A Sound Decision at Every Level: The Research Behind Starkey's Genesis AI Technology

Presenter: Maddie Olson, AuD, CCC-A

- Thanks, Anna. So, yes, I'm Maddie Olson. I'm a research audiologist at Starkey, and today we'll be talking about Genesis AI, all the research that went into Genesis AI so that you can make a sound decision on every level with your patients. So this presentation will cover all the research, what we've found from our interactions with our participants, so you can learn what to expect from your patients when fitting Genesis AI. So the learning outcomes for this course are listed here. So you'll be able to list the types of clinical studies employed by Starkey to validate Genesis AI benefits, describe the benefits of additive compression for the Genesis AI wearer, which were identified during the clinical trials, and describe telecoil use and speech intelligibility features investigated during the clinical trials.

So what goes into the research that we do at Starkey, and it's kind of two pieces. I think of it as being both carrying technology that's also backed by science. So to me what this means is that when there's features being proposed to the large team at Starkey, early in the process, we like to think, how does this benefit a patient, or how do we anticipate that this might benefit a patient? And if that benefit is meaningful, then we choose to pursue that feature or that product so that it can eventually get integrated into a program which eventually becomes a launched hearing aid. So we wanna make sure that there's a perceivable benefit and this is just something we're thinking about when developing products.

Then the next step is to make sure that we can develop studies to assess whether patients or participants perceive that benefit in either a measured or just perceived way. So it's both kind of thinking through what the needs are of our patients, and then using science to make sure that we're really meeting the needs for those real life hearing aid users. And the team that does this is the clinical and audiology research team who's pictured here. So we have a team of individuals of various backgrounds. So we have clinical audiologists, we have research audiologists, we have research

scientists, and we do have some team members who are both audiologists and research scientists. So all of those individuals together make up the team.

And I can say that this whole team was involved with the assessment of Genesis AI, which was really such a cool experience because that meant we were all collaborating, sharing knowledge, learning from each other, and collaborating to make sure that we were assessing various aspects of hearing aid benefit and performance with Genesis AI. So we could come to present all of this knowledge to help you fit your patients with Genesis AI. Altogether we had 11 team members who were critical in evaluating Genesis AI. That was a group of research audiologists, hearing scientists, a research coordinator, and clinical audiologists. And kind of the main difference between the clinical and research audiologists, the clinical audiologists are essentially in charge of maintaining our database.

So they see all of our participants for updated hearing evaluations, updated earmold impressions, they do the intakes, a cognitive screener so that we can make informed decisions when we're recruiting our participants. And that's because the research audiologists and scientists typically have not met these participants at this stage. We're just recruiting based on what we see in a database. The focus of today's presentation will primarily be on clinical validation. So that's the procedure we use to assess new products and features. And this is to support regulatory requirements and marketing claims and just to make sure that the feature is working or the product is working as expected for the intended population, once everything kind of comes together for one final study, which gates us from launching the product.

We also do pre and post-market clinical research studies. And these are to assess benefit and performance of our products. We often do these with external collaborators too. So that could mean a university, a research institution, hospitals or clinics. And really getting a lot of different investigators involved helps us to ensure that we're

meeting the needs of various investigators and clinicians as well as varied patient population pools. So all of that helps us to really prove that we can have good clinical performance. So what are the goals of clinical research? So essentially we all know how hearing aids are supposed to work, right? So the goal of clinical research is to make sure that when everything has all come together, that it really is working as intended for the intended user, hearing aid users.

And we answer those questions with a lot of other questions, right? So we ask, does it work as intended? Can people use it effectively? So we'll do that with usability testing. Do people like it? So we'll assess preference either from their own hearing aids or from a competitor. What benefit does it provide? And this could either be measured benefit that we get from a laboratory assessment or perceived benefit. Is it safe and does it meet the needs of both clinicians and patients? And really a goal of clinical research is primarily to assess, does it meet the needs of patients? But when we are thinking about meeting the needs of clinicians, we do turn to external collaborators to make sure that we're getting lots of opinions on how it's working for various clinicians of various backgrounds and specialties.

And that's really because the clinical research audiologists are typically a little bit too in the know to give an unbiased assessment of, is it meeting our needs? So we do really need to kind of source other opinions from various backgrounds to make sure that it's meeting lots of different clinicians' needs. What types of clinical studies do we do? So those include both laboratory and field studies. We typically start piloting a new product or feature with normal hearing participants just to make sure we work out any of the kinks before we kind of give this to a real participant, so kind of one that's recruited from our database and has hearing loss. So we kind of pilot and then if we determine that it's kind of working as expected and there's no bugs or hiccups, then we start to do studies with the hearing impaired participants.

And these clinical studies are various types. So there might be prototype evaluations where we're really just kind of looking at the shell of a hearing aid, making sure it sits on the ear correctly, is comfortable, the coupling to the receivers working as anticipated, that sort of thing. So in a prototype evaluation, there's likely not any functional pieces of the hearing aid, there's not a speaker or receiver, it's just kind of a prototype. We'll also do feasibility testing. And this happens early in the process with a new feature or product to ensure that it's really working as designed. So at this point, we'll have pretty strict criteria for what it means to pass a feasibility study. And if it doesn't pass, it either goes back to the drawing board or they totally reinvent the feature 'cause it didn't meet our expectations.

Or we might tweak it a little bit, and then do another round of feasibility testing. Or we might say, you know, this was a great idea on paper. When we tried to implement it, it did not work out as expected and we might scrap the idea. And to me, that's really a good sign of the research process. We should be having results that are either negative resulting in no feature or positive and the feature gets implemented into the program, and then later into the launched hearing aid product. We'll also do human factors or usability studies. And we have human factors engineers on staff who help us to kind of organize and ensure that we're doing quality, usability studies with our participants.

We'll do these for the hearing aids, for the mobile app. And the human factors group does this with the fitting software, but again, with external audiologists, we also do fit and comfort studies. And these also kind of involve cosmetic studies too, where we'll have either internal or external audiologists rate the cosmetics of hearing aid fitting as well so that we have a better idea of what our patients can expect once we do this on a larger scale. The primary focus of today's presentation will be on the product and feature validation. And that's really kind of the combination of all of these pieces. It's when we have the final hearing aids, the final fitting software, the final mobile app, and we're assessing them all as a whole and ensuring that it meets our performance and

clinical benefit outcome measures and our safety outcome measures before we decide to launch the product to you.

We also do competitive evaluations, and this can be in the lab typically. And in these types of evaluations, they're really quite informative and helping us to learn what our competitors are doing well, where we stand out and what we can learn from that to improve our products moving forward. A primary goal of feature validation is actually to substantiate claims. These do sometimes come from other types of studies, but it is kind of a main goal of feature validation. And the goal here is to make sure that any claim we make about a product that relates to clinical performance is substantiated by clinical researches' data on that product or feature. So anything you see in a claim sheet that relates to performance, we've validated.

We'll also sometimes do pivotal studies on clinical performance, safety or effectiveness. And these pivotal studies typically involve a little bit more control or we might have blinding or a control group to just make some more specific comparisons of our product to a different product in a more controlled study. And all of our research studies are subject to regulations and regulatory bodies with the goal of ensuring that we're doing good clinical practice and good research. So some of the regulatory bodies and regulations are listed here. The first is the Medical Devices Directive or Medical Devices Regulation. And this is a European regulatory body. So if you see a hearing aid that has the CE marking on it, that does indicate that MDD or MDR has approved that device for sale in Europe.

There's some ISO standards there as well. And that ISO is the International Organization of Standardization. The first is about clinical investigations of medical devices for human subjects with some guidelines to ensure good clinical practice is taking place. And then the second ISO standard is about the application of usability engineering to medical devices. So there's some good guidelines there to ensure that

we are assessing the usability of the medical devices in a high standard way. The HIA or Hearing Industries Association has some guidelines for hearing aid manufacturers for substantiation of benefit or performance claims. So we do have some guidelines we must meet in order to make those claims about our products. So lots of regulations all to ensure that we're doing good research.

So how do we do all of this research? It's with volunteers and a lot of them. So we do have a database of over 1500 individuals and we've recruited those individuals from the local community here in Minnesota. So we'll recruit hearing impaired individuals from ads that we post, from word of mouth, from refer a friend, so that we kind of have a pretty diverse database. And in that database, we've got lots of different characteristics of those participants that we have access to. So we have a good idea of what their audiogram looks like, their experience with hearing aids, what type of cell phone they have. We get a sense of their digital literacy from different intakes that we do.

We also do a cognitive screening on each of our participants so we know what their MoCA was the last time they came in. And all of that really helps us to make informed decisions on who we recruit for our studies. 'Cause like I mentioned earlier, we typically haven't met these individuals before we've recruited them. Participants are paid for their participation. However, we do not allow them to keep the hearing aids after the study is over. We're cautious of overly incentivizing the participants 'cause we do wanna make sure that they kind of have an open, honest communication with us, and we fear if we, you know, gave them too much incentive, they might just tell us what we wanna hear, right?

All of these protocols are approved through an external Institutional Review Board, and the IRB is really there to make sure that we're protecting the rights of our participants and conducting a safe and effective research study. All right, so let's share a little bit

about all the research that went into Genesis AI. So studies were conducted over a period of about 15 months. So that includes the pilots and the validation and any other feature investigation type studies that we did leading up to the launch of Genesis AI. All of those studies involved over 400 hearing impaired individuals. For the design validation alone, we had 550 hours of lab testing and 11,000 hours of field wear time. So that's just from the design validation.

We had much more on top of that with all of those other studies leading up to the launch. So the design validation, which is the primary focus of today's discussion, involved 75 participants, and they had a range of hearing losses from mild to severe. All of the hearing aid users did have sensory neural hearing loss in both ears. The participants came to see us for five to six laboratory sessions, and altogether they were in the study for about six to eight weeks. Each of those sessions was a full two hours long and it was really kind of jam-packed with lots of content, both interview, informal feedback, laboratory testing, questionnaires. So we really kind of kept them quite busy and got to know each other pretty well, which of course then helps for us to learn from each other what's working well, what needs to be fine-tuned, and how do these kind of compare to their personal hearing aids, too.

Now we had several audiologists on this study and altogether we conducted 450 pro fit sessions, we did 200 real-ear measurements, we administered 600 questionnaires, and conducted 850 speech test conditions. So as you can see, we were quite busy. These participants were also quite busy, too. Now on the right here, we'll see all of the different things that we conducted as part of the design validation that we won't focus on all of these today. I don't think we have enough time for that probably. So in a bit, I'll tell you which ones we are going to focus on, but at a high level, we did hearing aid fittings and real-ear measurements on all of the participants. We used the optimized vent recommender, we did usability testing of the hearing aid, the charger and the mobile app.

We tested specific My Starkey mobile app features like activity tracking, fall detection, and edge mode, plus we did speech intelligibility testing, both in quiet and noisy backgrounds. We did TeleHear appointments that involved us also collecting in-situ audiometry data. We tested the CROS system, we streamed with accessories and did a general streaming sound quality assessment. We tested the loop and telecoil and we evaluated performance for phone calls. Since we don't have three hours, we'll kind of focus on some key factors here. So we've kind of gone through and picked what we think are the most important pieces to talk about today. And these pieces really do pertain to Genesis AI's performance straight out of the box.

So these are kind of the baseline functionalities of the hearing aid that you can expect from kind of day one with Genesis. So those things will be additive compressions. So we've got the new dual compressor that's also using e-STAT 2.0 and how both of those things impact speech intelligibility and sound quality will be evaluated. We also will discuss the noise management system, which includes both noise reduction and directionality. And then we'll also discuss those impacts on speech understanding in noise. And finally we'll talk about the telecoil and phone performance. So first we'll dive into additive compression and e-STAT 2.0. So e-STAT 2.0 is our proprietary new fitting formula, which complements Genesis AI's new compression architecture. And what you can expect from e-STAT 2.

0 is 4-6 dB more gain for soft sounds and this results in a 10% higher speech intelligibility score for soft speech, one compared to our previous technology, which is really not trivial, that's statistically significant and I think it's clinically meaningful as well. With e-STAT 2.0, you can also expect more gain at higher frequencies and this is in the range that's above 3000 hertz. With improved vent modeling, you can also expect a more accurate fitting out of the box assuming you've run the feedback canceler. So for those not familiar with this feature, there's on the feedback canceler

initialization screen, there's a little toggle called optimize acoustic model. And if you have that enabled when you run the feedback canceler, what it will do is check what it believes the venting is in ear, crosscheck it with the venting you've selected in the fitting software under the acoustic options.

And if it thinks that there's kind of a discrepancy there, it will recommend that you update the modeling to better reflect what's measured on ear. So for example, if you have an occluded earbud and that's in the ear and selected in the software, if that occluded earbud's not sufficiently doing its job and it's more open than you'd expect, you might get an updated recommendation that would update the modeling to really reflect more accurately how that occluded earbud's behaving in the ear. So I'm really not anticipating that you'll see a recommendation 100% of the time. It's probably closer to 30 to 40% of the time that you might receive a recommendation. If you do when you apply it, we're expecting that fitting to be more accurate in the ear.

So here we're just looking at a plot comparing e-STAT and e-STAT 2.0. So we've got frequency on the x-axis and simulated real-ear on the y-axis, and the blue is soft inputs, purple is moderate, yellow is loud inputs. Now the solid line is e-STAT. The dashed line is e-STAT 2.0. So you can see kind of in some of those speech heavy regions you're seeing more gain with e-STAT 2.0 compared to e-STAT. And this is most prominent in soft and moderate inputs. You see there's not much change with loud, so you can expect that things are still quite comfortable for loud speech. And then you'll also see that there's extended bandwidth with e-STAT 2.

0. So in those highest frequencies where e-STAT starts to roll off, e-STAT 2.0 still applying gain there. So you might be thinking all of this added gain are things going to sound tinny or unnatural. And really what we've found is no, people seem quite pleased with the sound quality right out of the box and we attribute that to the extended bandwidth. So instead of it getting kind of peaky and sharp sounding, it sounds more

natural because of that extended bandwidth. And it really is kind of a delicate balance with new fitting formulas, of course, of trying to achieve good audibility and good sound quality. So we'll talk about both of those in a little bit more detail.

Okay, so let's talk about audibility first. So we assessed audibility in the lab using the CNC word recognition test where we were specifically targeting soft speech. So the CNC word recognition test is where you have the carrier phrase, say the dog again. The participant or patient repeats back the target word to be dog in that example. They either get the whole word correct or they can get partial points for getting some of the phonemes correct too. Now participants were tested in the sound fields at zero degrees azimuth where the speech was presented and it was presented at a pretty soft level of 50 dB A, which did make this task pretty difficult as it's quiet and there's not much context with the CNC word lists.

Now we had 30 participants who did this testing and they wore two pairs of RIC devices, one, Genesis AI RIC, and one, previous technology RIC. In each of those conditions, they listened to 50 CNC word list. Performance is on the plot on the left. So you'll see with Genesis AI, on average they scored about 73% correct. With previous technology, they scored 61.5% correct. So you see almost a 10 or more than a 10% increase in speech intelligibility for soft speech when comparing Genesis to previous technology. And this is statistically significant, and again I think would be appreciable to the patient as well. We did wanna look at this data in a little bit more depth. So what we did was we split up those 30 participants into groups that were grouped by hearing loss.

So you can see we did have quite a wide range of hearing losses here as we were able to group a PTA of 35 dB, which is quite mild, a PTA of 50 dB, which is more moderate, and a PTA of 65 dB, which is moderately severe to potentially severe. And what you'll see here is in each of those subgroups, we are seeing benefit over previous technology

with Genesis AI, but that benefit is quite large in the PTA 65 group, and those participants really did need the most help in these situations to understand that speech clearly. So with Genesis, they're not quite getting as good of performance as those with a lesser degree of hearing loss, though I'm not sure we would expect that, either given that they might have some degradation in their auditory system, meaning they can't achieve the same clarity as someone with a PTA of 35.

So overall for those who needed the most help here, they saw the most improvement with Genesis over previous technology. Now let's talk sound quality. So with Genesis AI, we also did a separate study outside of validation where we queried normal hearing listeners, and these listeners were totally blinded to what they were listening to. They were just tasked with sitting down to a self-paced listening test over headphones. They didn't know that they were listening to hearing aid recordings or what exactly they were listening to. They just knew that their task was to rate the sound quality on a scale from 1 to 100. And each time they were doing kind of a comparative listening task of A and B.

A and B did randomly switch as they were listening. So they weren't consistently picking A. Then what we did to make these recordings was we used KEMAR and we recorded Best Fit settings using Genesis AI and using previous technology, and we made several recordings. We used two audiograms, two levels, and three signal-to-noise levels, all to make sure that we weren't favoring one audiogram over the other and to kind of learn, did Genesis have difference in performance depending on audiogram or signal-to-noise ratio. And the results show that listeners rated Genesis higher in quiet and in noise. So 93% preferred Genesis compared to previous technology when listening in quiet and 84% rated Genesis sound quality higher than previous technology when listening in noise.

And that's really quite substantial. That's a pretty significant improvement in sound quality over previous technology. And the one thing we are really curious about here is with the new noise management system, we wanted to make sure we were still preserving speech intelligibility and sound quality because as we know, sometimes noise reduction can actually cause an interference in either of those things. It can degrade sound quality or degrade performance. And at least with this normal hearing listener group, it looks like with the updates to the noise management, the improvement in sound quality is quite substantial over previous technology. So that leads us into noise management. So for this, we used, back to the validation study, our hearing impaired participants and they did a speech-in-noise task in the sound booth.

Now they did four different conditions. They did without hearing aids, they did kind of amplification alone, which would be omni microphones and no noise reduction. Then we added in noise reduction. So that's kind of simulating an omni device though this was measured on RIC devices. And then we also did the adaptive directionality plus noise reduction on, which kind of simulates the personal program out of the box. Now 43 participants did this testing. They were seated in the sound booth where speech was presented at zero degrees azimuth, so the green speaker, and these were IEEE sentences. They were presented at a participant specific level and that was the level at which they achieved about 50% correct on the sentences or their SNR 50.

And this was measured using the omni plus noise reduction off condition, and then whatever level they achieved there, that was used to present for all four of the other conditions. And this just kind of ensured that we were avoiding ceiling and floor effects. So we wanted to be able to see really the improvement from one condition to the next in a pretty difficult listening environment. And we made it all the more difficult because we did surround them in noise. So that's playing from the red speakers in the picture. That was speech babble that was summated to a level of 67 dB A, so kind of

simulating a pretty moderately loud restaurant. Now the results on the right show that unaided on average participants were getting about 35% correct.

Then if we go to the yellow bar, you can see from amplification alone, we do see a significant improvement in speech intelligibility. It's closer to about 50% now. Then when we add in noise reduction, that's not a significant difference between the yellow and purple plots, but it is kind of trending in the upwards direction, though that is not statistically significant. Then you'll see a pretty big bump in performance when we go to the adaptive plus noise reduction on condition, which is also just the default personal program. So what you can see from the default personal program in gray to unaided is that we have a 30% increase in speech intelligibility in this very difficult listening environment over unaided.

And another good thing to see is that the yellow and purple plot bars are not significantly different from one another indicating that noise reduction is not having a negative effect on speech intelligibility. And we also asked for some subjective ratings for speech and noise. So they rated listening effort and clarity while they were doing this task and both kind of show a similar trend to the performance plots. So it seems to agree that how they're performing and how they feel they're performing are in agreement here. We also did a QuickSIN with these participants using Genesis AI. And if you're familiar with the QuickSIN, you know there's six IEEE sentences, each with five target words. The SNR do decrease from 25 to 0 dB SNR in 5-dB steps.

Now we presented speech from the front and noise from the rear, and then these are the results on the right. So we used their unaided performance to categorize them into what SNR loss group they fell into. So we've got a variety of categories from severe to normal. These are the categories provided by Killion et al in the literature. And what you can see is for the severe group, they jump two categories when you add in hearing aids. They go from a severe to a mild loss. The moderate group goes from a moderate

to a mild. The mild loss goes from mild to normal, and the normals just say normal. So this kind of shows that for those that need the most benefit, they were receiving more benefit from Genesis AI compared to unaided.

Finally, we'll quickly talk about telecoils and phone use. So for this setup, we were assessing all of the different telephone options that you have available as a clinician. And we were hoping to have some guidelines based on these results to kind of guide you which option's the best option for your patient. But as you can see, each option seems to be performing really well. So instead of which option performs the best, the conversation can be which option is most convenient for your patient. So what we did here was we had them listen to speakerphone, Bluetooth, autophone and autocoil, and they were listening to connected speech test sentences and rating how much of that passage they felt like they could understand using the speech intelligibility rating.

So really all of the performance they self-rated as in the 90 percentages. So upwards of 90% correct. So again, really no bad options here. It's more so which option is most convenient for your patient. We also did an assessment of the loop program in a laboratory environment. So we had three different types of recordings that patients listened to. An unobstructed recording, meaning there was nothing between the speaker and the listener, a mask recording where KEMAR actually wore a KN95 mask and a barrier recording where there was a plexiglass barrier between KEMAR and the listener. And we really tried to make this environment as difficult as we could to kind of show how these degraded listening conditions perform with a hearing aid microphone alone and with a loop.

And most of the places that have loops are also pretty difficult listening environments. So we felt this was pretty realistic. You can see performance significantly improves when adding in the loop when compared to the mic alone condition. And really I think that's probably not a surprise to anyone, but just kind of a shameless plug here to keep

utilizing loops because I know widely they're kind of underutilized and to continue participants to keep asking for loops in the environments that are important to them. And then just to demonstrate this point further, I do have some recordings that our team took at the Minneapolis airport where we recorded how the hearing aids performed with hearing aids alone and when adding in the telecoil.

So let's see if we can hear at all what this message is in the hearing aids only condition. Okay, maybe got a phrase or two. Let's see what happens when we add the telecoil.

- [Man's Voice] In order to prevent someone from placing dangerous items in your baggage. Do not leave your bags unattended and do not accept items from unknown persons to carry on to the aircraft.

- So it's really night and day, and I mean I much prefer at the airports when they have the banner with the information kind of posted visually so everyone can see. But that's not the case as you can see at Minneapolis. So when you're relying on hearing alone, adding in the telecoil can make a huge difference for your patients. All right, so the last slide here, we've just got some testimonials from our clinical validation participants. And a couple things that stand out here. People were complimenting the Bluetooth features, the new app, the battery strength, the sound quality, and the comfortable design. So we really received quite positive feedback from our participants and we hope you're receiving the same positive feedback from your patients.

That is all I have for today. We can open it up for questions if we have time. I know I went a couple minutes over.

- [Anna] I am not seeing anything in the Q&A right now. If anybody would like to, they can send in those questions. Otherwise, we can go ahead and end today's class.

- All right, thank you, Anna. Thank you everyone for attending. If you think of any questions, certainly let me know.

- [Anna] All right, thank you so much.

- Thank you.