collect data through at least a few methods. For example, we will systematically collect data through site visits that involve interviews with staff and other stakeholders, following written interview guides. Additionally, we will collect program documents and administer a (granting ?) survey.

After collecting data, evaluators would have to analyze it for themes using qualitative research methods. And finally, evaluators will need to summarize and communicate these findings, often in a report.

It's important to note that implementation and process studies have limitations. Researchers may have to make judgements on certain types of qualitative or subjective data. For example, assessments about changes in the culture of a program may be based on the researcher's observations and the opinions of interview respondents, and may not have statistical data to back up the assessment.

And perhaps the biggest limitation for these studies is that it can't tell you whether your intervention causes changes in claimant outcomes. In other words, implementation studies do not assess the impact of the program. However, as mentioned, implementation and process studies can nicely complement impact evaluations. They may help identify what to test and to help you understand exactly what you're testing. They can also help you understand potential reasons behind impact findings, which may raise further ideas for program changes and things to test.

Now my colleague Andrew will be discussing impact studies.

ANDREW CLARKWEST: Thanks, Tresa. As Tresa mentioned, implementation or process studies are useful in a lot of different ways.

One important thing that they can't do, which she mentioned, is demonstrate effectiveness of an intervention. Of course, demonstrated effectiveness is what the Bipartisan Budget Act of 2018 requires of interventions used by RESEA programs. So we want to spend a fair bit of time talking about impact evaluations here.

There are a number of different kinds of impact evaluation designs. In this presentation we're only going to introduce two broad types: random assignment designs, otherwise called experimental designs, so that's one; and the other are quasi-experimental designs, which encompasses basically everything else. They're called quasi-experimental because they try to replicate what a random assignment design does, but without using random assignment.

So here we're going to introduce the very basics of how they work, and in later evaluation TA that we're going to provide we'll get into those more deeply.

Now, before we start – well, we're going to address a few questions about impact design. I want to start by walking through specifically what we mean by impact. After that, we'll walk through considerations that you should take into account when you're thinking about how to choose an evaluation design, or thinking about a design that someone – maybe someone in your agency or an external evaluator – proposes to you, so that you can know, is this going to work for us or is this not going to work for us?

So first to talk about what we mean by impact. People talk about making a difference. And impact fundamentally answers the question, does our program, or a part of it, make a difference? There are two basic components involved in determining a program's impact. The first is what actually happened – or the outcomes with the intervention, in this case. The second is what the outcomes would have been without the intervention, but if everything else was the same. The second is what we call the counterfactual.

As Gloria said, we're not going to use a lot of really technical terms in our presentation. But "counterfactual" is a term that we're going to come back to repeatedly and it's an important concept to understand so that you can get really what impact designs are trying to do and assess whether an impact design is a strong design or not.

So the impact of RESEA specifically is defined as the difference between outcomes with the RESEA intervention versus outcomes without it, holding everything else equal. That is, what if the outcomes – what's the difference between what the outcomes would have been if everything else was the same in the person's life other than being – other than being assigned to RESEA?

And of course, this is discussing just the impact of RESEA as a whole. There are other impact questions you might be interested in, which we're going to talk about. Again, the first – the biggest one, does my RESEA program make a difference? Are the outcomes for the claimants better than they would have been if they hadn't been assigned to the program? How large of a difference?

Another thing for selection that you may be interested in – because sites – because states can't select – typically can't select every eligible UI claimant for RESEA, that only a subset can be selected. And so you want to know, OK, what are the claimants who would benefit most? And that's another question that can be answered.

A second question is, does a component of my RESEA program make a difference? This is important to understand in order to think about where should we focus resources? Or should we include this additional component or not? And so you want to know what is the effect of that particular piece? So that may be something like, do more intensive services matter? Or how much does our having a third RESEA meeting matter? Or something like that.

As a real-world example of a test that's been done like that, people may be familiar with the WIA gold standard evaluation. For that, you know of course, WIOA funds different tiers of services under its predecessor WIA. Those are called core services, intensive services, and training. And the Workforce Investment Act evaluation tested how much impact intensive services had over and above core services. And then how much impact training has over and above intensive services alone.

So that's the type of question that states may be interested in, in thinking about what sorts of services to provide or to focus on through RESEA.

The third kind of impact question compares differences in the impact of different ways of doing things. This gives you particularly direct kinds of evidence that can be used for program improvement.

As one real-world example, DOL a few years back sponsored a study that tested what would happen if WIA counselors gave customers stronger guidance in choosing training programs that are funded by ITAs, versus just letting customers choose basically any program that was on a workforce board's approved list? Because obviously if you're providing more guidance, this is going to require more resources. So you want to know, are we getting more bang for our buck in doing that? But it may be worth doing if it helps customers get better outcomes from training.

Those are three types of questions. And I want to focus on one particular issue right now, which is to clarify how impact is different from outcomes.

States are very used to measuring performance outcomes. And it isn't necessarily totally natural for people who are used to thinking about performance outcomes to distinguish those from impact. Now, performance outcomes are often used as a metric for program success, but they're a very different sort of thing. We're going to provide a hypothetical illustration here to show that difference.

So your performance measure is what claimants' actual outcomes are. So for example, in this chart we show the re-employment rate for claimants selected for RESEA. And in this example it's 70 percent, so that would be the performance outcome.

But as I noted earlier, impact is concerned with identifying how claimants' outcomes would have been different without the intervention. For that, you need to know what their outcomes would have been if they hadn't had access to it; in this case, if they hadn't been selected for RESEA. That's the counterfactual that I discussed earlier.

So suppose, hypothetically, that in the absence of the program RESEA-eligible claimants would have had an employment rate of 50 percent; that is, if they hadn't been selected. In that case, the impact is the difference between their actual outcomes and what their outcomes would have been if they had not been selected for RESEA. Which in this hypothetical example is the difference between 70 percent and 50 percent, or 20 percentage points.

So this leads to – having talked about what we mean by impact, this leads to the first question presented earlier is what to consider in choosing an impact design. And the very first issue to consider is, can this design give me valid results? Can it actually do what we have said before that an impact study has to do?

Specifically, in the real world we only see what happens under one condition. And our impact design has to answer the question, what would have happened otherwise? What would have happened under the counterfactual? And different impact designs have different ways to try to do that.

So we want our study to estimate impact. And the question is, do the results of our impact study – when we conduct it – reflect only the effects of the intervention and not other factors? For that, we need a strong counterfactual, which we've talked about. And the question is, how do we get that?

Now, as a – we're going to provide one hypothetical example here of a way that somebody might suggest going about estimating the impact of RESEA. That is, comparing the outcomes of RESEA claimants to the outcomes of claimants who were not selected for RESEA. The question is, is that going to work? Is that going to give us a good estimate of what the impact of the program is?

So we're going to return to this chart that we had before. We can see the same two bars to start with – the actual observed rate for RESEA claimants; then the counterfactual, what their rates would have been if they hadn't been selected. Now, of course, we don't observe this. What we need is a comparison group that has those same outcomes and who did not receive the program.

So the question is – so what we want, that's where this third bar is going to be. And what we want is these two bars to be the same. In this case, the third bar is outcomes for non-RESEA claimants.

The question is, so what are the outcomes of non-RESEA claimants? We know that they are likely not the same as what the outcomes would have been for RESEA claimants if they had not been selected. There's a pretty simple reason for this in most cases; the claimants who were selected for RESEA were selected, in most cases, because they have a higher risk of benefit exhaustion. That means we expect that their employment rates are going to be lower than the rates of claimants who were selected; at least if you don't do anything else, if you didn't provide any services.

So in this hypothetical example, the outcomes for non-RESEA claimants are probably going to be higher than what the outcomes – than the employment rates – sorry, what the employment rates would have been for RESEA claimants if they weren't selected. So these two bars are not the same.

So what does this mean if you were to use non-RESEA participants as the comparison group? We saw what the true impact was of the program before, which was 20 percentage points. But now if we've chosen a comparison group who has outcomes that are higher than the outcomes would have been for RESEA claimants if they weren't selected, that's going to lead us to have an estimated impact when we do our impact analysis that's lower than the true impact.

And why is that? Because they're not a good comparison. Because this comparison group – because claimants who weren't selected for RESEA are different from the claimants who were. And so their outcomes are going to be – when we compare the outcomes of RESEA claimants to non-RESEA claimants, their outcomes differ for reasons other than the program.

They differ both because of program participation, but they also differ because these other claimants who weren't selected maybe had stronger job histories in the past, they had different levels of educational attainment and all of those sorts of things impact their outcomes as well. And that means the comparison has a lot of other things that go into it. That is, not all else is equal; the groups differ in ways other than access to the program.

So again, this is obviously a complicated challenge to work out for impact evaluations. But there are standards that have been established to be able to tell, is this impact study a good impact study? And for DOL, through the Clearinghouse for Labor Evaluation and Research, has established standards for whether a study has a good counterfactual. And we're not going to talk about those in depth here, but we will in a later webinar.

But you will want your studies – if you're performing impact evaluations, you'll want those evaluations to meet CLEAR standards. CLEAR does have standards for both random assignment designs and for a few different types of quasi-experimental designs.

We're going to talk first about random assignment designs. We want to show you just basically how random assignment designs work and why it is that they produce a good counterfactual.

So the way that they work is for each person, you flip a coin. If you get heads, for example, they're assigned to get the intervention. So in the basic example we've had before, say that they are selected for RESEA. If you get tails, you're not selected for RESEA. In this case, the only systematic difference between the groups is that one was offered the intervention and the other was not offered the intervention. That means the random assignment holds everything else equal.

And the implication for that is when you compare differences and outcomes for the two groups, we can be confident that the differences between them are because of the program, that they reflect the impact of the program.

Of course, in many contexts random assignment isn't possible. In the context of RESEA, random assignments should generally be easier to implement than it is in a lot of other policy areas. But there may be cases where it's not – where you're not able to use random assignment.

So there are other types of impact evaluation approaches that have been developed to be used in circumstances where random assignment isn't feasible. Those are called quasi-experimental approaches, as I mentioned before. Each of those try to come up with some other way to produce a valid counterfactual; that is, each of those try to come up with a comparison group that's as similar as possible to individuals who receive the treatment.

And we're going to walk briefly through an example of how a very basic sort of quasi-experimental design works.

This chart here gives you a – presents a set of individuals. Each blue dot is an individual who was selected for the intervention. The scatter plot is meant to demonstrate basically that they vary on different characteristics. Some are higher on one characteristic, some are higher on another characteristics the ones on the bottom left are low on both of those characteristics. But this is a very high-level example.

Now suppose – so for each of these people who receive the treatment, what we need is people who are like them who didn't receive the treatment to use as comparisons. Suppose all these red dots are individuals who didn't receive the treatment.

Now, you use matching in a situation where those who received the intervention may, on average, be different from those who didn't. But there are at least some non-participants who do look similar to intervention participants on pre-existing characteristics that we can measure. So for example, in this simplified representation, the dots that are closer together are more similar to each other.

Now, you'll notice that there are a lot of red dots off to the right. Those are claimants who didn't receive the intervention who are very different from intervention claimants – intervention participants, sorry.

In the bottom right (sic), there are also some intervention participants who do not have any non-participants who are like them. So what a matching approach does is to look at only those participants for whom there's a similarity between the groups. So there's blue dots with red dots that are like them, and there are red dots who have blue dots who are similar; and they exclude the others, like the non-intervention claimants on the right and those intervention claimants on the bottom left.

And then at the analysis phase, you make further statistical adjustments to try to ensure similarities between the groups.

Now, there are important limitations to note when talking about matching and other quasi-experimental approaches. One major problem in the context of RESEA is, because claimants are selected by profiling models, there are also no non-RESEA claimants who are like them, who are good comparisons.

There are other quasi-experimental approaches that are designed for strong impact analysis when participant selection into intervention is done, using something like a profiling score. We're not going to talk about it here, but we will, again, cover it in other written materials.

So that's a basic on how you assess the credibility of an impact design. Now, credibility isn't the only consideration to take into account when you're thinking about what sort of impact design you might want to use. A second one is sample size; that is, how many claimants do you need to detect differences in the outcomes that you're trying to look at?

Now, this can be – this is a complicated question. It's intuitive, I think, to most people that a very, very small sample is not enough. Say, suppose you had only half a dozen claimants, that that probably isn't enough to demonstrate the effectiveness of your program.

As a maybe kind of silly example, suppose that you did have three comparison group members and three treatment group members. And say that two out of the three comparison group members get a job and suppose that one out of the three – only one out of the three treatment group members get a job. In that case, there's a big difference in their employment rates; that's a 33 percentage point difference. But the program probably did not actually have an impact of 33 percentage points.

So you know that you see that you say, OK, we see that that's what the difference is, but that's obviously not enough people to tell us. And this is, again, a silly example, but this is a problem that is very common in research, that people conduct impact evaluations but they don't have a large enough sample to draw firm conclusions about whether their program actually worked or not. It might work or it might not work, but it's not – you can't say for sure.

And it's really disappointing to come to the end of a high-quality study only to find that you're unable to tell whether or not there were impacts because you didn't have a large enough sample. So it's an important thing to think about beforehand, how large of a sample do I need for the impact analysis that I'm trying to conduct?

And the sample size that you need varies depending on several factors. One of them is which outcomes you're trying to look at. Another one is how powerful the RESEA – the intervention is that you're looking at. Like if you're examining the effect of RESEA as a whole, that's probably going to have a bigger impact than an individual component of RESEA. So it probably – if you're looking at an individual component, you probably need a larger sample.

And the study design also matters. And these study designs can be – or the required sample sizes can be larger than you might expect. So the sample sizes vary in particular by the outcome that you're looking at.

If you're looking at an intermediate outcome, like did a claimant attend a meeting, you may only need a few hundred claimants in your study. If you're looking at UI duration, you may need thousands of claimants. If you're looking at employment outcomes, like did they get a job or what were their earnings, then you're likely to need tens of thousands of claimants. And that's if you're doing a randomized control trial.

If you're using a quasi-experimental design, those typically require larger sample sizes than a random assignment design does.

The third consideration to take into account is logistics. What actually is involved in executing the evaluation design? And this again can vary across designs. And there are a few different things to think about what might be required. One of them are what are the changes you might have to make to your program implementation? For example, if you're testing alternative program approaches, you may need to train staff on a new approach because presumably you're trying out something new that you haven't done before. So that will take some effort.

For randomized control trials, you'll need to change the claimant selection algorithm. And some RCTs, depending on what you're trying to do, will involve oversight to ensure that staff are providing services appropriately.

One common concern with random assignment designs that's less likely to be a concern with RESEA is having to deny services to claimants. RESEA programs already select claimants, some claimants and not others. And random assignment would typically be instituted by altering the selection mechanism, by choosing at least some claimants at random, rather than selecting them solely by a profiling model or some other criteria that you're using currently.

So in that case, no one would be denied services. Like, if they walk into an AJC, no one's going to say, oh, you're in the comparison group; you can't get services. Rather, what would typically happen is that random assignment would determine who was offered or required to participate in services, rather than strictly denying them.

Another logistical consideration is how much time is this going to take? And this is where QEDs typically will have a big advantage over RCTs is that they can use data that you already have. And in that case, you just have to perform the analysis. Of course, not every question that you want to answer is answerable with data that you already have. But if it is, then that's faster.

Some QEDs and all random assignment designs will require you to collect new data, and this is going to take new time. And of course, any intervention that you want to try out that's new, that isn't something you've done before, is something that you have to do – is going to take more time to do, irrespective of whether you use a quasi-experimental design or random assignment design.

The third logistical issue is what sort of technical expertise do you need? For pretty much any impact design, most states will probably need to procure some outside evaluation assistance to supplement the internal expertise that you already have. And you'll want to make sure that the external evaluator that you contract with has demonstrated experience designing and carrying out the particular type of study that you're interested in.

And then the required expertise varies. QEDs are the most technically complicated because a lot more statistical work is involved in creating valid comparison groups and demonstrating the validity of those comparison groups to others, including being able to demonstrate to CLEAR – to again, the Clearinghouse for Labor Evaluation and Research – that you have a valid comparison group.

The final logistical issue to think about is what IT and data resources you're going to need. So for that you'll have to think about all of the – what are the outcomes you need to measure? What are the maybe participation characteristics or the service receipt measures that you need and what data required?

QEDs will typically require a richer set of background characteristics than random assignment designs were, again, because you need to do a lot more to establish that your treatment group and your comparison group really are similar, whereas a random assignment design takes care of that more automatically.

But you will, for random assignment design, need to incorporate, again, random assignment into the selection algorithm. That will require access to some IT and data resources to do that.

So now, again, this may have sounded like a lot of considerations, a lot of complex things. Going to then leave that with Tresa to talk a little bit more about, OK, so what do we do with all of this?

MS. KAPPIL: Thanks, Andrew. So now if you're interested in conducting an evaluation of your RESEA program, you may now be wondering, how are we going to do this? Well, DOL and Abt are here to help.

Stay tuned for more evaluation technical assistance. This includes webinars on a range of evaluation topics. We're also developing an RESEA-specific evaluation toolkit, as well as a series of brief evaluation-related topics. We'll be provided one-on-one customized evaluation technical assistance, starting with an impact analysis worksheet which walks you systematically through needed planning steps and provides feedback on ideas.

And you can also contact us at RESEA@abtassoc.com for more information.

Upcoming webinar topics include: what evaluation details do I need to plan for and how long will it take? That'll likely be the week of June 17th-21st. Another webinar on procuring and selecting an independent evaluator, which will be the week of July 15th through the 19th. And then we'll also have a webinar on using CLEAR, a demonstration – which Andrew mentioned earlier – and that'll likely be in early August.

Now to Gloria.

MS. SALAS-KOS: Thank you, Tresa and Andrew. We've covered quite a few details and approached several approaches that can be used by states to design and implement their own RESEA evaluations.

Please note that the specific dates and times for the upcoming webinars like this one will be shared via an email announcement from WorkforceGPS and through your RESEA coordinator Larry Burns at a later date and time.

So as we move forward, we will move on to answer some very thoughtful questions that were posed in the chat. We also want to let you know that we have three colleagues joining us today who can answer any policy-related questions. And they are Larry Burns, who's the re-employment coordinator with the Office of Unemployment Insurance; Megan Lizik, the senior evaluation specialist and project officer for the RESEA Chief Evaluation Office; and Wayne Gordon, who is with the Office of Policy Development, Division of Research and Evaluation.

So we'll start with the first two questions that we've received and then we'll open it up for any other questions. So please feel free to enter any questions that you may have about the material covered today in the chat box on the bottom left-hand corner.

So to start, the first question that came in is, "How do we address human subject review as we're measuring impacts on persons; especially if we're choosing to not provide a full menu of services to some participants to create a control group?" Andrew, can you respond to that question?

MR. CLARKWEST: Yeah, absolutely. And I really appreciate this question. This is a very important thing to think about in any kind of study, any type of evaluation you might be performing, both for random assignment design or any evaluation where you're collecting new data; or any situation where you might be sharing data with, say, an outside contractor or researcher who wouldn't normally have access to the data.

But it's really important to think about human subjects and get a review of the proposed approaches to make sure that they meet human subject standards.

If you're providing – for the specific question about random assignment, if you're not providing a full menu of services, that for most states is consistent with what is being done already. The states are already choosing to select some participants for RESEA and not, and so there are some who are being offered a particular menu – or required a particular menu of things – and others who aren't. And random assignment only changes who those people are. But nonetheless, it's very important to have an institutional review board review your designs.

And I realize that a lot of agencies won't necessarily have those internal resources. But any external evaluator who you contract with should have access to some kind of IRB – again, institutional review board – who can perform the human subjects review. And if you're not sure who to turn to for human subjects review, that's something that we can discuss with evaluation TA as well.

MS. SALAS-KOS: Thank you, Andrew. So we have another really good question. And this one we can pose to both Wayne and Andrew. And this is, "If we were to use randomized control trials to evaluate, how can we achieve randomness once the RESEA-eligible group is already selected based on their likelihood of exhausting their benefits?"

WAYNE GORDON: This is Wayne Gordon with ETA's Division of Research and Evaluation. This question makes me ask additional questions. And one just I wanted to point out that one of the most important things about randomization is the point at which you randomize. Because you want to make sure that you're selecting the two groups you're comparing from the same universe.

So it sounds like this would be after they've been selected for RESEA. So that's a good point. Stop right there. That's the group you want to come from – that you want to select from.

However, you must also ask yourself a question – and this touches upon what Andrew was speaking to before about sample sizes. You may be – depending on the client flow of participants in RESEA, you may be able to serve everybody who comes through the door or you may not be able to serve everyone that comes through.

So some other questions must be asked of yourselves as to whether there's enough client flow in order to do randomization where you will have a comparison group. Or if you don't, perhaps a different evaluation technique or methodology should be considered.

Andrew, do you want to add anything to that?

MR. CLARKWEST: Yeah. I think there's one point is that random assignment doesn't mean that claimant's still can't be intentionally targeted. It will require some changes. Wayne's point is absolutely right about thinking about the point of random assignment.

Obviously, states now aren't required to continue to use profiling models. But say that you want to. Say that the state still wants to select claimants for RESEA that have a high probability of exhausting benefits. You could do that with a random assignment approach, but you would need to expand – you need to lower the selection threshold. So say, rather than selecting those with profiling scores of .7 and above, you may need to select those with scores of .5 or above. And then within that set you can randomly assign some to RESEA and some to not.

And again, we've only been talking about selection for the program as a whole. If you want to randomly assign for individual components, which is likely to be the case, then the selection may be a little bit different. That is something that would have to be discussed where the appropriate point of random assignment is. And again, that we can potentially help with or an external evaluator could help with.

But there are ways to mix intentional targeting and random assignment and still have a valid control group.

MR. GORDON: Thank you, Andrew.

MEGAN LIZIK: And Wayne and Andrew, this is Megan Lizik with the Chief Evaluation Office. I would like to just emphasize the last point that you made, Andrew, about the experienced independent evaluators. These are the people that are going to be able to figure out the technical details that are the right way to proceed in your specific instance.

MS. SALAS-KOS: Exactly. Good point, Megan. So now we have another question related to IDRBs as well. And this is, "Do we submit any (2E ?) designs to an IRB for approval or have a licensed IRB evaluator on staff?" Andrew, is this one to you?

MR. CLARKWEST: Sure. I may misinterpret it. If so, the person who asked the question can ask the question again.

I think in most cases, particularly most cases where you have an external evaluator, the state will not necessarily be the one submitting the design to the IRB for approval. It would be the evaluator who does that.

Whether you have a licensed IRB evaluator on staff, I'm not sure if that means you have an internal IRB. It's the evaluation rather than the evaluator who would be IRB approved. But I may be missing something.

Megan, jump in if you have any other thoughts.

MS. LIZIK: I would just say that technically third-party evaluators will seek IRB approval, like you said, for those evaluation designs and sort of – because that provides sort of that independent review of what that evaluator's doing and sort of some – it's really in their best interest. So I would have a hard time imagining a case where you – a state would need to have a staff – an IRB staff person playing that role.

MR. CLARKWEST: Yeah.

MS. SALAS-KOS: Thank you, Megan. And just to clarify, quasi-experimental designs do involve those people who are being served. So there would be human subjects involved and you would also include informed consent as part of this process, which we haven't discussed but we'll probably be going into later as we move forward in the additional webinars.

So the next question I'm going to direct at Megan to see if she'd like to answer that. And I'll read this to you out loud. "As evaluation is envisioned, what is the role of individuals conducting the evaluations? For example, what would the role be of the UI administrators, reporting personnel, or research and statistics teams or others?" And then maybe Andrew can chime in as well.

MS. LIZIK: So thanks, Gloria. And thanks for that question. I think the first thing I would just say is something that we talked about a lot last time, in our last evaluation webinar, and that's that DOL is working on some evaluation guidance that might help clarify a little bit of sort of what this might look like in your specific case, in your specific state.

But I think this is going to be something that might vary from state to state, so I don't think that there's – it's hard to sort of answer the question exactly without having a lot of those details. But I do think that when we – it's important to make a note that when we talk about independent evaluators, third-party evaluators, what we're really talking about is contracting with an experienced evaluator who this is – that has a lot of this deep experience and methodological expertise for designing – (inaudible) – to answer specific questions that you have.

And so while internal state staff are likely to be instrumental in sort of outlining parameters for things like, what are your research questions, what do you want to answer, what's your theory of change for your program, how do you think it works, and those kinds of things and helping figure out what types of evaluations might be the best to answer those kinds of questions. And then helping sort of conduct it once you have a third-party evaluator in place. They probably are not going to be fulfilling the function of actually doing the evaluation.

But again, I think I would say – I would suggest to wait until some of that guidance is out and until you try to make some more specific decisions for how it might play out in your state. And like Gloria and others have mentioned, we are here. Larry Burns and the Office of Unemployment Insurance, myself – Megan Lizik – and Gloria, we are here, along with the RESEA evaluation team. If there are specific questions that you have, we can try to get you pointed to the right place sort of when some of that has come out.

Larry, is there anything else that you would like to add?

MR. BURNS: No. I think, Megan, you hit the nail on the head with the response that each state is starting from a different place as far as capacity goes. And each state's structure is different in how they administer RESEA.

So there's really no one answer to this question other than, as you can tell by how we structured this technical assistance, we've left it open to the workforce system in general. So we have UI, workforce, local; we're trying to basically do an all-hands approach to getting people to understand the requirements. And that way, we'll be ready to implement them once they take input.

I think you covered it pretty well, Megan. Thank you.

MS. SALAS-KOS: Here's a really good question. Andrew, can you answer this one for us? It is, "My state has some high population areas where we should we be able to deploy a random selection. But there are also lower population areas where random assignment would result in a low sample size. Any ideas?"

MR. CLARKWEST: So this is definitely an issue if a state is interested in knowing what the impact is specifically within a particular part of the state. Is it bigger in this part of the state versus another part of the state? Then this is definitely a problem because, say, you don't have a lot of people in one part of the state, so it's tough to know what the impact is in that part of the state.

But if what you're interested in is the impact overall, statewide, then this is fine, because then the data for the small areas get mixed with the data from the larger areas and they get analyzed all together and that's what matters in that case. If you're looking – if you're interested in statewide impacts, all that really matters is how many people you have in the state – or statewide, rather than in each individual area in the state.

MS. SALAS-KOS: We've got about five minutes left and we can maybe take one more question. Megan, can you answer this next question? "Can DOL or a contractor provide peer review services for QED impact studies?"

MS. LIZIK: That's a great question. So we haven't talked too much about providing that kind of evaluation TA broadly yet. But we do have some resources to be able to help with individual design challenges like you're talking about.

And so what I would suggest for now would be to reach out to the RESEA evaluation TA team email box with your specific question and we can try to make sure that we get you an answer at this point in time.

MS. SALAS-KOS: (In progress) – answering at a later point in time. I'm sorry. You caught me in mid-sentence here.

So thank you very much, everyone. We're about three minutes out and we did get quite a few questions that we aren't going to be able to get to today.

Just as a reminder, we do have – we recognize and appreciate that you're very interested in continuing this dialogue, and you may do so by reaching out to us via the contact information on this slide. And in particular, if you're interested in a state's RESEA grantee or program manager, please send an email to the resea eval TA inbox that's on this slide.

Any questions related to RESEA program policy, reach out to Larry Burns. And of course Megan Lizik is a project officer for this study and so you may reach out to her as well and/or I and Andrew and Tresa are also available to consider any technical assistance questions that you may have.

We thank you for being part of the session today.

MR. BURNS: This is Larry. Just wanted to add one thing. (Inaudible) – quite a few questions we didn't get to get through all of them. But I will say that we're currently in the process of developing some guidance for the upcoming year.

So we're working on fiscal year 2020 operating guidance for the RESEA program, which will be similar to guidance you've seen in previous years. But we're also working on a separate guidance document that's going to be specific to the evaluation requirements. So we'll be taking the feedback we received during the session back with us as we help draft and finalize that guidance.

I also just wanted to give a heads-up that we are working to finalize the schedule for the events. Right now we have kind of a range of dates. But we are planning, once we have those dates finalized, to send an e-blast out; that way you can plan accordingly for the upcoming events.

I just wanted to mention we have those couple things in the works. So thanks, Gloria, just wanted to –

MS. SALAS-KOS: Thanks for chiming in.

MR. BURNS: Yeah.

MS. SALAS-KOS: And that's true. We do have a couple of other things in the works that you should be looking for as well. We thank you for being part of this session today and look forward to – please look forward to getting new announcements as they come forward.

Thanks again, everyone, and we appreciate your time.

(END)