**Meta-analysis of single-case experimental designs: An overview of current synthesis methods**

Presenter: Dr. Mariola Moeyaert

Webcast Chair: Dr. Oliver Wendt

Sponsored by The Campbell Collaboration and by AIR’s Center on Knowledge Translation, Disability and Rehabilitation Research (KTDRR)

[https://ktdrr.org/training/webcasts/webcast81/index.html](https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Fktdrr.org%2Ftraining%2Fwebcasts%2Fwebcast81%2Findex.html&data=05%7C01%7Ctbauman%40air.org%7C70583e1cde7c4eca585b08db1106826a%7C9ea45dbc7b724abfa77cc770a0a8b962%7C0%7C0%7C638122491870724215%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=U49sG9liw%2B31C%2FDI4no7VflD3fKGQiVI9nEQrOdxG3A%3D&reserved=0)

Transcript for audio/video file on YouTube:

<https://youtu.be/bgQ1kO7JGdI>

OLIVER WENDT: Well, hello. And welcome, wherever you are around the world. We certainly appreciate that you're tuning into our first event to kick off the 2023 Campbell Webinar Series. My name is Oliver Wendt. And together with Joann Starks, I'm co-chair of the Campbell Disability Coordinating Group. We are very grateful to the Campbell Collaboration and the Center on Knowledge Translation for Disability and Rehabilitation Research for sponsoring this webinar.

For this session, we have chosen the exciting topic of synthesis and meta-analysis of single-case experimental designs, which has received ever-growing interest in the fields of disability and rehabilitation research. Single-case designs and their synthesis are essential methodologies to evaluate treatment effects and document evidence-based practices. Over the last decade, we have seen a proliferation of new approaches. But as you know, there is often some uncertainty among applied researchers when selecting suitable procedures.

Today, we are very fortunate to have with us one of the world's foremost experts on the topic, who will present us with an overview on the current state of the art. It is my great pleasure to introduce Dr. Mariola Moeyaert. Mariola is associate professor of statistics in the School of Education at the University at Albany in New York.

Her major research interests relate to multilevel analysis, meta-analysis, large-scale data analysis, and interrupted time series, which has resulted in over 60 international publications to date. Her research specifically on multilevel meta-analysis of single-case experimental designs has been funded by three grants from the Institute of Education Sciences in the US. And I can say from my own experiences, whenever it comes to data analysis questions and single-case research scenarios, Mariola is definitely my go-to person.

Some of the content in the presentation will require more of an intermediate to advanced level of knowledge in regression-based statistics. You may not be able to fully understand everything in one session. But we will point out some resources for further study. And there will be a follow-up webinar in the fall, where will be showing some applications with real data from current research projects.

The results for our webinar are as follows. Mariola will present for about 45 minutes. And we ask that you mute your microphones and hold your questions till the question and answer session at the end. If you'd like, you can type a question into the chat room, and I will collect the questions to be answered after the talk. So, without any further ado, let me turn it over to Mariola.

MARIOLA MOEYAERT: Thanks so much for the introduction, Oliver. And welcome, everyone. I just would like to acknowledge support from the Institute of Education Sciences. So through their support, I was able to contribute to the development and application of meta analytic techniques for the synthesis of single-case experimental data. And I also would like to acknowledge Dr. James Pustejovsky for collaboration on some of the content slides.

So, after a general introduction, my aim is to provide just a general, broad overview of three major approaches that can be helpful to synthesize single-case experimental design research. And then, as Oliver mentioned, at the end, there will be some time for questions.

Now, to fully understand the three broad approaches that I will be talking about, it is important to understand some main features related to single-case design experiments. And instead of just reading the text here, I'm going to actually display an example of a single-case experimental design study. So here, you can see one important feature. That is that we have repeated measures over time, and this during different conditions.

Traditionally, we have a baseline condition that is followed by an intervention condition. And that, again, can be followed then by other conditions, such as maintenance condition. What is very important to understand is that each participant of a single-case study serves as its own control. Because we can contrast for each participant its own baseline data with its own intervention data. So that is a very important feature, repeated measures on control.

And a second one that I want to highlight is replication. Because in one single-case study, we always have actually multiple participants, most of the time. So here, we can see one single-case study with one subject 1, subject 2, subject 3.

So why am I highlighting this? Well, if you're new to single-case design research, it's important to understand that we have repeated measures and that we have replication, either within a participant or across participants. And that is important because when we are thinking about an approach to synthesize multiple single-case studies, we need to account for that.

We will deal with dependencies because we have measurements nested within a subject, and we have multiple subjects that are part of one single-case study. And then we want to pool that data together across multiple studies. So that is why I just take a little bit of time to introduce this design.

Well, if we then calculate an effect size-- and I'm going to talk more about that in a bit-- then we can do that for participant 1, 2, and 3. However, can we truly generalize that the intervention is working using data from only one to three subjects from just one single-case study? Well, no, it has very limited external validity.

So in order to increase that external validity, the researchers have been proposed to synthesize data from multiple studies. And the good news is that that is really possible. Why? Well, here you can see a graphical display of the number of publications. On the left-hand side, you can see the number of publications. And that is the number of single-case experimental design study publications.

You can see that there is an exponential increase in the number of publications. So the good news is, there's more evidence available for research synthesis meta analysts to combine the data. And that is really what we're going to focus on during this presentation. So how can we think about combining data from single-case experimental design studies?

And it's something that is really needed. Because then we can really contribute to decisions that are made in policy and practice and theory and research. So we can really focus on generalizability-- and not only generalizability. By looking into summarizing the magnitude of the intervention effect part, we can also look into-- well, is there a lot of heterogeneity?

So, across all the studies, is there a lot of variability in intervention effectiveness between the studies, but also within cases from studies? And then the next step is, well, if there is a lot of variability, how can we explain it so that the focus is on moderators?

I'm not going to go into all those details. That will be for a future webinar. But I'm going to talk about the approaches that are available to us. And then we can further expand that to deal with heterogeneity and to add moderators. But first of all, it's important to understand which approaches are there for us.

Before we talk about those approaches, I just want to mention that it's really nice to work as a meta analyst in the field of single-case experimental design research. Why? Well, the raw data is always available to you. So you do not overly rely on a statistic that is reported in the paper. Why? Because there is a tradition to graphically display single-case experimental data.

And we do have those graphs. We can just digitize the raw data from those graphs and put that nicely in a data file. So we can calculate literally anything we'd like based upon that raw data.

There are many data retrieval programs available for this purpose. I personally like WebPlotDigitizer a lot. So what you would do is, you just take a screenshot of the graph that is reported in the single-case study. You import it in WebPlotDigitizer.

And then after calibrating your x- and your y-axis, you can just click on every data point. It s automatically is exported in a format that you can use in Excel or Word or any program you'd like. And eventually, what we want is a meta-analytic data set. And that is important to understand because that will be used as input to calculate our effect sizes and then to apply our three approaches that I'm going to talk about.

You would have a column with each study name, then the cases or subjects that are part from that study. Then we need an indicator telling whether your measurement Y is part of your baseline. So phase would be 0 or 1 if your measurement Y is part of your intervention. So Y is really that Y outcome score. So we can see here, it would be 3, 3, 4, 2. Those are all baseline measures, and then we have some intervention measures. And then we have the session number also indicating things.

Once you have your raw data, you have that nice meta analytic data set, we need to think about, how are we going to synthesize that now? And we organized it according to three broad categories.

So you can conduct a meta-analysis using study-level summary effect sizes. And that is actually not that different compared to group comparison design research. You would have an effect size at the study level with its standard error. And we can just use traditional meta-analytic approaches.

But in single-case design research, we have not the study as a unit of analysis; we actually have the cases that are nested within studies as a unit of analysis. So we can run the meta-analysis of case level effect sizes. That's a second approach.

And then the third one is, because we do have that raw data, we can actually go a little bit further. And we can actually model different functional forms and so on at the measurement level. We would call that the raw data meta-analysis. So that would involve a three-level model. Because we do have measures nested within our cases, cases nested within the studies.

The three approaches are quite different, one to another. And so we really need to think about, what is the level of analysis? What are my research questions? What do I want to address?

And then look into, OK, so if I use this approach, what effect size metric could be useful? And then also, what assumptions am I willing to make? So I'll talk about all those things in a bit.

And what could be helpful for you in order to make your decision is to perhaps use this flowchart. First of all, your question should be, what are my study types? So am I exclusively going to synthesize data from single-case designs? Or am I eventually interested in combining data from single-case designs with group design studies?

And that's possible because there are studies published addressing the same research question, the same outcome, and so on, but they just use a different design type. So, can we combine those design types in one meta-analysis? And if so, which effect size should we be using for that? And for that, so that would be called the study level effect size. And there's just one metric available for that, which I'm going to talk about in a bit.

Now, if you just focus on single-case design research, we actually can take the nature of the single-case design types into account. Because then we can either summarize that case level or we can summarize the raw data.

And in order to make a decision between those two approaches, we need to look into our outcome types. Do we have very varied and heterogeneous dependent variables, even within one study? Well, then it's better to summarize case by case your intervention effect and then apply your meta-analytic technique.

However, if you have common and homogeneous dependent variables, then I would strongly encourage to consider the raw data synthesis. Because this has a lot of potential flexibility and so on, but more about that in just a bit.

This is just a general overview, and it could help you understand the meaning of this presentation. So I'm going to talk about those one two three approaches. And after I introduce each of those approaches, I'm also going to provide a very brief illustration using real data.

And this real data comes from a meta-analysis that is currently under review. But I just randomly selected 10 studies out of that meta-analysis for demonstration purposes. The meta-analysis solely has single-case design data. So we can either use any of the three approaches.

Then the population would be K-12 students. The intervention is a graphic organizer. The outcome is, hopefully, that graphical organizer is helping increase the reading comprehension. We really wanted to look into studies that have a clear baseline and clear intervention and so on, so we can easily calculate the effect sizes. There's a total of 23 studies. But again, I'm just going to use 10 studies randomly sampled out of those 23.

I already talked about the data organization, so I can move on to the next slide. Just one thing that I wanted to mention is, in addition to the study, case, session number, the Y outcome, and the dummy variable indicating baseline of intervention, you can add other variables, and those are all called moderators. That would be in the next step to explain heterogeneity and effect sizes between cases between studies-- but more about that in a future webinar.

So using this meta-analytic data set, we can apply any of the three approaches that I just briefly introduced. Let's start with the study-level effect size. So with the study-level effect size, there's actually only one option. And that is the between-case standardized mean difference.

And that is sometimes also called the design comparable effect size. Why? Because this effect size has been developed, so it could be assumed on the same scale as Cohen's d and Hedges' g from group design studies. And that is the reason why we can combine group design studies and single-case design studies in one meta-analysis using this between-case standardized mean difference metric.

Now, what is important here to understand is we just have that one single summary effect size for each individual study. So here, you can see three studies that I randomly sampled out of the 10 studies that I talked about earlier. So we have study 3, 25, and 45. So they just given those numbers.

And so because I'm having here three studies, for each of those three studies, we have g, which is an effect size, and SE, which is a Standard Error. So g here represents that between-case standardized mean difference for study 3. Then we have that for study 24, and so on and so on.

So here, the goal-- and you really need to think about, when you're selecting your approach, here-- the goal really is to summarize study level, so average effect sizes across the cases within part from one study. And the main reason to do that is because then you can apply that between-case standardized mean difference that is comparable to Hedges' g from between group studies. Your intention might not be to combine with group design studies.

But your intention might be, I actually do want to have an effect size and metric that I can make similar interpretations compared to Hedges' g from group design studies. So I kind can work with that scale, what is considered to be small, moderate, and large. So that could be your intention. Or your intention could be, I actually do have a lot of group design studies as well. So I want to combine that in one meta-analysis.

So here, you can see some resources. You can look into them at your own time. Because there are some resources that help you to first calculate that between-case standardized mean difference before you run your meta-analysis. So just know that we're making some assumptions.

Because in order to calculate that between-case standardized mean difference, we're applying a two-level hierarchical linear model-- but more about that in just a little bit. If you don’t think you quite understand this metric, think about the standardized mean difference from group design studies.

So what we've been doing traditionally, we're subtracting the average outcome from the control from the average outcome in your intervention group. And we're dividing that by a source of a standardizing factor. Here, usually it is the standard deviation of your control group. So that is here, the general equation.

And if we translate that, then, in the between-case standardized mean difference using single-case designs, well, the numerator is not any different. We still have intervention and baseline control group data. But now, the standardization factor becomes different.

Because now we have not only the standard deviation, the variability within your groups, but we also have the variability within our participants because we have those repeated measures. Think about the beginning of the session. Here, we can see that the numerator becomes different. We have within-participant variance and between-participant variance.

This might get a little bit more technical, so I'm going to just give you the main idea. But just know that there is actually an entire webinar developed talking about this single metric. So I would guide you towards that resource if you want to understand more about the underlying mechanics. So here, we're using a two-level hierarchical model to calculate between-case standardized mean difference. What are those two levels? The level 1, that is actually modeling variability over time within participants. So that's a regression model.

Wwe have Yij. So that's outcome on measurement occasion i nested within subject or case j, that equals this right-hand part of the equation. And you will see Trt is one independent variable. It's the dummy variable. That's that base indicator, 0 if measurement occasion i from case j is part of the baseline and 1 if it's part of the intervention.

So why am I mentioning that? Well, if Trt equals 0, then we have an estimate of the baseline level for case j. And if it's 1, then B1j, that represent that change in level. And that B1j and B0j, they can be different from different participants.

And that is, we add level 2. So at level 2, we're looking into that variability in our coefficients at level 1. So, for instance, here, beta 0j, which is that baseline level, equals an overall average baseline level across all the cases from that study and some deviation.

Then we need to make a decision also for the treatment effects of beta 1j. We can also allow that to vary between participants. But the most basic model is saying that we have a common effect, a common intervention effect for my cases that are part from the same study.

And then this theta 1, that is a common effect across the cases from one study. And that is what is used here on the right-hand side to calculate our design comparable effect size, so DCES, or the between-case standardized mean difference.

Theta 1 is not standardized. Think about the numerator on the previous slide. We still need to divide that by a standardization factor. And I already mentioned that is a function of the between-participant variance and within participants. So sigma squared is within-participant variance. Tau squared is between-participant variance.

This is nice because we have now that standardized mean difference. Delta is that standardized mean difference. And that is assumed to be on the same scale as Cohen's d, Hedges' g from group design studies. So again, more details-- I'll share a resource with you. But that is really the basics to understand. And then we can do the same as with group compares and design search. We add a bias correction because we're dealing with small sample sizes. So this is displayed here.

But we're not yet there. Because we only have here effect size that is bias corrected. We still need to combine that across our studies. So here, you can see we can then run a traditional random effects meta-analytic model. That's the same as with traditional meta-analysis.

So here, you can see we have our effect size. That's the between-case standardized mean difference. We have the overall average effect size. And then we have those sources of variability that could be very interesting.

So those deviations follow some assumptions. We have normally distributed. And they have a variance of tau squared. And that tau squared-- I'm going to show that in the output-- that is the between-study variance and effectiveness.

So tau squared is that heterogeneity, that eventually, we probably want to add moderators to explain heterogeneity. That is the between-study variance and effectiveness. We're also making some assumptions about those errors as well. But that's not any different than traditional meta-analysis.

There are some limitations that you need to be aware of, some assumptions we're making.

But also, we need to have single-case design studies that have more than three participants. Why? Because we need the between-case variance for the standardization.

It's also available for limited design. So you might need to change your inclusion criteria and limit them to multiple baseline data and some other design types. Because otherwise, you cannot apply that between-case standardized mean difference.

Then the data organization-- so remember, we have that raw meta-analytic data set. First we need to calculate for each of the studies that between-case standardized mean difference with a standard error. And then we can use that as input to run our meta-analysis, which is that random effects model that I just talked about.

Don't worry too much about those two screenshots here. This is just the R output that was obtained by running that random effects meta-analysis using the between-case standardized mean difference.

As we can see here, k equals 10 studies. We are using the restricted maximum likelihood. And what we can deduce from this is the model result. The estimate of the between-case standardized mean difference is very large. It's 3.14.

If we are assuming we can use this scale from Cohen's d, Hedges' g, that's extremely large. And then we also have the standard error reported, the 95% confidence intervals. We see 0 is obviously outside that confidence interval. So you would deduce at the 0.05 level two-tail testing that we do have a very effective, statistically-significant intervention effect.

That between-study variance, that heterogeneity, it's 4.22. We can see also the heterogeneity statistics indicate that there's a lot of variability in our effect size between studies-- no surprise, because look at those studies here. Those are the 10 randomly selected studies. And we can see, some effect sizes are very different, one from another. So the results do make a lot of sense.

I already mentioned that there's a whole lot going on behind the scenes when thinking about that design-comparable effect size. There is a webinar I presented in 2020 about this. It's an hour long. And there, you can learn about more details about the whole process in preparation to calculate that standardized mean difference, some interpretations. And there's also a demonstration using a Shiny application for the calculations. I strongly encourage you to have a look at this if that is something you're interested in.

So that was, in a very brief nutshell, the first approach, which was the meta-analysis of study-level effect sizes. We're ready to transition to a second one. And that is a meta-analysis of case-level effect sizes.

It has a very different goal. Because now, we're not interested in combining with group design research. We're also not interested in having a metric that is somehow comparable to the metric for Cohen's d. Here, we're really focusing on single-case designs solely.

So what we can do now is calculate an effect size, not at a study level, but for each of my cases, which is very appropriate to do. And then we would end up with having an effect size f subject 1 from study 3, subject 2 from study 3, subject 3 from study 3. We would have a lot of effect sizes.

I already talked about the goal here. The available metrics, that's something to talk about a little bit. There's a lot of them. Some might be better than others, and there's a lot of papers published about that. I'm going to just introduce a couple of them.

But the assumptions we're making here, we're really focusing on effect-size metrics that are case specific. Some of those effect sizes have standard errors, but most of them do not have that. That's important to understand.

And most effect sizes here, they assume no time trends and no real data complexities are modeled-- for most of them, at least. Because we will have, for each study, multiple effect sizes because we have multiple cases within the study; we naturally deal with a hierarchical structure.

So in thinking about that, if we're going to apply our meta-analytic approach, we do have dependency between effect sizes from the same study. So we need to think about how we can model that.

We cannot just use a basic meta-analytic model as the one that I introduced before, that random effects model is traditionally used also in group design studies. We just can't use that because we have dependency. So we'll have to think about that.

And luckily, there are some researchers who helped us with that. And they are Dr. James Pustejovsky and Chen. They had two approaches to deal with that. We can just do the simple average and apply the cluster-robust variance estimation.

Or-- and this is what is very nice-- I would recommend that one, the multilevel meta-analysis and apply that cluster-robust variance estimation. What is that cluster-robust variation estimation all about?

Well, it's dealing with that dependency. It's correcting your standard errors for that. So those are those two approaches. But I would guide towards-- we shared it at the end of the presentation in the reference list, the resources. They talk about in detail and compare those two approaches as well.

What is nice about the multilevel meta-analysis model is that we have all those sources of variance. And I talked a little bit about that previously. But now, we have an additional source of variance. Remember, that tau squared. We talked about that. That is that between-study variance.

But we also have omega squared that we get estimated, which is that between-case variance. And then we have the within-participant variance, which we don't have if we applied the simple average. We don't have those additional sources of variability.

Before applying those two approaches, we need to make a choice on which statistic to use. And you will see, depending on which statistic you use, you can either apply the simple average or the multilevel meta-analytic model. Unfortunately, that multilevel meta-analytic model can only apply using very limited number of effect sizes. Most of the time, we'll end up using the simple average with that cluster-robust variance estimation.

But here, you can see those available effect sizes. I'm not going to go into detail. Just know there's a lot of them. I classified here in three groups. But there's also another group I should have called "others" of remaining. But the non-overlap indices are very popular. There are some mean-based ones and then the regression-based effect sizes, which I tend to like a lot because of its flexibility.

Now, the idea of a non-overlap indice is, you're contrasting your baseline points with your intervention points. And you're basically going to see how many data points are overlapping or non-overlapping. And all those metrics are a variation, one of another. They're just easy to interpret.

Because here, for instance, the PND, you basically calculate how many points in the intervention are exceeding the maximum of the baseline points. And then you can see here, would be 7 points out of the 11 or exceeding that maximum, so 60%. It's so easy to interpret, but it also is very limited. That might be an outlier. There's a lot to tell about that. But just know that one is always a variation of the other.

Then the mean-based ones, they're not going to count data points at all. They're going to use the mean of the baseline and the mean of the intervention to calculate an effect size. And here, the response ratio is just the ratio of the mean YB here, YB. That is the mean of your intervention data. And then YA is the mean of your baseline data. So we're going to look into that ratio. It's a change, a proportional change statistic.

So we take the natural logarithm of that because then we have a nice sampling distribution to work with. We actually do have a standard error. For instance, here, the PND and some of the other non-overlap statistics, they don't have those nice standard errors. So it's limited in things we can do at a meta-analytic level. Basically, for those metrics, most of them will have to use the simple average with that robust variance estimation.

Whereas for the log response ratio, we can actually use that multilevel meta-analytic model. That's a very nice feature. And then the standardized mean difference is also an indice that is used quite a bit-- not to be confused with the between case. So within case, we have that metric for each of the cases separately. That between-case standardized mean difference pulls that summary statistic across the study. So that's different.

Here, for each of your cases, we can calculate the difference in means from baseline and intervention. And we divide it by the standard deviation of that baseline data. We apply that Hedges' bias correction. And then we have the standardized mean difference within for each of the cases, and we use that as input for the meta-analysis.

The last group are those regression-based effect sizes. So here, we can see that regression model. Trt, remember, that's that dummy variable, 0 if y is part of the baseline, 1 if it's the treatment. So that beta 1 estimated by B1, that is really looking into a change in level.

But here, we're adding some time components too. So we can see time, that here. So beta 2, B2 is the estimate of beta 2, is the slope in the baseline. And we can also look into an interaction between treatment and time. So B3 would be the estimated changes in baseline and intervention slope.

Most of the time, we would then have B1 and B3 as indicators of intervention effectiveness. Because, is there a change in level when we start the intervention? And is there actually a change in slope? So those are indicators of intervention effectiveness. And we can use that together with its standard errors as inputs of our meta-analysis.

So it’s quite different from the approach I introduced first, which was a study-level effect size. Now we have effect sizes for each of the cases within a study. And then once we make a choice, do we want to model time trends? Am I just interested in non-overlapping indices? Maybe I want to see a ratio of change.

So that's a choice that needs to be made. And once you make a choice, you can choose the approach. Now, what approach to use? I already mentioned that, well, most of the time-- so the Chen and Pustejovsky study indicated-- for some of the non-overlap statistics-- actually, most of them-- and the within-case standardized mean difference, we have to use that simple average.

The log response ratio, though, because of its desirable statistical properties, we can use that multilevel meta-analysis model. And we have those different sources of variability that I talked about. That could help if you're interested in running a meta-analysis and you are only combining single-case research data. You can calculate case-level effect sizes. Then you can apply either the multi-level meta-analysis or the simple average.

And here, the data organization is also quite different compared to the study-level effect size. Because now for study three, we don't have just one effect size. We have an effect size-- so here, for case 1, 2, and 3 from study 3. Then we have an effect size for case 1, 2, 3, 4, 5, 6, from study 7, and so on.

Here, I used the NAP, the number of non-overlapping pairs. And this one is a metric that also has a sampling distribution. It has a standard error. So the estimate and the standard error. I first calculated the estimate and standard error for each of my cases, from each of my studies, together with these confidence intervals. You don't have to do that. You basically only need your NAP estimate and its standard error as input of that multilevel meta-analysis.

On the right hand side, you jcan see that forest plot. I always like to look into that because it gives a visualization of your effect sizes for each of your cases within the study. And it gives you an idea. It might anticipate a lot of variability or not. And I would say yes here.

Within one study you can see the same colors indicate effect sizes from the same study. So you can see the greens, the blues, the red ones, they all belong to one study. And then we can see there's a lot of variability in those effect sizes within the studies. But what about the between the studies? I would anticipate quite a bit of variability.

And this is the same data that I used for the study-level effect sizes. By looking into the case level, the cases united analysis, you can dive a little bit deeper. And we can get a little bit more information out of our model, even.

The NAP here would be estimated. So here on the right-hand side, we can see that multivariate meta-analysis. It's basically running a multilevel model with that-- you can see a note here at the end. It's using the cluster-robust variance--to deal with dependency, right? Because we have those multiple effect sizes within one study, we really need to do that.

And we get our estimate here is 0.94 null, that non-overlap statistics, their scale goes to 100%, 0 to 1. So that's very large. And that is confirming that very large effect size that we obtained using the study-level effect size. So that was 3 point and a little bit. That was high. We talked about that. But that's confirmed here too using the NAP. It's very high. And you can see also, here, it's statistically significant.

And then we have that omega squared, so there's some variability. And there's not a whole lot of variability going on between the studies. That would be the type of output we obtain and interpretations we can make using the meta-analysis of case-level effect sizes. And here, I used that multilevel meta-analysis and not just a simple average with the robust estimation.

But it just gives you an idea. I could have illustrated that using a regression-based effect size. I could have used a mean-based effect size. But I just applied it to the NAP. A lot of possibilities and a lot of choices to make.

And then the last approach. I do have still a bit of time to talk about that, because that's my favorite one. It's the raw data meta-analysis. I did quite a bit of research in this particular meta-analytic technique.

It's also sometimes called the Individual Patient Participant-- the IPD meta-analysis. And here, we really have that raw data that we're modeling. We have the repeated measures within-participants' part of the same studies. And then we're pooling the study together.

So instead of first calculating a case-specific effect size and combining that, here we're using the raw data. And that allows us a lot of flexibility. We model a lot of complexities. So, I’m happy to talk about that in a future webinar, but just know that it's a very flexible approach.

This is the model that we're really looking at, which is different from all the models already introduced. I don't want to overload you with too much technical details, but it's actually quite straightforward to understand. Because now, instead of two levels, we have three levels in this approach.

At level 1 are variations of data within participants. And you can see here yijk has three indices. Because we have measurements i nested within cases j nested within the studies k. And this equation, I already introduced that.

So beta 1jk, that represents the changes in level for participant j from study k. It's very, very unreasonable to assume that this change in level, which is our effect size of interest, is the same for all the cases from all the studies.

So we add a level 2, which we are looking into variations between participants from the same study. So here, for instance, beta 1jk equals a study-specific intervention effect-- so that is the intervention effect across all j cases from study k-- and then how the individual case j is deviating from those overall estimates.

These deviations are assumed to be normally distributed and having a certain variance. And that variance is exactly what we're going to get from our output. Sigma squared u1 is actually between-case variance in intervention effectiveness. So that is something important that we want to look at.

Here, we can also not assume that all the effect sizes for all the parts for all the studies are the same. So we add that study level. Then we can see that theta 1oj here equals an overall average intervention effect across all the studies and then how individual study k is deviating from that. Again, we can look into that between-study variance by looking at that variance-covariance matrix. So sigma squared V1 is between-study variance in intervention effectiveness.

Long story short, what we really are interested in and what we are obtaining by running this model is an estimate of the overall average baseline level across all my studies, the overall average weighted intervention effect across all the studies, and then also the between-study variability, the between-case variability, and the within-case variability.

That gives you additional information compared to, for instance, the previous approaches. Because here, we have that baseline level. We also have intervention effect, different sources of variability. And we can extend this model as we'd like. We can add linear trends, quadratic trends. We can model something which is called "autocorrelation" and so on and so on, all things that are not necessarily possible with the previous approaches. And we can play around with moderators at level 2 and level 3 and so on and so on.

The next slide is combining those equations in one, because that is the model that is really run here. I already talked about those variances. I already talked about the quantities of interest that we're particularly interested in.

The next slide, I'm just going to skip that one. But I just wanted to note that we also need to standardize our data when using the raw data meta-analysis. Because although we specified very strict inclusion criteria-- we have a research question, we were looking at a certain outcome, a certain intervention for a certain population-- we still will have studies included that have different outcome scales. So one study might measure a certain variable, that outcome, on a scale from 0 to 10 and another from 0 to 100.

So we actually investigated standardization methods and also looked into bias-correction methods when we're applying the standardization, so all the techniques are very well developed. And, again, ‘m happy to talk about that in a future webinar. Or if you're interested in that, you can just also reach out to me. But standardization is definitely needed.

I already mentioned the major advantages of using that IPD meta-analysis or the raw data meta-analysis-- heterogeneity, autocorrelation, and the outcome type. We can also run count outcomes and deal with nonlinear trends. I talked about the moderators as well. But the biggest and the major limitation is that we're making a lot of assumptions because we're using a parametric model. So we also might miss-specify our model and so on.

The data organization, it's nice. We don't need to do anything too much because we do need all the raw data. We don't need to first calculate effect sizes at the study or case level and then run our meta-analysis. We can just use that raw meta-analytic data set. I highlighted here something in red, because we do need to have the standardized data.

The results. Now we're running that linear mixed-effects model. Again, we're using the restricted maximum likelihood. And now we have variance of the random effects between studies and between cases within studies. And we do have that for our baseline, which is the intercept. We do that for our treatment effects.

That's different from the previous approach that I showed. Because there, we don't have baseline variability that we're modeling and so on. Here, we can see under the fixed effects that are estimates, our overall average weighted baseline level would be 5.73 in standardized units. And our treatment effect would be 8.26 and standardized units. It's statistically significant, and it's a large effect.

Then we can see there's quite a bit of heterogeneity between our cases from studies and between our studies, which is an ideal opportunity to add moderators at level 2, which is a participant level, and level 3, which is a study level, to explain why we are seeing so much heterogeneity. It could be study quality, it could be the age of the participant, gender, it could be a certain disability type, and so on and so on.

So you have your over average weighted effect size. But I'm more interested in whether that is consistent across all the study and the cases. Or if there is a lot of variability. And if there is a lot of variability, how can I explain this?

Because it could be that the intervention is actually not working-- obviously, it's not going to work for everyone if there's a lot of variability. There might be high, there might be low effect sizes. And this is just the weighted average that we're looking at. We can dive a little bit deeper and come up with explanations.

So I'm going to end it here. Just know that this presentation is organized according to three general approaches. Those three different approaches are not yet that well understood or disseminated. They also deserve a little bit more study. Further metallurgical developments are also still needed in this area. And it's not always that convenient and easy to make a choice.

But I hope with that flowchart and giving this presentation, it helps a little bit thinking about the approaches are there and how can I make a choice? You really need to look into the level of analysis, the purpose or goal of your study. Can I really make these assumptions if I choose a certain approach? And if I choose a certain approach, what effect sizes can I really use?

I'm actually also a very strong advocate for collaboration between more applied research synthesis and methodologists to further develop this type of model. So we can think together about things such as, what are we really having here as a level of analysis? Is it effect size really going to help us provide an answer to the research question?

Through having those conversations, the most appropriate approach for your specific study can be really motivated and rationalized. Also, by having those conversations, methodologists can be informed about what is further needed in this field.

There are actually quite a bit of references here. A lot of work has been done in the field throughout the years. So you can always have a look at that. And with this, I'm going to pause. Oliver, can I hand it over to you to say a couple words about the additional resources we're having?

OLIVER WENDT: Right. Well, Mariola, first of all, thank you so much for this very enlightening presentation. This was just wonderful. And for the future, a couple resources are shown here on the slides. There has already been a special issue in the Journal of Evidence-Based Communication Assessment Intervention that talks in-depth about advances in statistical analysis and meta-analysis of single-case experimental designs.

The March 2023 issue will specifically dive into meta-analysis of single-case experimental designs. So, again, there is the link. Or please feel free to email me if you need more information in this regard. And the next slide.

We are planning a more hands-on demonstration type of webinar in the fall 2023, where we will be taking the procedures that Mariola has shown us today and showcase how to apply them with the real-world data set. So watch out for that to come in the fall.

I want to use the rest of the time for a few questions. There's already the first question here in the chat, and it is from Andreas [INAUDIBLE]. He is asking, ‘it appears DCES are typically larger than comparable Cohen's d, Hedges' g derived from group design studies for the same intervention. How would you explain this difference in effect size magnitudes? Are SCED methods of intervention delivery really more effective? Does it have to do with SCDES metrics, statistics use, measurement, et cetera? ‘

MARIOLA MOEYAERT: Yeah, thank you so much for that question. And it's something very important that you're highlighting here, the scales. Because a couple of times, I mentioned, assumed to be on the same scale as Cohen's d and Hedges' g-- assumed, so that's an assumption we're making.

And can we really use the 0.2, then from 0.2 to 0.80 is moderate, and then larger than that is large? Can we really do that? Well, I do have my strong hesitations about that because of this. So if we look, the metric is still quite different. I mean, it has a same logic. You look into the changes between baseline, so control and intervention, so the change in that mean, if you'd like. And then we're using a standardization factor.

Now that standardization factor is going to provide some insights into your question. We're now using two sources of variance. We have now the between case and the within-case variance, and we take the root square of that. So the standardization factor is crucially different compared to the one of group comparison design. Because there, we were only dividing it by the between-case variance, right? So that is one thing to highlight.

And then a second thing, if we indeed look into the numerator right here, because participants themselves, they serve as their own control. So there, we also have a difference so that the mean difference, basically, is fundamentally different than if you just look into the average of all participants within a study.

What I want to highlight here is that both the numerator and denominator are fundamentally different. And because of those differences, we tend to see, indeed, effect sizes that are really larger in single-case experimental design research.

Now, regardless of using this metric, if we're really thinking about how we designed a single-case experiment, we, by purpose, select participants with specific needs. And then we apply that intervention for those participants. We have all those repeated measures over time.

And the nature of the design is also fundamentally different. There are a lot of differences. I haven’t yet talked about that much. But in that webinar here-- so

if you follow this link, we actually talked quite a bit about all the limitations and about things-- are they truly on the same scale? Can we really assume that something that is large for group comparison design, is that also large for single-case experimental design studies? So there's still a lot to do in that field related to that.

But I can absolutely explain why the magnitudes look so different if we compare the current study with the design-comparable effect size. Andre, does that answer your question? If you have a follow up about that, I'm happy to hear it. You can always reach out to me afterwards as well. Are there any other questions?

OLIVER WENDT: All right. Do we have any further questions? I also want to mention that we will be sending an email around to everybody who was registered and giving you the information where to find the recording of this webinar.

If there are no further questions, then let me thank Mariola one more time. This was really fantastic, Mariola, such a great overview on the three most prominent approaches that are currently available. And you've done it in a way that it was very succinct, yet very comprehensible, even to some of the more laypeople in the audience. I think we all took away quite a good amount of new information. So thanks again for enlightening us.

All right. I would wish everybody a good rest of the day, wherever you are. And we hope to see you again on a future webinar. Bye-bye for now.