eAudiology: Dissecting the "Evidence" in Evidence-Based Practice

Presenters: Aaron Roman, AuD, Ryan McCreery, PhD

Everyone, My name is Dr. Aaron Roman and I'm going to be talking to you today about assessing the evidence in evidence based practice. This talk was completed in conjunction and in consultation with Dr. Ryan McCreery, who aided in developing some of these slides and developing some of the content. So without any further ado, let's begin.

First of all, just some of the niceties. I have no financial disclosures to disclose. I do serve on the American Academy of Audiology's Research Initiative Committee, and I chair the State Relations Committee for the American Academy of Audiology. Dr. McCreery's disclosures can be found here as well.

Okay, so before I begin and before I dive into the slides, let me talk to you a little bit about what brought me to propose this talk. I am an audiologist by trade. I hold my audience and I work as a researcher and as a faculty member at a university. And so I get to do research a lot. And I realized coming out of grad school and starting my own sort of research program that there was a lot of things that I either needed to tune up to get to my ability, get my ability on par and learn some new stuff as well.

And when I spoke to some of my colleagues and some of my friends, you know, there were things that were kind of universal across both clinical and academic research based audiologists that people struggled with, particularly in the concept of evidence based practice. You know, what constitutes good evidence? How do we dissect articles? What do all those stats mean? So the purpose of today's talk is to aid in examining these tasks.

Now, this is not a stats lecture. I'm not going to be talking much about statistics at all. And that's because I want the average clinical researcher or the average clinician rather to understand that while you don't need a necessarily in depth knowledge of stats, sort of critical thinking and critiquing the content can get you a really long way. And that's what I'm going to be talking about today. First of all, how do we read a scientific paper?

Many people think that we approach a paper much like you would a novel, where you start at the beginning and you work your way through. However, there have been studies and there's a really wonderful infographic on EL Salvier, which I probably said incorrectly that I would highly recommend. And what's recommended for most scientific reading is not to start at the beginning and work your way through each paragraph. Instead, there's sort of an approach that you can take to dissect an article effectively, rapidly and in a way that is digestible to a clinician as a reader. So for clinicians, the way evidence is recommended to be dissected.

The first part is to skim the paper. And by skimming, it's more like screening it. Is the paper relevant to the question that you're curious about, either for academic and research purposes, or you have a client and you need to figure out what's best for my patient, given their unique circumstances. So skim the paper to see if this paper is appropriate for it. The title may mislead you.

Sometimes also review the publication date. Modern research or current research, rather, is really important because there have been studies that may have come out that either negated it, disproven it, or elaborated upon it. So you want to look into that a little bit more. And while you're reading through the paper in kind of a skimming basis, identify things that you're like, I don't really know what that means. Through skimming and identifying, a lot of that will be found in the background section of a paper or the introduction section of a paper.

You can kind of use that as your dictionary or glossary. So on second reread, that's where you're going to look through a bit more critically. You're going to look at sort of the what's and the whys and the who's of the study. So, for example, you may look at what was the population that they assessed, how did they assess it, and is it relevant to me and how I'm assessing my patients? Is this study generalizable?

Do they have a big sample size? Did they test a lot of people, or do they have a very small sample size? Based off of that, you can move to the third step, which is interpret. And that's where you actually look at the visualizations, the graphics, the tables, the charts, and see if you can kind of come to a conclusion on your own without diving into the stats, looking at the P values, the F values and things like that. Look at those to determine if on your own you could determine if something is an acceptable finding.

And a lot of those main ideas will be found really in those visuals, because people tend to highlight the main ideas of their study through visualization. It's more consumable that way. And then if you are doing this for future research or to collect for a presentation, for example, you may want to summarize it in your own terms, just to make it a bit more digestible to you. And it also aids in learning. So another method that people tend to use to address Research is this idea of pico or the pico can be pronounced different ways depending on the second letter.

But pico is an acronym and it stands for the way we look at a study, or each acronym. Each piece of the acronym rather stands for a way we look at a study. So the piece could stand for the patient, the problem, the population. It's really like the target that we want to look at. The I or the E, again, depending on your pronunciation of pico, will either stand for intervention or exposure.

So what was done to separate the population into groups? Or what differentiates the population? The next is how did they look at it? Did they use a comparison? Did they control something along those lines?

And then what was the outcome measure? So this is a really common way of assessing evidence now in audiology specifically. It can get a little fuzzy. So I have a few examples that I'd like to run through with you. So example one, here's a research question that I came up there that I kind of developed.

Do patients diagnosed with COVID 19 experience more intrusive tinnitus compared to those who. Those not diagnosed with COVID 19? So let's take a peek. If we look at, apply our PICO standards, let's first identify our patients. Who are we looking at?

So our patients are those diagnosed with COVID 19. That's what we're really looking at. What is the intervention here? This one's a little tricky, right? The intervention could be the experience of tinnitus based off of how we measure tinnitus.

But that's also going to factor in when we talk about the outcomes. What's the comparison? Well, those not diagnosed with COVID 19, that one's pretty obvious. And then of course, the outcome is experience of tinnitus. So here the intervention versus the outcome is not quite as clear as, let's say if we had a placebo and we were looking at a medical or like ototoxicity, something like that.

Another example here, this is something that I'm personally very academically interested in. Do hearing aids improve performance on the Mini Mental State examination? The MMSE is a screener for cognitive impairment. So do hearing aids improve performance on the Mini Mental State examination? So in this question, it's a little bit difficult to say who the patient is.

Patients could be the hearing aid, those who wear hearing aids. It could be those who are not wearing hearing aids compared to those who are. We don't know much based off of this research question about what the patient population is. What's the intervention? I would argue that the intervention here is probably hearing aids.

But again, it depends on the study design. What's the comparison? Well, it says improved performance improve. So we have to have some sort of comparison. We don't know if that means before or after, but that would be the comparison here.

And what's the outcome? Well, that one we do know, the mini Mental state examination. So here's one where the research question at face value is very valuable, very nice. But if we apply the Pico guidelines of assessing it, there's some gaps in the logic, there's some gaps in how we look at it. Let's look at one more.

Do children with auditory processing disorders or APD or capd, do they demonstrate lower reading scores than those without apd? So again, patient population are children. The intervention would be potentially reading. But again, this is a lot like that first one where there's not a clear intervention set up. The exposure again could be exposed to APD if, if we look at it that way, comparison is again, those without APD and again the outcome is low reading scores.

So using this Pico guideline, it's nice to say, well, if it's here and it fits here, but there's some gaps in the knowledge and that's really beneficial. However, using this topics don't cleanly fall into the each acronym in audiology perfectly sometimes and that's okay. There was actually a study done fairly recently, I believe it was 2018 and it found that Pico guidelines are really good for evaluating research for therapy. But for diagnostics and for behavioral research, Pico guidelines leave something to be desired. So that means we have to rely on our own intuition and our own knowledge to figure out whether or not a study is well developed, if it's strong and if it makes sense for our client.

So what I will be spending the rest of this talk on is helping you identify some of those skill sets that may help you in critiquing articles and determining if the evidence you're presented is good evidence. And we're going to look at it in three different ways. We are going to look at the patient population or the sample population. We're going to look at the study design and we're going to look at interpreting the outcome. So let's first start with the population.

So first, what do we need to know about the population? You're always going to see a sample size or you should always see a sample size in any study that is going to look at group populations or group

statistics. So your sample size also symbol as N refers to the entire group that is being assessed in your study. So what makes for a good sample size? That's tricky.

What you would really like to see is the term power analysis. A study that has completed a power analysis and holds to it indicates that they used mathematical modeling to determine how many people it would require to make the study as it stands effective. Now, not all do that. And so we kind of have to use our own deduction to figure that out. So if we see a small sample size, if we see a group of 10, for example, that is potentially damaging to the study because it increases the chances of finding what's called a false premise as true.

So what it does is it presents false evidence. It might say that something's effective when in reality it's not. Or it might say something's not effective when in reality it is. So what this means is we can't globalize the findings, we can't generalize them. And so small sample sizes weaken studies such as that.

There's a common belief that you can't have too big of a study, but you actually can also have too big of a study. So depending on the statistics that's being used, if you have a population of 300,000 people that you're assessing, first of all, what may happen is it can cause kind of negligible results and still call them statistically significant. Think of it this way. If I had a group of 10 and I was administering a test and 10 people got an 80% or, sorry, eight people got an 80% and two people got 100%, well, the average is going to be roughly an 85ish or so. Now imagine if I had 100 people and 80 people scored below the range that we were looking for, and 20 people scored above the range we're looking for.

Even though percentage wise, that makes sense, the number impacted is much larger now. And so the larger you expand it, the more that's going to impact. And so the large sample sizes may give you statistically significant results that are really kind of. When we say clinically negligible, think of it like this. It may show that there's a significant difference between a PTA of 35 and 37.

And clinically, we may not say that's significant. So we do want to be cautious of overly large sample sizes. We want to be cautious of overly small sample sizes. We want kind of the Goldilocks effect. We want to find something that is just right.

And the best way to do that is if the researcher reported a power analysis. Now, I'm not advocating that you sit at home and do a power analysis by yourself. Instead, just kind of sit back and think logically. How many people are in this study? Okay.

What are the demographics of the people? Where did they get them? And that's going to influence the results. So you could use those skills to determine what's in that middle ground. And that's what we need to focus on.

To help you with that, I want to highlight some common biases that you see in studies. Now, when you hear the term bias, especially in terms of research, you may feel like it's inherently a bad thing or completely negative negates the study. And in some circumstances, yes, that's true. However, in most cases, there is always going to be some level of bias. It's up to the clinician who interprets it to determine how and how much that bias is going to influence the findings.

So I just wanted to highlight some of the four most common biases you may see in clinical research. So the first is something called selection bias. And selection bias is that you use a criteria that is very specific for one cohort and then a different criteria for a different cohort. So that's not super common in prospective studies, which I'm going to get to. But if you use different criteria, you're obviously going to find different things.

So you want to be careful about, okay, what was the criteria that they used to recruit kind of all subjects? And that will help you. Another one is this channeling bias. So that means that whatever the

patient's current medical status is, it will dictate the cohort. So it might just be a matter of convenience, like, oh, you have hearing loss, you're going to go into the hearing loss group.

Irregardless of what you're necessarily looking for in the hearing loss group. An interview bias might be you may change your interviewing techniques. So an example that I thought of or that I've seen in clinic is, you know, if you're asking someone a patient history file and you look over their intake form and they seem to have, they indicate that they have no contraindications to hearing. So let's say they're like 25, they have no negative history, or they have no past history of any sort of trauma, noise exposure, things like that, you're going to go through that interview rather quickly compared to someone who's very high risk for hearing loss. Let's say you're interviewing someone who's 75, who has heart disease, who has things like that.

You're going to ask a lot More questions, a lot more in depth. Well, that's interview bias. You're changing your questions based off of what's presented in front of you. And so a good study or a well developed study should ask the same questions of all participants, irregardless of what category they would fall into. And the final one just to be aware of is if there are any confounds to the study.

So that means that there may be a third piece of information that would explain the connection between two factors. And I'm going to get to that in just a little bit. So let's talk about some study designs. But before we do, I just want to give you a little task. So I created a study here.

I'm conducting a study to see if long term hearing aid use correlates with lower perceived stress scores on a stress questionnaire, the perceived stress scale, if you're familiar with it. In this case, the sample size is 33 hearing aid users. Between 33 and 85 participants are grouped into two groups. Long term hearing aid users greater than 5 years and short term hearing aid users less than 5 years. All participants complete a face to face interview and fill out an eight item questionnaire regarding to perceived stress.

So looking at that study, what are some problems that could arise with the design? Well, first of all, sample size. Right. Only 33 people. What about my groups?

If you look at the, the overall sample, it goes from 33 to 85. Do you think that stress is going to change between 33 year olds and 85 year olds and anyone in between? Absolutely. So I should probably do something that helps separate those a little bit better. The two groups, long term hearing aid users of greater than five years and short term hearing aid users less than five years.

What did I use to justify that? Well, it doesn't say. It could just be an arbitrary guess. So these are things you want to look at when you're first looking at a study. Because if you take those at, if you take the results at face value without considering them, you may be misled.

And that's not what we want. So let's look at some study designs that are common in clinical research. The first is a randomized control trial. You may have heard of these very common or an RCT randomized control trials. What they do is they take a participant and they place them blindly into one of two groups.

In one group they will be administered the treatment or the intervention and in the other they will be administered, usually a placebo. So a very common example that's probably relevant to modern Day is the development of the COVID vaccines. So they would take a group of people and they say, all right, this person you're going into group A, this person you're going into group B. One person will get the placebo, the other will not. So this allows for blinding, meaning that the participant doesn't know if they're getting the placebo or the actual intervention which is going to benefit because they're going to exhibit how they imagine the results or the behaviors that they would perceive and will the researcher.

So the researcher won't interpret. And that stops that interview bias, that stops all those sort of biases that we talked about a little bit before. And so for that reason, RCTs are generally considered one of the strongest models of research design. More commonly in audiology, you may see a cohort study. And a cohort study is a study in which people are placed based off of whether or not they may meet a certain disease or if they're either in an experimental group or a control group.

So, for example, you may see one where you have normal hearing individuals and you have individuals with hearing loss, and you're going to do the same interventions on both and see how the outcomes differ. So that might be a good example of a. Excuse me, prospective study or a cohort study.

The final one that I'm going to talk about today is a case control study. Case control are a little less common. You may see them. I read a study recently that looked at heart disease and hearing loss, and they used a case control study. So a case control study is retrospective, meaning that they look at something that's happening currently and they look back on it to see what has changed.

So, for example, in the retrospective or in the case control study, they were looking at people who were diagnosed with diabetes, and they wanted to see if some of the symptoms of diabetes were more common in a group of hearing. A group of individuals with hearing loss versus a group of individuals without hearing loss. And that's what a case control study is. So you are looking at a certain factor and you're looking back on the experiences of one group and another. The groups kind of counter each other in that one will have the disorder or a.

Some sort of target behavior, while the other group will control for it, thus the control group. So you'll see those occasionally. I would say, sort of anecdotally, the most common one I see is a cohort study or prospective cohort study.

Okay, so let's look at my original study again. We're looking at a sample size of 33, 33 to 85 year olds. And these are grouped into two people, two groups based off of their hearing aid usage, whether they're long term or short term. So based on that study, would you say that this is best described as a randomized control trial? Would you say this is a case control trial or is this sort of a prospective cohort or prospective study?

What would you say? So I would say this is a cohort study, right? Because I'm not randomly putting people into a long term or a short term hearing aid use. That is a characteristic that they just exhibit at the time of my study. And I want to say that these three study designs are not the only three.

There are many different types of study designs. These are probably the three most common that you'll see, but they're not necessarily the be all, end all. However, I would argue that for this study, we're looking at a cohort design.

So how can I alter this study to make it a randomized control trial? Well, I can get rid of that whole third bullet, right? And instead I could take a group of people that are 33 to 85 and I can randomly put them into two groups. One, I would say maybe I do an intervention. And so I say something like, okay, I'm going to intervene with a hearing aid at this point in time and I want to see how your stress levels change.

So in this modification, what I would do is I would take a group of people who have hearing loss who are considering hearing aids. In some maybe I would get hearing aid A, some I would get hearing aid B, and I would fit them all. And I would look to see, do people in the hearing aid a category perceive less stress than those in the hearing aid B category. That's one way to do an RCT based on this sort of theme.

Okay, so now let's talk a little bit about interpreting results. So one of the most common questions, I actually fielded some questions out to my colleagues and I asked, you know, when you're reading research, what kind of holds you up? What confuses you? One that I saw over and over again is what is a p value? So this is probably the most statsy I'm going to get tonight.

But what is a p value? First, just let's get some basic research terminology down. There are two things you need to know before we talk about what is an actual p value. You have to know the idea of what is a null hypothesis and an alternative hypothesis. A null hypothesis is the idea that there is no difference that is significant or there's no significant difference between group A and group B, or observation A and observation B.

So whatever you're looking at, there's no difference between them. The alternative hypothesis is that there's a significant difference between group A and group B. So a P value is a value that's found statistically. That refers to the probability that the null hypothesis is correct. So the probability that there is no difference.

So you can kind of think of it like this. If you're a big fan of like Law and Order, something like that, you probably have heard the term innocent until proven guilty. Same thing here. Null hypothesis until proven alternative. So the P value refers to the probability that the null hypothesis is actually true.

So we set what is called the P value, the cutoff P value in research, at 0.05 or 5%, meaning that there needs to be a 5% chance that the null hypothesis is true or less. That does not mean that there's a 95% chance that, that the alternative hypothesis is true, because there's a lot of confounds that can happen. However, it means that in general, there is a 5% chance or less that whatever you're looking at, the null hypothesis is true. So it's very unlikely that the null hypothesis is true and therefore the results are significant.

As I said before, just beware, because sometimes we'll see that a finding is statistically significant, but it's clinically insignificant, like that PTA example I gave earlier, and vice versa. Sometimes we'll find something that is significant, statistically insignificant, but clinically significant.

Another thing to be kind of cautious about is correlations. So correlations are used to find whether or not a relationship between two factors exists. You've all probably heard the phrase correlation does not equal causation. And so to highlight this, I came up. This is just random data that I generated, but this is hours of hearing aid usage related to an outcome score between 0 and 30.

And they found that the longer the hearing aid usage, the higher the outcome score. If you look at the correlation, the correlation is 72%. It's a very strong correlation. However, hearing aid hour usage is really a strange measurement, right? Because hours of usage is based off of motivation, it's based off of severity of hearing loss.

So all of those factors might be going into the outcome score. And hearing aid usage is just something I picked. So correlation is not a be all end all. It's not a causal finding. And so I do want.

If you're ever seeing a correlative thing, it's not that you have to disagree with it or disregard it, but just be cautious because there's a likely other factors that are influencing it, other factors to consider. You know, look, you don't have to be a stats master to interpret the results. The first thing you want to look at is what are the variables that are being compared. So independent variables, going Back to Stats101 or Research Methods101, independent variables are variables are not influenced by study design. So time, group, gender or gender expression, anything like that, whatever you want to look at, that's your independent variable.

Dependent variables are manipulated by the study. So like the MMSE that I said earlier, you know, literacy scores, when I talked about the children with apd, those are dependent variables. So what you

first want to do is identify the independent and the dependent variables. Next thing you should do is decide whether or not the comparison makes sense. So a comparison will make sense.

The first thing you have to do is identify what comparisons are being present. Is there a between subject or group design, meaning that you have group A and group B and you are comparing the performance between the two, or is it a within subject group design, meaning that you are looking at the same person just in different time points or on different point on different measures? So if I said I want to look at perceived stress score, throwback to that at the beginning of the pandemic versus now on the same people, that would be within some studies have both. You have. I want to look at perceived stress scores between those who identify as male and those who identify as female from the beginning of the pandemic to the end.

Well, there's two measures there. There's time, that's an independent variable. And then there's gender identity, which is a second independent variable. And then the dependent variable is perceived stress measure. So identifying those will help you out a great deal.

So to wrap up, let's evaluate one more time. Looking at my study that I proposed here, what are the independent and dependent variables of this study? Now, that's not really fair because I kind of gave it away. But think about it. Looking at my study, I would say that the hearing aid groups are the independent variables and perceived stress is the dependent variable.

Now, looking at this, would you say that this is a between design or a within subject design or both? Based off of the language that's written here, it's probably between. We have two groups and we're just comparing scores at one time point between those two.

All right. And that's 30 minutes on the dot. So thank you all so much for attending my talk. If you have any questions at all, my information is here. I'm happy to talk about this further.

I love talking about research methods and research development, so please reach out. The information for Dr. McCreary, who again served as a great consultant and a great contributor to this project, can be found here. And all references will be here. So thank you so much and have a wonderful day.