Comparison of In-Situ and Retrospective Self-Reports on Assessing Hearing Aid Outcomes

Yu-Hsiang Wu[1]   Elizabeth Stangl[1]   Octav Chipara[2]   Anna Gudjonsdottir[3]   Jacob Oleson[3]
Ruth Bentler[1]

[1] Department of Communication Sciences and Disorders, The
    University of Iowa, Iowa City, Iowa
[2] Department of Computer Science, The University of Iowa, Iowa City, Iowa
[3] Department of Biostatistics, The University of Iowa, Iowa City, Iowa

Address for correspondence Yu-Hsiang Wu,
yu-hsiang-wu@uiowa.edu

Abstract

Background   Ecological momentary assessment (EMA) is a methodology involving repeated surveys to collect in-situ self-reports that describe respondents' current or recent experiences. Audiology literature comparing in-situ and retrospective self-reports is scarce.

Purpose   To compare the sensitivity of in-situ and retrospective self-reports in detecting the outcome difference between hearing aid technologies, and to determine the association between in-situ and retrospective self-reports.

Research Design   An observational study.

Study Sample   Thirty-nine older adults with hearing loss.

Data Collection and Analysis   The study was part of a larger clinical trial that compared the outcomes of a prototype hearing aid (denoted as HA1) and a commercially available device (HA2). In each trial condition, participants wore hearing aids for 4 weeks. Outcomes were measured using EMA and retrospective questionnaires. To ensure that the outcome data could be directly compared, the Glasgow Hearing Aid Benefit Profile was administered as an in-situ self-report (denoted as EMA-GHABP) and as a retrospective questionnaire (retro-GHABP). Linear mixed models were used to determine if the EMA- and retro-GHABP could detect the outcome difference between HA1 and HA2. Correlation analyses were used to examine the association between EMA- and retro-GHABP.

Results   For the EMA-GHABP, HA2 had significantly higher (better) scores than HA1 in the GHABP subscales of benefit, residual disability, and satisfaction ($p = 0.029$–0.0015). In contrast, the difference in the retro-GHABP score between HA1 and HA2 was significant only in the satisfaction subscale ($p = 0.0004$). The correlations between the EMA- and retro-GHABP were significant in all subscales ($p = 0.0004$ to <0.0001). The strength of the association ranged from weak to moderate ($r = 0.28$–0.58). Finally, the exit interview indicated that 29 participants (74.4%) preferred HA2 over HA1.

Conclusion   The study suggests that in-situ self-reports collected using EMA could have a higher sensitivity than retrospective questionnaires. Therefore, EMA is worth considering in clinical trials that aim to compare the outcomes of different hearing aid technologies. The weak to moderate association between in-situ and retrospective self-reports suggests that these two types of measures assess different aspects of hearing aid outcomes.

Keywords
► Outcome measures
► hearing aids
► ecological momentary assessment

Retrospective self-reports such as questionnaires have been widely used in audiological research to assess real-world outcomes of hearing aids. Although retrospective self-reports have consistently supported the benefit of hearing aids relative to unaided listening,[1,2] they rarely demonstrate the effect of hearing aid features in aided listening conditions. For example, retrospective questionnaires used in previous clinical trials have shown a small or lack of difference in real-world outcomes between wide dynamic range compression and linear processing,[3–5] between hearing aids with different numbers of channels,[3] between hearing aids with and without directional microphones[6–12] or noise reduction algorithms,[13,14] and between hearing aids with advanced and basic technologies.[1,15–17] The insensitivity of retrospective questionnaires in detecting the outcome difference between hearing aid technologies could be due to, at least in part, two reasons. First, retrospective questionnaires are subject to recall bias. Because retrospective questionnaires are typically administered several weeks or months after hearing aid fitting, respondents have to recall and summarize their listening experience across a long period of time. Long-term recall could be inaccurate and unreliable.[18] Second, retrospective questionnaires often suffer from poor contextual resolution. For example, the effect of modern hearing aid technologies, such as directional microphones, often depends on the characteristics of listening activities, situations, and environments (i.e., listening context). If the listening context described in a questionnaire is not sufficiently specific, the questionnaire will not be able to demonstrate the benefit of the technology.

Several techniques have been developed to overcome the disadvantages of retrospective self-reports. The ecological momentary assessment (EMA) is one of them. EMA is a methodology that asks respondents to repeatedly report their experiences during or shortly after the experiences in their natural environments (i.e., in-situ self-reports).[19] EMA provides a rich description of a sample of moments in respondents' lives, while avoiding the distortions that affect the delayed recall and evaluation of experiences. As a result, EMA is considered to be less affected by recall bias. Also, because detailed contextual information can be collected in each assessment, EMA has high contextual resolution. EMA has been implemented using paper-and-pencil journals,[20–23] daily diaries,[14] portable computers,[24] and smartphones[17,25] to assess listening difficulty or hearing aid outcomes in the real world.

Because EMA is less subject to recall bias and has higher contextual resolution, EMA could have a better ability to detect the effect of hearing aid technologies than retrospective questionnaires. However, to date there has been no empirical evidence to support this conjecture. Previous studies have used both EMA and retrospective questionnaires to measure hearing aid outcomes.[14,17] However, directly comparing the data of EMA and retrospective questionnaires collected in the previous studies is less meaningful, as in these studies EMA surveys and retrospective questionnaires often had different wordings and response formats. For example, Wu et al[17] compared the effectiveness of advanced hearing aids relative to basic-level hearing aids. Participants' satisfaction with hearing aids was measured using the retrospective questionnaire Satisfaction with Amplification in Daily Life (SADL)[26] and EMA. The SADL had 15 questions (e.g., "*Do you think your hearing aid(s) is worth the trouble?*") and the response was collected using a 7-point scale. In contrast, only one question was asked in the EMA survey to assess hearing aid satisfaction ("*Were you satisfied with your hearing aids?*") and the response was collected using a visual analog scale with two anchors from "not at all" to "very satisfied."

The first objective of the present study was to compare the ability of in-situ self-reports collected using the EMA methodology to detect the outcome difference between two hearing aid models relative to that of retrospective questionnaires. This ability is called *responsiveness* in some literature[27] and is referred to as *sensitivity* in the present study. The same wording and response format were used in the EMA survey and retrospective questionnaire to ensure that data could be directly compared. It was hypothesized that EMA would have a higher sensitivity than the retrospective questionnaire.

The second objective of the present study was to examine the association between in-situ and retrospective self-reports. Robinson and Clore[28] proposed the *accessibility model* to explain the discrepancy in self-reported data collected using different types of techniques. The accessibility model suggests that when online self-reports (such as EMA) that ask respondents to report their current experiences are used, respondents can directly introspect on experiential knowledge as it is generated. Experiential knowledge, however, can neither be stored nor retrieved after it is generated. Therefore, when short-term (hours to days) self-reports such as EMA or daily diaries are used, people tend to use episodic memory (the memory collection of past personal experiences that occurred at a particular time and place) to support reporting their experiences and related context. Although episodic memory can be biased by the most intense and recent events (peak and recency effects), it is relatively similar to experiential knowledge. Episodic memory, however, decays and rapidly becomes inaccessible. Thus, when long-term (weeks or months) self-reports such as retrospective questionnaires are used, respondents abandon the episodic recall strategy and start accessing relevant semantic memory to help guide their reports. While episodic memory is specific to an event from the past, semantic memory is not tied to any particular event but rather consists of certain beliefs or attitudes that are rarely revised. As a result, retrospective questionnaires could be biased by situation-specific beliefs (e.g., "vacations are enjoyable") or identity-related beliefs (e.g., "women are more empathetic than men"). Therefore, the accessibility model suggests that because self-reports collected using EMA and retrospective questionnaires involve different memories, these two types of measures likely assess different aspects of hearing aid outcomes. The former reflects what hearing aid users actually experience, while the latter reflects what users believe or remember.[29] According to the accessibility model, it was hypothesized that the association between hearing aid outcomes measured using EMA and retrospective questionnaires would be significant but small.
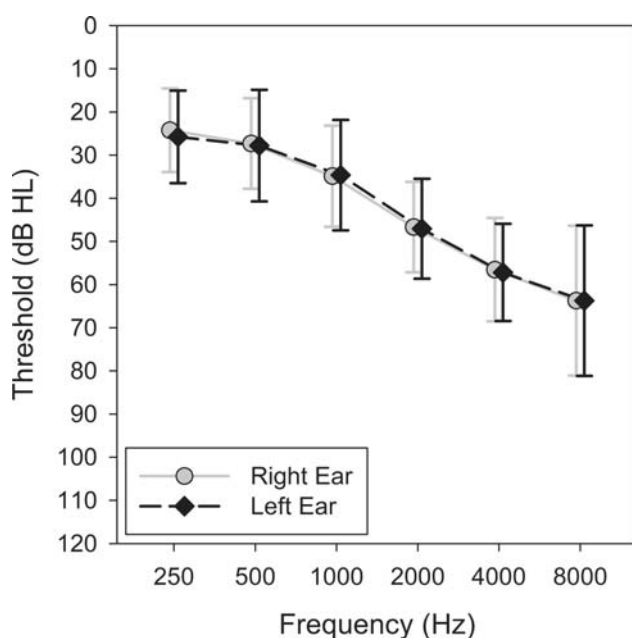
Methods

Overview

The present study was part of a larger clinical trial designed to compare the outcomes of a prototype hearing aid (denoted as HA1) and a commercially available device (HA2). The clinical trial was sponsored by the manufacturer of HA1. Older adults with hearing loss were recruited and fitted with bilateral hearing aids. A single-blinded, crossover repeated measures design was used. During the field trial of each hearing aid condition, participants wore the devices in their daily lives for 4 weeks. Hearing aid outcomes were measured using laboratory tests, retrospective questionnaires, and EMA. To answer the research questions posed in the present paper, the self-reported questionnaire and EMA data were compared.

It is of note that the protocol of the clinical trial required the order of the test condition (HA1 vs. HA2) to be counterbalanced across participants. However, because the delivery of HA1 was delayed by the manufacturer, most participants (37 out of 39) started with the HA2 field trial condition.

Participants

Thirty-nine participants (21 males and 18 females) were recruited from the community and completed the present study. The participants were recruited through the participant registry maintained by the researchers and word of mouth from the participants in the study. Their ages ranged from 48 to 83 years with a mean of 71 years. The participants were eligible for inclusion in the larger study if their hearing loss met the following criteria: (1) postlingual, bilateral, sensorineural type of hearing loss (air-bone gap < 10 dB); (2) pure-tone average across 0.5, 1, 2, and 4 kHz between 25 and 60 dB HL; and (3) hearing symmetry within 20 dB for all test frequencies. The mean pure-tone thresholds are shown in ►Fig. 1. All partic-

ipants were native English speakers. Upon entering the study, 17 participants had previous hearing aid experience for at least 1 year. Although the EMA was implemented using smartphones (see later sections for details), previous smartphone use was not part of the inclusion criteria. Previous smartphone experience was not required in the clinical trial because participants were thoroughly taught how to use the smartphone and completed a real-world smartphone practice before any study assessments were begun.

The subject number of the larger clinical trial was determined by a power analysis. To calculate the power, it was determined that a 5% score difference in self-reported outcome measures would be a clinically relevant difference. Based on the pilot data and literature, the standard deviation of paired difference in self-report score was estimated to be 10%. With these estimations, the larger clinical trial required 34 participants, assuming $\alpha = 0.05$ and $\beta = 0.2$.

Hearing Aids and Fitting

Participants were fitted with two behind-the-ear hearing aid models (HA1 and HA2). The hearing aids were receiver-in-the-canal style instruments and were coupled to the participants' ears using noncustomized domes. ►Table 1 compares the features of the two hearing aid models. The main differences between the two devices were that HA2 had fewer channels and did not have low-level expansion processing or an impulse noise reduction feature, and HA1 was not equipped with a feedback suppression feature. Both hearing aid models had smartphone applications (i.e., apps) that allowed users to use smartphones to change the volume, user program, and frequency response of the hearing aids. These apps are referred to as the hearing aid specific apps (in contrast to the EMA app described below). Only HA1's app could use the smartphone as

Fig. 1 Average audiograms for left and right ears of study participants. Error bars = 1 standard deviation.

Table 1 Differences, as described by the manufacturer, between the two hearing aid models

	HA1	HA2
Number of channels	16	9
Wide dynamic range compression	Yes	Yes
Expansion	Yes	No
Number of programs	4	4
Adaptive directional microphone	Yes	Yes
Gain-reduction noise reduction	Yes	Yes
Wind noise reduction	Yes	Yes
Impulse noise reduction	Yes	No
Frequency lowering	Yes	Yes
Feedback suppression	No	Yes
Smartphone app		
Volume control	Yes	Yes
Changing program	Yes	Yes
Equalizer (bass and treble)	Yes	Yes
Remote microphone	Yes	No

Abbreviation: App, application software.

a remote microphone. Both hearing aid models could have up to four user programs. Four programs recommended by the manufacturers were used in the study: Speech, Speech in Noise, Noise, and Music for HA1, and All Around, Restaurant, Outdoor, and Music for HA2.

Note that although HA1's app could use smartphone as a remote microphone, the participants were not explicitly encouraged to use this feature during the clinical trial. Also note that although HA1 had more channels and had expansion processing and an impulse noise reduction feature, the effect of channel number[3] and impulse noise reduction[17] is small and the evidence supporting the benefit of expansion is mixed.[30] In contrast, audible feedback could significantly reduce hearing aid satisfaction[31] and HA1 was not equipped with a feedback suppression feature. Therefore, although the manufacturer of HA1 wished the clinical trial to demonstrate that HA1 would outperform HA2, the opposite was hypothesized by the researchers.

It is also of note that although the hearing aids used in the study were same in style (receiver-in-the-canal), color, and size, they were not identical in the appearance of the device cases and receivers. The hearing aid specific apps also varied in the interface. HA1's app did not have any identifying information about the manufacturer or hearing aid model. HA2's app did show the name of the manufacturer but did not show the name of the model of the hearing aids. The case of HA2 had the name of the manufacturer and was removed. Therefore, participants were not completely blinded regarding the test condition. However, they were not aware of technology details of the two devices.

The hearing aids were fitted bilaterally. The fitting was conducted with the devices set to the default program (Speech or All Around). The devices were first programmed to meet real-ear aided response (REAR) targets (± 3 dB) specified by the second version of the National Acoustic Laboratory nonlinear prescriptive formula[32] and were verified using a probe-microphone hearing aid analyzer (Audioscan Verifit; Dorchester, Ontario, Canada) with a 65-dB SPL speech signal presented from 0-degree azimuth. Then the devices were fine-tuned based on participant preference. Whenever possible, the same gain fine-tuning adjustments were made to both HA1 and HA2. Noncustomized ear domes that were appropriate for the hearing loss were selected by audiologists for each device in each ear. The noncustom domes were chosen based on the ability to match the REAR targets, participant comfort, and reduction of feedback. HA1 was often fitted with a more occluding dome and less high-frequency gain than HA2 due to the absence of the feedback suppression feature in HA1. The frequency-lowering feature of both devices was disabled. All other features, including the volume control, remained active at default settings during the study.

Laboratory Test

To assess participants' aided speech recognition performance, the American-dialect version of the Four Alternative Auditory Feature test (AFAAF)[33] was used. The AFAAF was selected because the British-dialect version of the Four Alternative Auditory Feature test[34] is suggested to be a reliable and

sensitive way to gauge speech recognition performance.[35] It was expected to be sensitive to audibility difference between HA1 and HA2. The AFAAF is made up of 80 test sentences. Each sentence has an embedded key word: "Can you hear __ clearly?" "Can you hear STREAM clearly?" is an example of a test utterance. Four alternatives such as SCREAM, SCHEME, STREAM, and STEAM were presented to the participants on a computer screen. Participants would then click the word they thought they heard with a computer mouse to select their answer.

The AFAAF was administered in a low-reverberant sound field (reverberation time = 0.21 second) created using eight Tannoy (Coatbridge, Scotland) i5W loudspeakers. The loudspeakers were placed 1.2 m from the seated participant at 0, 45, 90, 135, 180, 225, 270, and 315-degree azimuths. Speech was presented from 0-degree azimuth and the level was fixed at 62 dBA. Uncorrelated AFAAF masking noise, which is a speech-shaped steady noise, was presented from all eight loudspeakers. The overall level of the noise was fixed at 57 dBA. Eighty sentences of the AFAAF were used in each hearing aid condition. Sentence order was randomized across participants. Performance was scored based on the percentage of words selected correctly.

Retrospective Self-Reports

Five standardized questionnaires were used in the larger clinical trial to measure hearing aid outcomes in various domains. The questionnaires were administered in a paper-and-pencil form.

Abbreviated Profile of Hearing Aid Benefit

The Abbreviated Profile of Hearing Aid Benefit (APHAB)[36] is a 24-item inventory designed to evaluate benefit experienced from hearing aid use and to quantify the degree of communication difficulty experienced in various situations. The questionnaire consists of four subscales. The ease of communication, background noise, and reverberation subscales are focused on speech communication and therefore the global score of the APHAB is the mean of the scores of these three subscales. The aversiveness (AV) subscale evaluates the individual's response to unpleasant environmental sounds. The global score (referred to as the APHAB-Global) and the AV subscale score (APHAB-AV) were used in data analysis.

Hearing Handicap Inventory for the Elderly or for the Adult

The Hearing Handicap Inventory for the Elderly or for the Adult (HHIE/A)[37,38] is a 25-item inventory designed to evaluate the social and emotional impact of hearing loss on an individual's life. The questionnaire is divided into two subscales, ranging from 12 to 13 items in length: the social subscale, which assesses the extent to which social aspects of an individual's life are impacted by hearing loss, and the emotional subscale, which measures how emotional responses in an individual's life are influenced by hearing loss. The user rates the degree of the impact with "Yes" (equal to 4 points), "Sometimes" (equal to 2 points), or "No" (equal to 0 points). Scores are added for each subscale. In the present study, the HHIE was used for participants older

than 65 years. The global score is the sum of the scores for all 25 items and was used in data analysis.

Satisfaction with Amplification in Daily Life

The SADL[26] is a 15-item inventory designed to evaluate an individual's satisfaction with his/her hearing aids. The questionnaire is divided into four subscales. The positive-effect subscale quantifies improved performance while using hearing aids, such as reduced communication disability. The personal-image subscale evaluates the domain of self-image and stigma. The negative features subscale assesses undesirable aspects of hearing aid use, such as feedback problems. The service and cost subscale measures the adequacy of service provided by the professional and the cost of the devices. The mean of the scores for all items (except for the item related to cost, as hearing aids were provided at no cost in the present study) forms the global score and was used in data analysis.

Speech, Spatial, and Qualities Hearing Scale

The 49-item Speech, Spatial, and Qualities hearing scale (SSQ)[39] is a validated questionnaire designed to measure a range of hearing disabilities across several domains. The SSQ consists of three subscales that measure the ability of an individual to understand speech, to localize acoustic events, and to evaluate auditory experience including music perception and the clarity and naturalness of sound. In the present study, the mean of the scores for all items was defined as the global score and was used in data analysis.

Glasgow Hearing Aid Benefit Profile

To answer the research questions of the present study, the aided portion of the Glasgow Hearing Aid Benefit Profile (GHABP)[40] was included in the larger clinical trial. The aided GHABP assesses four outcome domains (hearing aid use, hearing aid benefit, residual disability, and hearing aid satisfaction) in four predefined listening situations (TV listening, small conversation in quiet, conversation in noise, and group conversation). Each domain is assessed on a 5-point scale. To ensure that participants did not evaluate their own hearing aids in the GHABP, the wording "your hearing aid" in the original GHABP was replaced by "the study hearing aids." The four subscale scores were used in data analysis. Patient-nominated listening situations of the GHABP were not used in the present study. Because the GHABP described in this section was a retrospective self-report, it is referred to as the retro-GHABP.

In-Situ Self-reports

The EMA methodology was used to collect in-situ self-reports. EMA was implemented using Samsung (Seoul, South Korea) Galaxy S6 smartphones. A smartphone app was developed to deliver EMA surveys.[41] This app is referred to as the EMA app. During the last week of each hearing aid field trial (see the next section for details), the participants carried the study smartphones with them as they went about their daily lives. The EMA app prompted the participants to complete surveys at randomized intervals, approximately every 1.5 hours, within a participant's specified daily time window (e.g., between 8 a.m. and 9 p.m.). Participants answered the survey questions based on their listening experiences during the past 1.5 hours. If a participant knew that they would be unable to take a survey in the next half-hour, they were able to "snooze" the survey to ensure that no notifications would come for at least 30 minutes. If a participant missed a survey or it came at an inconvenient time, the survey was skipped as there was no way for participants to initiate their own surveys.

To ensure that the results of EMA and retrospective questionnaires could be directly compared, the wording and response format of the GHABP was used in EMA surveys. The exception was that the abbreviation "HAs" was used to represent the wording "hearing aids" in the question and responses of the benefit subscale item (to save space on the smartphone's screen so that a larger font could be used in the EMA app). At the start of an EMA survey, the app first asked if the GHABP's predefined listening situation (i.e., TV listening) happened during the past 1.5 hours. Participants tapped a button (Yes or No) on the smartphone screen to indicate their responses. If the response was No, the next predefined situation would be presented. If the response was Yes, the EMA app would present the four GHABP questions sequentially to assess outcomes in hearing aid use, benefit, residual disability, and satisfaction. The five answers from the GHABP were displayed on the smartphone screen as five response buttons and the participants selected the applicable response. Note that if participants indicated that they did not use hearing aids in the past 1.5 hours in the first GHABP question, the three questions about benefit, residual disability, and satisfaction would not be presented. The scoring was identical to the retro-GHABP. The GHABP described in this section is referred to as the EMA-GHABP.

Procedures

The study was approved by the Institutional Review Board at the University of Iowa. After signing the consent form, the participants' hearing thresholds were measured using pure-tone audiometry. If the participant met all the inclusion criteria, hearing aids were fitted. Next, demonstrations of how to work with and care for the smartphone were provided by the laboratory. Special attention was focused on taking EMA surveys on the phone. Participants were instructed to respond to the auditory/vibrotactile prompts to take surveys whenever it was possible and within reason (e.g., not while driving). Once all of the participants' questions had been answered and they demonstrated competence in the ability to perform all of the related tasks, they were sent home with one smartphone (with the EMA app installed) and a pair of hearing aids for the first trial condition (HA2 for most participants) and began a 7-day practice session. Each participant was given a set of take-home written instructions detailing how to use and care for the phone, as well as when and how to take EMA surveys.

Participants returned to the laboratory after the practice session. If participants misunderstood any of the EMA/smartphone-related tasks during the practice session, they

were reinstructed on how to properly use the equipment or take the surveys. If needed, hearing aid gain adjustments were provided based on participants' reports and preferences. The gain adjustment was conducted under the guidance of probe-microphone measures. The settings of the features were not adjusted. Before participants left the laboratory, the hearing aid-specific app (to change user programs, for example) of the first field trial condition was installed to the Samsung smartphone and the phone was paired to the hearing aids. Demonstrations of how to use the hearing aid-specific app were then provided. The EMA app on the smartphone was deactivated for the next 3 weeks.

Next, the first field trial condition began (HA2 for most participants). Participants used the hearing aids and the hearing aid-specific app for 3 weeks. Participants then returned to the laboratory and the EMA app was activated. A brief retraining on the EMA app was provided. The assessment week in which participants conducted EMA surveys then began. Participants were encouraged to go about their normal routines during the week. One week later, participants returned to the laboratory and the AFAAF and retrospective questionnaires were administered. While completing the AFAAF, participants wore hearing aids with their typical volume and setting (i.e., as-worn measures). When answering the retrospective questionnaires, participants were asked to recall their experiences during the past 4 weeks. Participants were then interviewed face-to-face about their experience with the study hearing aids. Open-ended questions regarding their likes and dislikes about the devices and whether they would purchase them were asked of the participants.

After finishing the first trial condition, the second condition immediately followed. The general procedure for the second condition was identical to the first condition. First, the gain frequency responses of the hearing aids of the second condition were adjusted to match the gain frequency responses of the devices used in the first condition under the guidance of probe-microphone measures. Next, the hearing aid-specific app for the second trial condition was installed to the phone and hearing aids were paired. The EMA app was deactivated. Participants then left the laboratory and began the second field trial (HA1 for most participants). In the first week of the second trial condition, participants could request additional hearing aid gain adjustment, although this adjustment was not encouraged (to ensure that HA1 and HA2 had similar gain frequency responses). Additional gain modifications were made for four subjects who reported excessive feedback from HA1. Three weeks later, participants returned to the laboratory and the EMA app was activated. The assessment week in which participants conducted EMA surveys then began. One week later, participants returned to the laboratory and the AFAAF and retrospective questionnaires were administered.

After participants completed the second trial condition and returned to the laboratory, a second interview was conducted. In addition, participants were asked to indicate their overall hearing aid preference (HA1, HA2, or no preference) and the reason for their preference. Monetary compensation was provided to the participants upon completion of the study.

Results

Because 37 out of 39 participants started with the HA2 field trial condition, the test order could have biased the results of the clinical trial. Literature has shown that people tend to report better outcomes for the devices they experienced more recently.[42,43] Because only two participants started with HA1, it makes less sense to control the effect of test order in statistical analysis. Instead, to shed light on how test order could affect the outcome difference between the two devices, the effect of hearing aid experience was controlled in the statistical model. This decision was made based on the study by Naylor et al[43] that compared the outcomes of two identical hearing aids. Naylor et al found that most (81%) first-time hearing aid users preferred the second device, while this order effect was not observed in experienced users. In the present study, 22 and 17 participants were first-time users and experienced users, respectively.

Laboratory Test

The mean AFAAF score of each hearing aid condition averaged across all participants is shown in ►**Fig. 2A**. Higher scores represent better performance. A linear mixed model with random intercept for subject was created to examine the effect of hearing aid model (HA1 vs. HA2), hearing aid experience (first time vs. experienced), and their interaction on the AFAAF score. Results indicated that the AFAAF score of HA2 was significantly higher (better) than HA1 ($p = 0.0053$). The interaction between hearing aid model and experience was not significant ($p = 0.86$). See Appendix A1 in the Supplementary Material (online only) for detailed statistics.

Retrospective Questionnaires

►**Fig. 2B** shows the mean score of retrospective questionnaires (excluding the retro-GHABP) of each hearing aid condition
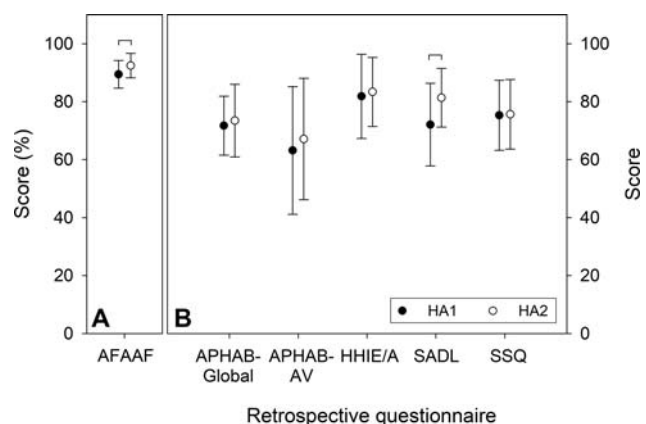
Fig. 2 (A) Mean score of the American-dialect version of the Four Alternative Auditory Feature test (AFAAF) of each hearing aid condition (HA1 and HA2). (B) Mean outcome scores of retrospective questionnaires of each hearing aid condition. Higher scores represent better outcomes. Brackets represent significant difference. Error bars = 1 SD. APHAB, Abbreviated Profile of Hearing Aid Benefit; AV, aversiveness; HHIE/A, Hearing Handicap Inventory for the Elderly or for the Adult; SADL, Satisfaction with Amplification in Daily Life; SSQ, Speech, Spatial, and Qualities hearing scale; SD, standard deviation.

averaged across all participants. All scores shown in the figure have been linearly transformed so that the score ranges from 0 to 100, with higher scores representing better outcomes. Separate linear mixed models with a random intercept for subject were created for each measure to determine the effect of hearing aid (HA1 versus HA2), hearing aid experience (first time vs. experienced), and their interaction on the measure. Results from the models indicated that the SADL score of HA2 was significantly higher (better) than HA1 ($p = 0.017$). No other significant main effect of hearing aid was found. None of the interactions was significant. See Appendix A1 for detailed statistics.

EMA versus Retro-GHABP: Effect of Hearing Aid and Measure

Across the two hearing aid conditions, a total of 3,200 EMA-GHABP surveys were completed by the 39 participants (HA1: 1557, HA2: 1643). On average each participant completed 5.9 surveys per day. Across all surveys, the frequency of occurrence of each GHABP-predefined listening situation (TV listening, small conversation in quiet, conversation in noise, and group conversation) was 23.2, 47.7, 15.0, and 16.1%, respectively. In 851 (26.6%) surveys, none of the four predefined listening situations occurred. Because these surveys did not contain any outcome data, they were excluded from analysis. The remaining 2,349 EMA-GHABP surveys were used in analysis (HA1: 1,087, HA2: 1,262). It is of note that because EMA involves repeated sampling, there were more data points for the EMA-GHABP (on average 30.1 assess-

ments per participant per hearing aid condition) than the retro-GHABP (one assessment per participant per hearing aid condition). To ensure that the results of the EMA- and retro-GHABP could be directly compared and could be analyzed in the same statistical model, the EMA-GHABP scores of individual surveys completed by a participant in a hearing aid condition were averaged. For both the EMA- and retro-GHABP, the scores were further averaged across the four predefined listening situations, each subscale separately. The GHABP data were not examined in each listening situation because (1) the numbers of individual EMA surveys completed in some situations were quite low (e.g., conversation in noise: 4.5 assessments per participant per condition) and (2) EMA relies on a large amount of data from each respondent to derive a clear pattern of human experiences.[19] The averaged EMA-GHABP and retro-GHABP scores (ranging from 1 to 5) were then used in data analysis.

►**Fig. 3A** shows the mean subscale score of the EMA- and retro-GHABP of each hearing aid condition averaged across all participants. For all scores shown in the figure, higher scores presented better outcomes. To compare the ability of the EMA- and retro-GHABP to detect the outcome difference between HA1 and HA2, a linear mixed model with a random intercept for subject was created. The independent variables were a variable termed "group" (four levels, which are the factorial combination of two hearing aid conditions and two measures: retrospective/HA1, retrospective/HA2, EMA/HA1, EMA/HA2), hearing aid experience (first time vs. experienced), and their interaction. The dependent variable was the GHABP subscale score. The models were created for each
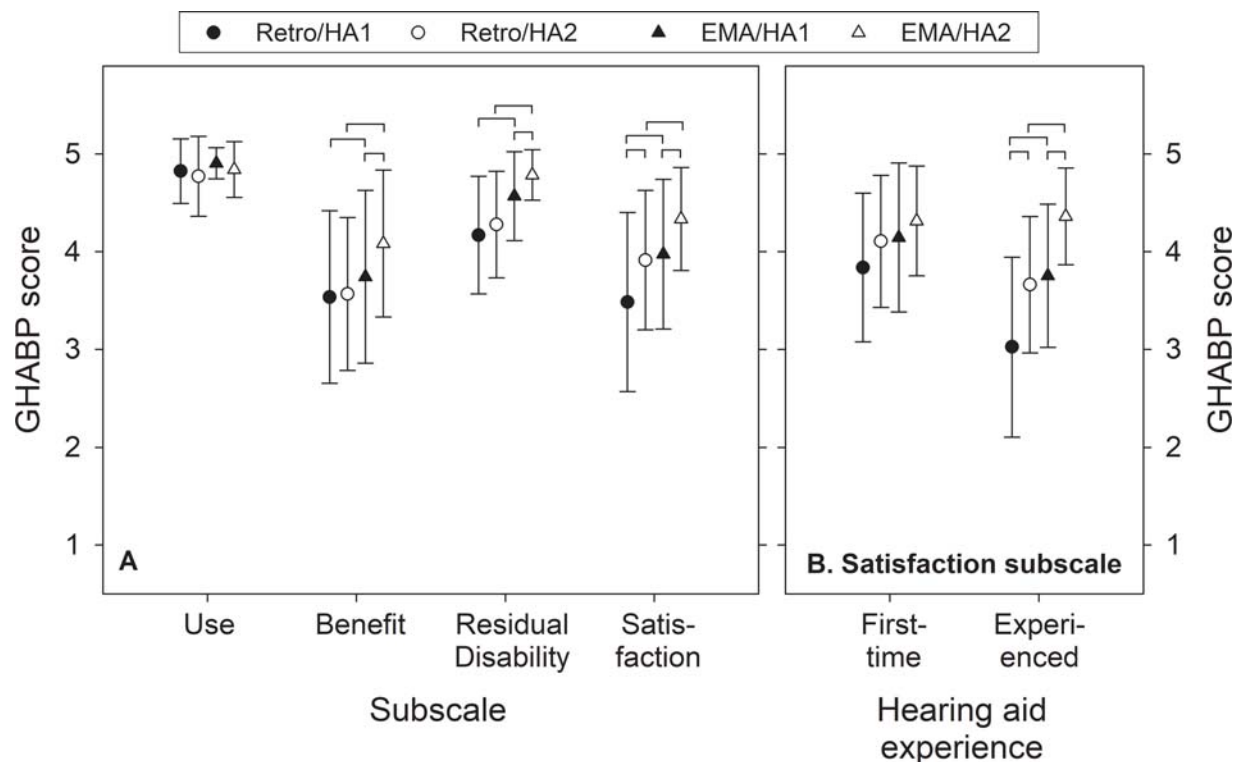
Fig. 3 Mean subscale score of the EMA- and retro-GHABP of each hearing aid condition (HA1 and HA2). Higher scores represent better outcomes. Brackets represent significant difference. Error bars = 1 SD. EMA, ecological momentary assessment; GHABP, Glasgow Hearing Aid Benefit Profile; Retro, retrospective; SD, standard deviation.

GHABP subscale (excluding the use subscale). For the use subscale, the scores were close to ceiling and were not normally distributed. Approximately 55 and 72% of the participants reported using hearing aids "all the time" (a score of 5) in the EMA- and retro-GHABP use subscale, respectively. Therefore, the use variable was converted to a categorical variable that had two categories: "all the time" for a score of 5 and "not all the time" for a score of lower than 5. A generalized linear mixed model with a logit link and random intercept for subject was then created. The dependent variable was the categorical use subscale. The independent variables were group, hearing aid experience, and their interaction. The results indicated that the effect of group was significant ($p < 0.0001$) in all GHABP subscales, except for the use subscale ($p = 0.079$). The interaction between group and hearing aid experience interaction was significant in the satisfaction subscale ($p = 0.0004$) but not in other subscales.

The statistics of the created models were then used to conduct pairwise comparisons of the four levels of the group variable (retrospective/HA1, retrospective/HA2, EMA/HA1, EMA/HA2). To adjust for the evaluation of multiple comparisons, the p-values were adjusted using the false discover rate.[44] This adjustment method was selected because it controls the proportion of significant results that are incorrect while preserving the power of statistical models.[45] Results first indicated that, for the use subscale, none of the pairwise comparisons were significant. The models then revealed that the EMA-GHABP scores of HA2 were significantly higher (better) than the EMA-GHABP scores of HA1 in the subscales of benefit, residual disability, and satisfaction (indicated by narrow brackets in ►**Fig. 3A**, adjusted $p = 0.029–0.0015$), suggesting that HA2 yielded better outcomes than HA1. In contrast, the difference in the retro-GHABP score between HA1 and HA2 was significant only in the satisfaction subscale (adjusted $p = 0.0004$). Unexpectedly, the models further indicated that, for a given hearing aid condition, the EMA-GHABP score was significantly higher than the retro-GHABP score (indicated by wide brackets in ►**Fig. 3A**, adjusted $p = 0.0049$ to < 0.0001) in all subscales. The exceptions were the use subscale (adjusted $p = 0.74$ and 0.99). The results of the pairwise comparisons across hearing aid model and measure (e.g., retro-GHABP of HA1 vs. EMA-GHABP of HA2) were not of interest to the present study and are not shown in ►**Fig. 3A**.

Because the interaction between group and hearing aid experience was significant in the satisfaction subscale, pairwise comparisons of the four levels of the group variable were conducted between first-time and experienced users separately using the statistics of the created mixed linear model. ►**Fig. 3B** shows the mean score of the satisfaction subscale of these two types of participants. Results from the statistical models indicated that experienced users reported more satisfaction with HA2 than HA1 in both the EMA- and retro-GHABP ($p = 0.0046$ and 0.0011, respectively). However, the difference in the EMA- and retro-GHABP between the two devices was not significant for first-time users ($p = 0.36$ and 0.17, respectively). See Appendix A1 for detailed statistics.

EMA versus Retro-GHABP: Association

►**Fig. 4A–D** shows the scatter plots of the EMA-GHABP score as a function of the retro-GHABP score in the four subscales. The symbols in the figure are slightly shifted on the x-axis so that they do not overlap. It is of note that most symbols shown in ►**Fig. 4A–D** are above the diagonal line that represents a perfect match between the EMA- and retro-GHABP scores. To examine the association between the EMA- and retro-GHABP while controlling for the effect of hearing aid model (HA1 and HA2), partial correlations were used. Because the use and residual disability subscale data were right censored, Kendall rank correlations were used for all variables. All correlations were found to be significant ($p = 0.004$ to < 0.0001) and the results are shown in ►**Fig. 4**. The strength of the association ranged from weak to moderate ($r = 0.28–0.58$).

To shed light on how hearing aid experience affected the association between the EMA- and retro-GHABP, partial correlations were conducted on first-time and experienced users separately in each subscale. All correlations remained significant, except for the use ($r = -0.0059$, $p = 0.96$) and residual disability ($r = 0.19$, $p = 0.12$) subscales of experienced users (►**Fig. 4E, F**). See Appendix A1 for detailed statistics.

Overall Preference

Participants were asked to report their overall preference in the exit interview. Among the 39 participants, eight of them preferred HA1 and 29 preferred HA2. Two participants had no preference (►**Fig. 5**). A binomial test was conducted on all participants to determine if one of the hearing aid models was more likely to be preferred. The two participants who had no preference were excluded from the analysis. The result indicated that participants were more likely to prefer HA2 ($p = 0.0008$). ►**Fig. 5** also shows the hearing aid preference of first-time and experienced users. HA2 was evenly preferred by both groups of users, while HA1 was mainly preferred by first-time users. The binomial test was then conducted for first-time and experienced users separately. Results suggested that the likelihood for first-time users to prefer HA1 or HA2 did not significantly differ from the chance level ($p = 0.13$), while experienced users were more likely to prefer HA2 ($p = 0.001$). See Appendix A1 for detailed statistics.

Participants were also invited to indicate the reason for their preference. For participants who preferred HA1, this device had better sound clarity and more comfortable fit. For participants who preferred HA2, this device had better sound quality, was less likely to generate acoustic feedback, stayed connected to the app/phone better, and had longer battery life.

Discussion

Outcomes: HA1 versus HA2

For the larger clinical trial designed to compare the outcomes of HA1 and HA2, the biggest flaw is that the test order of the two devices was not counterbalanced. Among the 39 participants, only two started with HA1 (one first-time user and one experienced user). As mentioned, an order effect could bias hearing aid outcomes such that research participants,
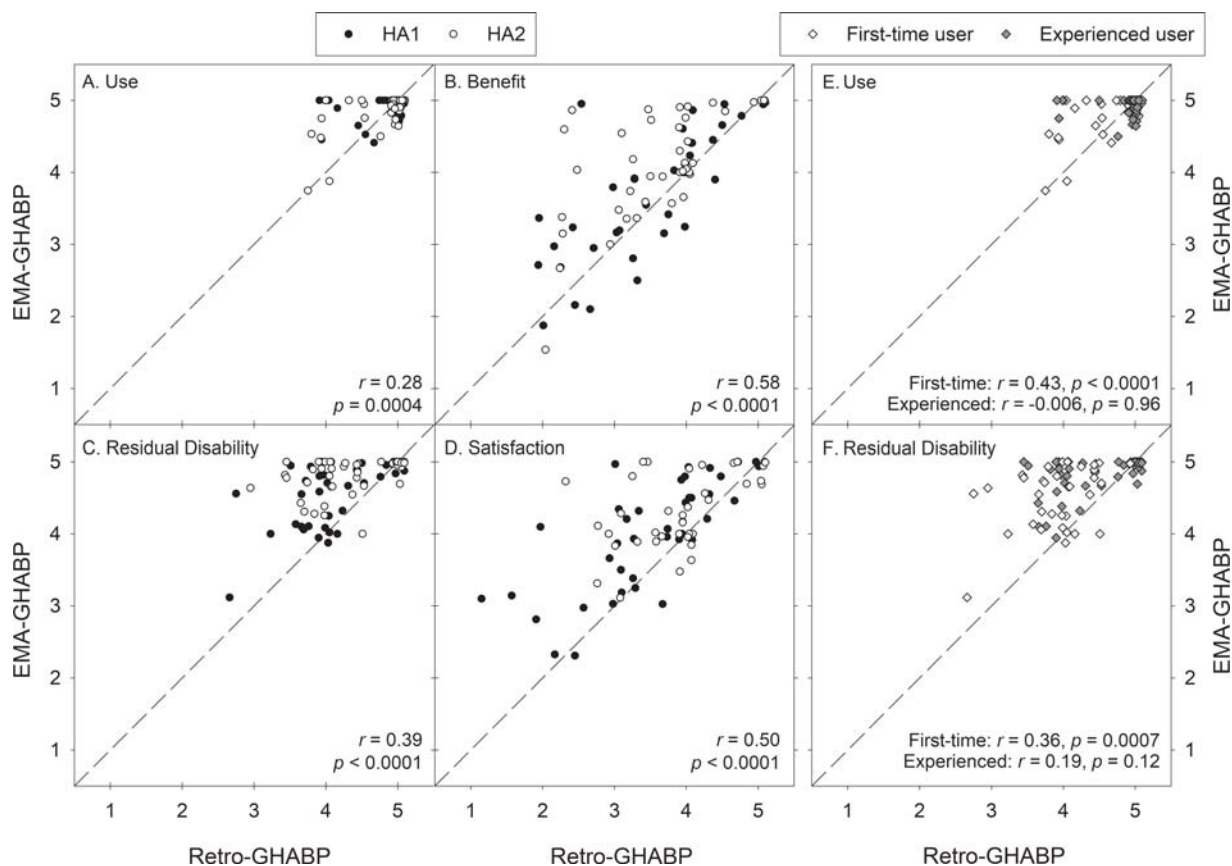
Fig. 4 Scatter plots of the EMA-GHABP score as a function of the retro-GHABP score for the two hearing aid conditions (HA1 and HA2). Symbols are slightly shifted on the x-axis so that they do not overlap. Diagonal lines represent perfect match between the EMA- and retro-GHABP scores. EMA, ecological momentary assessment; GHABP, Glasgow Hearing Aid Benefit Profile; Retro, retrospective; SD, standard deviation.
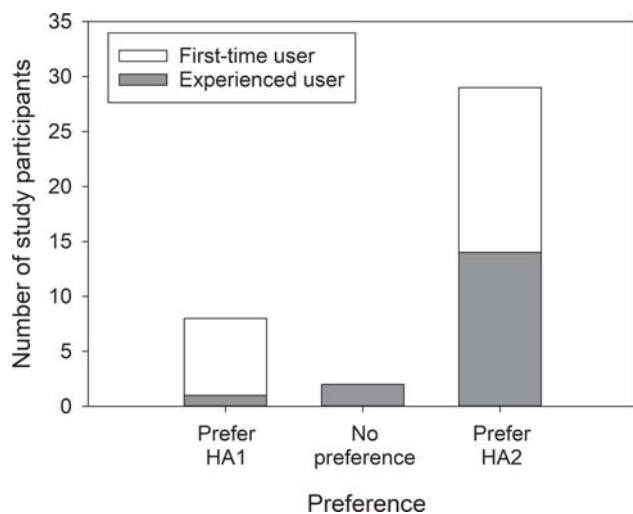
Fig. 5 Overall hearing aid preference of study participants.

especially first-time users, tend to prefer the devices experienced more recently.[42,43] This order effect could be observed in hearing aid overall preference data shown in ►**Fig. 5**. Specifically, only 1 out of 17 experienced users preferred the device of the second trial condition (i.e., HA1), while 7 out of 22 first-time users preferred HA1 in the second condition. Therefore, the order effect could have biased the results of the larger clinical trial toward the direction of favoring HA1.

Despite this bias, all outcome measures used in the study trended toward HA2 outperforming HA1 except for the use subscale of the GHABP (►**Figs. 2** and **3**). The trend of HA2 outperforming HA1 was also consistent with the literature regarding the potential impact of feedback suppression (which HA2 had) and minimal effects of higher channel numbers, expansion processing, and impulse noise reduction (which HA1 had) mentioned earlier. Therefore, it seems reasonable to conclude that HA2 yielded better outcomes than HA1 (likely across multiple outcome domains) and that the degree of outcome difference was underestimated in the larger clinical trial, especially for first-time users. The underestimation in first-time users is consistent with the GHABP satisfaction subscale results showing that experienced users reported more satisfaction with HA2 than HA1, while the difference between the two devices was not significant in first-time users (►**Fig. 3B**).

Retrospective Questionnaires
If HA2 outperforming HA1 across multiple outcome domains is regarded as the ground truth, then none of the retrospective questionnaires, except for the SADL and the satisfaction subscale of the retro-GHABP, were able to detect the outcome difference between HA1 and HA2 (►**Figs. 2B** and **3**). This finding is consistent with literature showing the insensitivity of retrospective questionnaires on detecting the effect of hearing aid technologies in aided listening conditions. It is of

note that the two retrospective self-reports that were able to detect the outcome difference between HA1 and HA2 (i.e., the SADL and the satisfaction subscale of the retro-GHABP) were designed to assess outcomes in the satisfaction domain. This could result from the main reasons for participants disliking HA1 being more in line with the questions asked in satisfaction measures. For example, many participants reported that they disliked HA1 because this device had shorter battery life and poorer smartphone connectivity. These dislikes could be better reflected by questions that assess hearing aid satisfaction, such as "*Do you think your hearing aids are worth the trouble?*" in the SADL and "*How satisfied were you with your hearing aids?*" in the retro-GHABP. In contrast, although participants had better speech recognition performance with HA2 than HA1 in the laboratory (likely resulted from HA2's better audibility brought by its feedback suppression feature) (►**Fig. 2A**), the difference could be too small to be detected by retrospective questionnaires that assessed speech communication in the real world (i.e., the APHAB-Global and SSQ).

EMA- versus Retro-GHABP: Effect of Hearing Aid

Except for the use subscale that had scores at the ceiling level, all three subscales of both the EMA- and retro-GHABP trended toward HA2 outperforming HA1 (►**Fig. 3**). This trend was statistically significant in all three subscales for the EMA-GHABP, while the retro-GHABP only revealed a significant effect of hearing aid model in the satisfaction subscale. To minimize differences between the in-situ and retrospective measures in the present study, the EMA- and retro-GHABP questionnaires used the same wording and response format. Although the EMA- and retro-GHABP were administrated using different platforms (smartphone and paper, respectively), Gwaltney et al[46] suggest that self-report data collected using electronic and paper-and-pencil administration are equivalent. Putting all this information together, the results of the present study suggested that in-situ self-reports collected using the EMA methodology could be more sensitive than retrospective questionnaires in detecting the outcome differences between different aided listening conditions. The higher sensitivity of the EMA-GHABP is likely due to the in-situ and repeated sampling nature of the EMA methodology. Because the participants reported on their experience during the past 1.5 hours in the EMA surveys, the experiences tied to specific listening situations were more accurately recalled. Because the surveys were repeated multiple times during the assessment week, the random variation in the factors that could affect participants reporting their experience (e.g., participant's mood in the moment) would be minimized when multiple survey data were aggregated. In contrast, because the participants only had one opportunity to report their experience in the retro-GHABP in each hearing aid condition, no data aggregation was available to reduce bias.

EMA- versus Retro-GHABP: Association

The correlation between the EMA- and retro-GHABP was significant in all four subscales when the data from all participants were pooled (►**Fig. 4A–D**). When the analysis was conducted for first-time and experienced users separately,

the correlation was not significant for experienced users in the use and residual disability subscales (►**Fig. 4E, F**). The lack of association likely resulted from the smaller data range on the EMA-GHABP for experienced users.

Despite the significant correlation, the score difference between the EMA- and retro-GHABP varied considerably across participants. For example, although many participants reported similar hearing aid satisfaction in the EMA- and retro-GHABP (symbols on the diagonal line in ►**Fig. 4D**), several participants reported that they were "reasonably satisfied" (score = 3) by the hearing aids in the retro-GHABP while they reported that they were "delighted with aids" (score = 5) in EMA surveys. The weak to moderate correlations shown in ►**Fig. 4** are consistent with the accessibility model,[28] which states that self-reports collected using EMA and retrospective questionnaires involve different types of memories (i.e., episodic and semantic memories, respectively). Therefore, the EMA- and retro-GHABP likely assessed different aspects of hearing aid outcomes. The former reflected what participants actually experienced, while the latter reflected what participants believed or remembered. However, note that the weak to moderate correlations could also be due to the difference in the assessment time window of the two measures. The retro-GHABP was administered 4 weeks postfit and reflected the overall outcome across the 4 weeks, while the EMA-GHABP was administered in and assessed the fourth week of the trial.

EMA- versus Retro-GHABP: Effect of Measure

The difference in assessment time window could also explain why the scores of the EMA-GHABP were consistently higher (better) than the retro-GHABP—an unexpected finding of the present study. Humes et al[47] used several retrospective self-reports to measure hearing aid outcomes at 7, 15, 30, 60, 90, and 180 days postfit. The results indicated that the effect of postfit interval was significant on the score of a satisfaction survey, with a lower score (lower satisfaction) at 7 days postfit compared with the scores at the rest of intervals. In the present study, the retro-GHABP reflected the overall outcomes across the 4 weeks of the trial, which included the first week postfit in which hearing aid satisfaction could be lower. Therefore, participants could report a lower satisfaction level in the retro-GHABP compared with the EMA-GHABP that measured the outcomes of the fourth week in the trial. However, Humes et al[47] did not find a significant effect of postfit interval on self-reported hearing aid usage, hearing aid benefit measured using the questionnaire Hearing Aid Performance Inventory,[48] and residual disability measured using the HHIE. Thus, the poorer initial outcome that occurred right after hearing aid fitting could not completely explain why in the present study the retro-GHABP had lower (poorer) scores in all subscales (except for the use subscale) and a lower global score than the EMA-GHABP.

Another explanation for the discrepancy between the EMA- and retro-GHABP involves the accessibility model.[28] As described in the introduction, respondents use semantic memory to help guide their long-term retrospective self-reports. Because semantic memory is often shaped according to criteria such as coherence with personal beliefs or attitude, people preferentially remember information that supports their

coherent beliefs or attitude in long-term retrospective self-reports. Therefore, the lower scores of the retro-GHABP in the present study could be due to participants' stigma or negative attitude toward hearing aids. If this is the case, participants might obtain more benefits from hearing aids in real time than what they believed or remembered in retrospective questionnaires. However, because hearing aid owners (i.e., experienced users) are less likely to have a negative attitude toward hearing aids compared with nonowners (first-time users in the present study),[49] the data showing that the retro-GHABP scores of experienced users tend to be lower (poorer) than that of first-time users (solid symbols in ►**Fig. 3B**) do not support this explanation. More research to clarify this hypothesis is suggested.

Limitations

In addition to the execution flaw of the larger clinical trial, the present study has limitations that concern the generalizability of the study results. First, because the EMA- and retro-GHABP were administered in the way that is typically used in clinical trials, the two measures had different assessment time windows (1 vs. 4 weeks). If the outcome is assessed in the same time window (e.g., the retro-GHABP asking participants to report the experience of the past week), the association between the EMA- and retro-GHABP could be stronger and the score discrepancy between them could disappear. Second, although the present study suggested that the EMA-GHABP was more sensitive than the retro-GHABP, it does not guarantee that all in-situ self-reports collected using the EMA methodology will have high sensitivities. For example, because it is impossible to strictly control real-world environments, EMA relies on a large amount of data from each respondent to derive a clear pattern of human experiences and behaviors. In the present study each participant on average had 30.1 EMA surveys in each hearing aid condition for analysis. The findings of the present study may not generalize to the clinical trials in which much fewer EMA surveys are completed by study participants due to, for example, low motivation.

Conclusion

The present study suggests that in-situ self-reports collected using the EMA methodology could have a higher sensitivity than retrospective questionnaires in detecting outcome differences of hearing aids. Therefore, EMA is worth considering in clinical trials that aim to compare the outcomes across different aided listening conditions. The association between in-situ and retrospective self-reports was found to be weak to moderate, suggesting that these two types of measures assess different aspects of hearing aid outcomes. The former likely reflects what hearing aid users actually experience, while the latter reflects what users believe or remember.

Note

Portions of this paper were presented at the Academy Research Conference of the American Academy of Audiology, March 27, 2019, Columbus, OH.

Conflicts of Interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

References

1 Cox RM, Johnson JA, Xu J. Impact of advanced hearing aid technology on speech understanding for older listeners with mild to moderate, adult-onset, sensorineural hearing loss. Gerontology 2014;60(06):557–568

2 Humes LE, Rogers SE, Quigley TM, Main AK, Kinney DL, Herring C. The effects of service-delivery model and purchase price on hearing-aid outcomes in older adults: a randomized double-blind placebo-controlled clinical trial. Am J Audiol 2017;26(01): 53–79

3 Humes LE, Ahlstrom JB, Bratt GW, Peek BF. Studies of hearing aid outcome measures in older adults: a comparison of technologies and an examination of individual differences. Semin Hear 2009; 30:112–128

4 Johnson JA, Cox RM, Alexander GC. Development of APHAB norms for WDRC hearing aids and comparisons with original norms. Ear Hear 2010;31(01):47–55

5 Larson VD, Williams DW, Henderson WG, et al; NIDCD/VA Hearing Aid Clinical Trial Group. Efficacy of 3 commonly used hearing aid circuits: a crossover trial. JAMA 2000;284(14): 1806–1813

6 Walden BE, Surr RK, Cord MT, Edwards B, Olson L. Comparison of benefits provided by different hearing aid technologies. J Am Acad Audiol 2000;11(10):540–560

7 Cord MT, Surr RK, Walden BE, Olson L. Performance of directional microphone hearing aids in everyday life. J Am Acad Audiol 2002; 13(06):295–307

8 Cord MT, Surr RK, Walden BE, Dyrlund O. Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids. J Am Acad Audiol 2004;15(05):353–364

9 Surr RK, Walden BE, Cord MT, Olson L. Influence of environmental factors on hearing aid microphone preference. J Am Acad Audiol 2002;13(06):308–322

10 Palmer CV, Bentler R, Mueller HG. Amplification with digital noise reduction and the perception of annoying and aversive sounds. Trends Amplif 2006;10(02):95–104

11 Gnewikow D, Ricketts T, Bratt GW, Mutchler LC. Real-world benefit from directional microphone hearing aids. J Rehabil Res Dev 2009;46(05):603–618

12 Humes LE, Ahlstrom JB, Bratt GW, Peek BF. Studies of hearing-aid outcome measures in older adults: a comparison of technologies and an examination of individual differences. Semin Hear 2009; 30:112–128

13 Bentler RA. Effectiveness of directional microphones and noise reduction schemes in hearing aids: a systematic review of the evidence. J Am Acad Audiol 2005;16(07):473–484

14 Bentler R, Wu YH, Kettel J, Hurtig R. Digital noise reduction: outcomes from laboratory and field studies. Int J Audiol 2008;47 (08):447–460

15 Johnson JA, Xu J, Cox RM. Impact of hearing aid technology on outcomes in daily life II: speech understanding and listening effort. Ear Hear 2016;37(05):529–540

16 Johnson JA, Xu J, Cox RM. Impact of hearing aid technology on outcomes in daily life III: localization. Ear Hear 2017;38(06):746–759

17 Wu YH, Stangl E, Chipara O, Hasan SS, DeVries S, Oleson J. Efficacy and effectiveness of advanced hearing aid directional and noise reduction technologies for older adults with mild to moderate hearing loss. Ear Hear 2019;40(04):805–822

18 Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: the impact of memory and inference on surveys. Science 1987;236(4798):157–161

19 Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1–32

20 Preminger JE, Cunningham DR. Case-study analysis of various field study measures. J Am Acad Audiol 2003;14(01):39–55

21 Walden BE, Surr RK, Cord MT, Dyrlund O. Predicting hearing aid microphone preference in everyday listening. J Am Acad Audiol 2004;15(05):365–396

22 Wu YH, Bentler RA. Impact of visual cues on directional benefit and preference: Part II–field tests. Ear Hear 2010;31(01):35–46

23 Wu YH, Bentler RA. Do older adults have social lifestyles that place fewer demands on hearing? J Am Acad Audiol 2012;23(09):697–711

24 Galvez G, Turbin MB, Thielman EJ, Istvan JA, Andrews JA, Henry JA. Feasibility of ecological momentary assessment of hearing difficulties encountered by hearing aid users. Ear Hear 2012;33(04):497–507

25 Timmer BHB, Hickson L, Launer S. Do hearing aids address real-world hearing difficulties for adults with mild hearing impairment? Results from a pilot study using ecological momentary assessment. Trends Hear 2018;22:2331216518783608

26 Cox RM, Alexander GC. Measuring Satisfaction with Amplification in Daily Life: the SADL scale. Ear Hear 1999;20(04):306–320

27 Wright JG, Young NL. A comparison of different indices of responsiveness. J Clin Epidemiol 1997;50(03):239–246

28 Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. Psychol Bull 2002;128(06):934–960

29 Conner TS, Barrett LF. Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. Psychosom Med 2012;74(04):327–337

30 Plyler PN, Hill AB, Trine TD. The effects of expansion on the objective and subjective performance of hearing instrument users. J Am Acad Audiol 2005;16(02):101–113

31 McCormack A, Fortnum H. Why do people fitted with hearing aids not wear them? Int J Audiol 2013;52(05):360–368

32 Keidser G, Dillon H, Flax M, Ching T, Brewer S. The NAL-NL2 prescription procedure. Audiology Res 2011;1(01):e24

33 Xu J, Cox RM. Recording and evaluation of an American dialect version of the Four Alternative Auditory Feature test. J Am Acad Audiol 2014;25(08):737–745

34 Foster JR, Haggard MP. The Four Alternative Auditory Feature Test (FAAF)–linguistic and psychometric properties of the material with normative data in noise. Br J Audiol 1987;21(03):165–174

35 Gatehouse S, Naylor G, Elberling C. Linear and nonlinear hearing aid fittings–1. Patterns of benefit. Int J Audiol 2006;45(03):130–152

36 Cox RM, Alexander GC. The abbreviated profile of hearing aid benefit. Ear Hear 1995;16(02):176–186

37 Newman CW, Weinstein BE, Jacobson GP, Hug GA. The hearing handicap inventory for adults: psychometric adequacy and audiometric correlates. Ear Hear 1990;11(06):430–433

38 Ventry IM, Weinstein BE. The Hearing Handicap Inventory for the Elderly: a new tool. Ear Hear 1982;3(03):128–134

39 Gatehouse S, Noble W. The speech, spatial and qualities of hearing scale (SSQ). Int J Audiol 2004;43(02):85–99

40 Gatehouse S. Glasgow hearing aid benefit profile: derivation and validation of a client-centered outcome measure for hearing aid services. J Am Acad Audiol 1999;10:103

41 Hasan SS, Lai F, Chipara O, Wu YH. AudioSense: Enabling real-time evaluation of hearing aid technology in-situ. Paper presented at: The 26th IEEE International Symposium on Computer-Based Medical Systems; Porto, Portugal; June 20–June 22, 2013

42 Cox RM, Johnson JA, Xu J. Impact of hearing aid technology on outcomes in daily life I: the patients' perspective. Ear Hear 2016;37(04):e224–e237

43 Naylor G, Öberg M, Wänström G, Lunner T. Exploring the effects of the narrative embodied in the hearing aid fitting process on treatment outcomes. Ear Hear 2015;36(05):517–526

44 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 1995;57:289–300

45 Oleson JJ, Brown GD, McCreery R. Essential statistical concepts for research in speech, language, and hearing sciences. J Speech Lang Hear Res 2019;62(03):489–497

46 Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(02):322–333

47 Humes LE, Halling D, Coughlin M. Reliability and stability of various hearing-aid outcome measures in a group of elderly hearing-aid wearers. J Speech Hear Res 1996;39(05):923–935

48 Schum DJ. Test-retest reliability of a shortened version of the hearing aid performance inventory. J Am Acad Audiol 1993;4(01):18–21

49 Meyer C, Hickson L, Lovelock K, Lampert M, Khan A. An investigation of factors that influence help-seeking for hearing impairment in older adults. Int J Audiol 2014;53(Suppl 1):S3–S17