The Baltimore HEARS Study: Hearing Equity Through Accessible Research & Solutions

Statistical Analysis Plan (SAP) version 1.2 – 8/13/2021

ClinicalTrials.gov Identifier: NCT03442296

Protocol version 6.0 – 7/9/2021

Principal Investigator: Frank Lin, M.D., Ph.D.

IRB Application Number: IRB00144968

Section 1: Administrative Information

Baltimore HEARS: Hearing Equality Through Accessible Research & Solutions
ClinicalTrials.gov Identifier: NCT03442296
Protocol version 6.0 – 7/9/2021
Principal Investigator: Frank Lin, M.D., Ph.D.
IRB Application Number: IRB00144968

SAP Version: 1.2

SAP Contributors:

Joshua Betz, Senior Statistician

Frank Lin, Principal Investigator

Carrie Nieman, Coinvestigator

Emmanuel Garcia Morales, Biostatistician

Jami Trumbo, Senior Research Program Coordinator

All statistical analyses were pre-specified by individuals without knowledge of treatment or outcome information.

SAP Changes and Justifications: version 1.1 to 1.2 – 08/13/2021

The SAP was updated to reflect incrementing the protocol from version 5.0 to 6.0. Protocol changes reflect the addition of measures to address COVID-19 and updating the statistical analysis portion to reference the SAP.

SAP Changes and Justifications: version 1.0 to 1.1 – 07/02/2021

Note that all changes after version 1.0 occurred after unblinding of outcomes. No changes pertain to the primary analysis.

1. **Section 1**: The protocol information was updated.
2. **Section 5**:
    a. Inclusion criteria in version 1.0 indicated a Hearing Handicap Inventory for the Elderly Short Form (HHIE-S) > 8 at baseline: this has been corrected to HHIE-S ≥ 8 at baseline.
    b. For the reporting of study participation, a category was added for individuals who were eligible to participate and consented to be randomized in addition to those who were ultimately randomized.
    c. The reporting of withdrawals, missed visits, and loss to follow up in the CONSORT diagram was clarified. In the primary analysis paper, attrition will be reported up to 3-months post baseline, where the primary outcome was obtained. Subsequent publications will report attrition after the primary outcome.

3. **Section 6**:
    a. The Adapted Listening Self Efficacy Questionnaire was accidentally omitted from the baseline characteristics listed in Table 3.
    b. The SF-12 Mental and Physical Component Scores were accidentally omitted as secondary outcomes in version 1.0: this has been corrected.
    c. Further description was added to the primary analysis, including the commands in Stata and R for the outcome regression model.
    d. The formula for calculating the weights in the Complier Average Causal Effect model was corrected, and additional explanation of the method was added.

Section 2: Introduction

Background and Rationale: Hearing aids, along with adequate counseling and education, form the foundation of a comprehensive approach to hearing health care. Hearing aids can improve audibility, communication, and may help promote better general health and social engagement in older adults (Boi et al., 2012). While an important component of hearing health care, hearing aids are underutilized. Only 15% of older adults undergo hearing screening and less than 20% of older adults with demonstrated hearing loss use hearing aids (Kochkin, 2009; Popelka et al., 1998; Gates et al., 1990). This demonstrates a gap in care that the NIDCD, Healthy People 2020, and the National Academies of Science, Engineering, and Medicine (NASEM) have identified and, subsequently, called for the development of additional models of affordable and accessible hearing health care (Donahue et al., 2010; NASEM, 2016).

One NHANES study found that 8.3% of Blacks and 12.9% of Mexican or Hispanic elderly participants as compared to 19.9% of Whites utilized hearing aids on a regular basis and only 12.5% of those with < $20,000 household income report hearing aid use as compared to 22.9% with household income of ≥ $45,000 (Lin et al.,2011). This study uses the Baltimore HEARS intervention that was developed to aid in NIDCD's charge to deliver hearing health care for all. Baltimore HEARS is an intervention that provides affordable and accessible hearing care to minority and low-income older adults. In previous pilot studies (IRB# NA_00088278 and IRB#00079481), we used a community-engaged research approach to develop a culturally-tailored, community-delivered hearing care intervention that included hearing screening, education on age-related hearing loss, communication strategies, and provision of a low-cost over-the-counter amplification device. The initial pilot study was delivered by a medical expert (i.e., an ENT resident) and demonstrated the feasibility and acceptability of the novel accessible and affordable approach to hearing care intervention for low-income, minority older adults. The follow-up pilot study tested the feasibility of delivering the hearing loss intervention (HEARS program) in the community through trained CHWs.

The current study aim is to conduct an RCT of the HEARS intervention delivered by an audiologist-CHW care team utilizing an immediate treatment and 3-month delayed treatment group of low-income and primarily minority older adults living independently in Baltimore. Lessons and insights acquired through direct encounters with pilot participants and advisory board meetings informed the design and procedure of the proposed study. Active engagement with community representatives through a Community Advisory Board will be ongoing.

Objectives: The primary objective of this study is to conduct a randomized controlled trial studying the efficacy of the HEARS intervention in reducing self-reported hearing handicap as delivered by trained community health workers (CHW). The primary endpoint is change in score on the HHIE-S from baseline to 3-month post-randomization in the immediate vs. delayed (waitlist control) group. The project aims to recruit and enroll eligible community-dwelling older adult residents from independent, affordable housing to receive the intervention. After 3 months of follow-up post-randomization, participants in the waitlist control group will also be offered the HEARS hearing intervention. All participants (also referred to as "clients" in the HEARS intervention materials) will then be followed observationally for up to 12 months post-intervention. Participant outcomes will be collected through phone calls and in-person visits. Long-term follow-up will provide an understanding of both short- and long-term effects of a community-based hearing care program on participants' communication, social, and emotional function. Outcome measures will assess various domains such as communication function, depression, loneliness, and social connectedness, with the primary outcome being communication function as measured using the Hearing Handicap Inventory for the Elderly-Screening (HHIE-S).

Section 3: Study Methods

Intervention Description: The HEARS intervention is a community-delivered hearing care intervention that provides education on hearing loss, communication strategies, and provision of an over-the-counter amplification device. The program is designed for older adults and has been previously piloted locally in partnership with Weinberg Senior Living. The intervention involves two one-hour meetings between the participant and his/her CHW. CHW's will be trained to deliver the HEARS intervention to participants through a separate ongoing study (IRB#00152093).

During the first meeting, the CHW will introduce the HEARS program and review communication strategies and basics of age-related hearing loss, as well as options for amplification including the differences between over-the-counter amplifiers used in the program versus conventional hearing aids typically acquired through audiologists. The second meeting will consist of a step-by-step fitting and orientation to the participant's amplifier of choice. Based on the needs of the participant and the discretion of the CHW, additional meetings may occur to ensure the participant is comfortable using his/her amplifier independently.

Study Design: The study design is an open label, two-arm, parallel group trial: participants are randomized 1:1 to a group which receives the HEARS intervention immediately (the 'immediate group', or 'Wave 1') or receives the HEARS intervention after the 3-month primary endpoint (the 'delayed group', or 'Wave 2'). Randomization is stratified by study site to maximize within-building balance in treatment allocation.

Intervention Description: The intervention ("treatment") consists of aural rehabilitation, which includes basic education on age-related hearing loss and communication strategies, and provision of an over-the-counter amplifier. The amplifier devices utilized in this study are low-cost, self-fit amplifiers that are currently available over the counter, directly to consumers: the Sound World Solutions (SWS) Sidekick and Sonic Technology's SuperEar SE9000. Neither of these devices are medical devices regulated by the FDA. The session will include fitting the device and a step-by-step orientation to the device. The intervention will be delivered by a trained community health worker (CHW) in the participant's community. Treatment would be considered a failure if individuals do not complete the two-session intervention (i.e., did not attend all scheduled intervention meetings).

Sample Size Justification: The primary outcome will be the Hearing Handicap for the Elderly Screening Version (HHIE-S) measured at 3-months post-intervention. The HHIE-S is a validated 10-item questionnaire (range: 0-40 points) to assess perceived effects of hearing loss among older adults (Weinstein et al., 1986). The Baltimore HEARS pilot study showed excellent test-retest reliability of the HHIE in the target population and demonstrated a 'large' effect size on this outcome (Cohen's D=-0.74, N=15).

We hypothesize that assignment to the immediate treatment group will have higher average HHIE scores at follow-up compared to those assigned to the delayed treatment group, who will not receive the intervention prior to the 3-month primary endpoint: a superiority hypothesis.

Effect Size	N per Arm	Total Sample Size	10% Attrition	20% Attrition
0.60	60	120	134	150
0.65	51	102	114	128
0.70	44	88	98	110
0.75	39	78	88	98
0.80	34	68	76	86
0.85	31	62	70	78
0.90	27	54	60	68

Table 1: Sample sizes necessary to achieve 90% power while maintaining a type I error rate at 0.05 using a two-sample t-test. These sample sizes are adjusted to account for potential loss to follow up.

A target sample size of 150 participants would ensure adequate power to test a clinically meaningful effect on the primary outcome, even with 20% attrition.

Framework: All hypotheses are being tested within a superiority framework. Individuals will be randomized to receive the intervention immediately (i.e. prior to the 3-month post-randomization visit, where the primary outcome is obtained), or to a wait-list control (i.e. receive the intervention after the primary outcome is obtained). For the primary outcome, it is hypothesized that individuals randomized to receive the intervention immediately will have greater declines in hearing handicap from baseline, as measured by the HHIE-S, than those assigned to the wait list control. Similarly, immediate treatment is hypothesized to be superior to wait list controls on all secondary outcomes.

Interim Analyses and Stopping Guidance: Given the short follow up window and minimal risk to participants, no interim analyses for efficacy or futility will be conducted. There is no anticipated need to stop the trial for safety concerns.

Timing of the Final Analysis: Data for the primary outcome will be considered complete after the last participant's last visit window for the primary outcome has closed. These records will be locked, although post-intervention follow-up data will continue to be collected. Efficacy of the intervention will be assessed only using information obtained at baseline and at the primary endpoint (3 months post-randomization). All data will be complete after the last participant's last window for the 12-month post-intervention visit has closed.

Timing of Outcome Assessments: A table of study visits and their corresponding visit windows is given in Table 2.

Study Visit	Immediate Treatment: Wave 1	Delayed Treatment: Wave 2
Wave 2: 3-Months Post-Randomization (Repeat Baseline*)	N/A	Goal: 3-months post-randomization Window: Up to 6 months post randomization
3-Months Post-Intervention	Goal: 3-months post-intervention Window: Up to 6 months post intervention	Goal: 3-months post-intervention Window: Up to 6 months post intervention
12 Months Post-Intervention	Goal: 12 months post-intervention Window: Up to 18 months post-intervention	Goal: 12 months post-intervention Window: Up to 18 months post-intervention

Table 2: Windows allowed for each follow-up visit. The efficacy of the HEARS intervention will be assessed at the 3-month post-randomization visit*. After this visit, the delayed treatment group will be offered the intervention, and they will have a 3 months post-intervention study visit. Both immediate and delayed treatment groups will have a 12-month post-intervention follow-up visit.

The primary endpoint is the 3-month post-randomization visit. After this assessment, individuals in the delayed treatment group ('Wave 2') will be offered the intervention and followed up 3 months after the intervention has been administered (with a window up to 6 months post-intervention delivery). Both immediate ('Wave 1') and delayed ('Wave 2') treatment groups will have a 12-month post-intervention follow-up visit. Only the baseline visit and 3-month post-randomization visits will be used to assess the efficacy of the intervention.

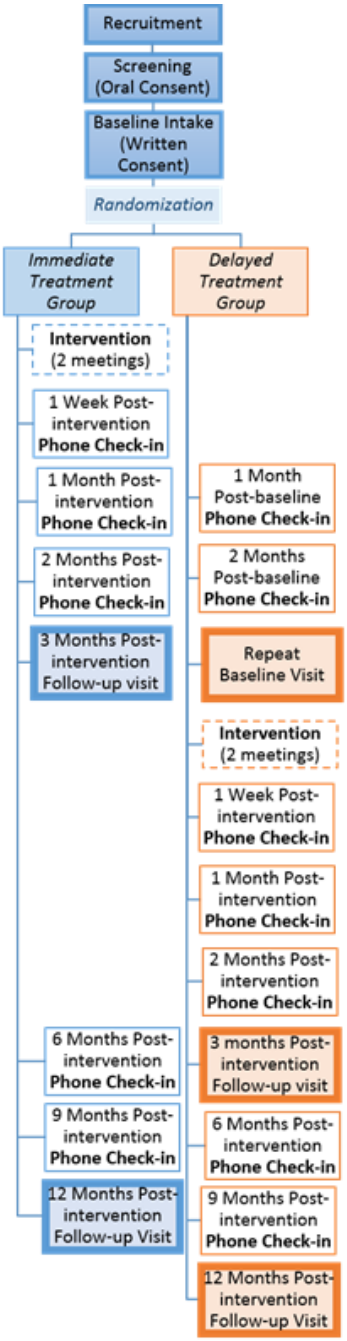

Figure 1: A diagram of study visits by treatment arm. Dotted boxes indicate visits with community health workers (CHWs) and solid boxes indicate visits with study staff.

Section 4: Statistical Principles

Hypothesis tests and Confidence Intervals: This study was designed to achieve at least 90% power to detect an effect size of $\delta \geq 0.6$ on the Hearing Handicap Inventory for the Elderly Short Form (HHIE-S) while maintaining a familywise type I error rate of $\alpha = 0.05$. Since there is only one contrast of interest in one outcome at one endpoint, the average treatment effect at 3-months post randomization between immediate (Wave 1) and delayed (Wave 2) treatment groups, no adjustment for multiplicity is needed. A 95% confidence interval will be reported for the primary outcome, and for all hypothesis-generating secondary outcomes. All hypothesis tests will be two-sided, using a 5% significance level.

Adherence: All analyses will be conducted according to the intention to treat (ITT) principle: participants will be analyzed as they were randomized, irrespective of the receipt of treatment. Individuals would be considered as not adhering to treatment if they do not complete the two-session intervention (i.e., did not attend all scheduled intervention meetings) or do not use the amplification device 1 or more hours per day at 3-months post-intervention. Patterns of adherence will be tabulated and summarized. Additionally, once the delayed treatment group receives the intervention, patterns of adherence will be summarized in aggregate and by treatment arm.

Protocol Deviations: The following protocol deviations and adverse events will be tabulated and summarized.

HEARS Protocol Deviations and Violations by Category

DATA COLLECTION PROCESS:

- Data collector failed to save data leading to data loss (re-surveying not an option).
- Data collector failed to save data but participant is re-surveyed
- Missing data points due to data collector entry error (within measures)
- Only partial outcomes measures were collected due to study window

INTERVENTION DELIVERY:

- Participant did not receive intervention
- Intervention started but not completed
- Intervention delivery outside window
    - Study team delays intervention
    - Interventionist delays intervention
    - Participant delays intervention

3 MONTH STUDY VISIT:

- 3 MONTH STUDY VISIT outside window
    - Study team delays visit
    - Participant delays visit
    - Participant unable to schedule visit (due to traveling, physical health, unable to contact)

ADVERSE EVENT

- Adverse event during study visit (study related)
- Adverse event during study visit (non-study related)
- Adverse event outside of study visit (study related)
- Adverse event outside of study visit (non-study related)

Analysis Populations:

The intention to treat (ITT) sample will include all individuals who were randomized to receive the intervention, irrespective of their receipt or adherence to treatment. No safety analyses will be performed due to the minimal risk nature of the study. When estimating the complier average causal effect (CACE), adherence with the intervention will be defined as attending both scheduled intervention meetings and using the amplification device 1 or more hours per day at 3-months post-intervention.

Section 5: Trial Population

Screening data: Variables that will be tabulated to describe the representativeness of the sample are summarized in table 3.

Eligibility Criteria: Participant inclusion and exclusion criteria are as follows:

- Inclusion criteria:
    - Age 60 years or older
    - English-speaking
    - Aural-oral verbal communication as primary communication modality
    - Post-lingual hearing loss (Audiometric pure tone averages [0.5-4kHz] in both ears >25 dB)
    - Does not currently use a hearing amplification device or hearing aid
    - Signed informed consent to participate in all study related activities
    - Willing to regularly use listening device once provided for the remainder of their time in the study
    - Hearing handicap as measured by HHIE-S score ≥8
    - Able to follow study instructions
- Exclusion criteria:
    - Evidence of ear disease or pathology requiring further medical evaluation

CONSORT Criteria:

- Assessed for eligibility: Those who were screened for eligibility in the study
- Excluded: Those who did not meet study eligibility criteria
    - Age: Number of people who did not meet age eligibility
    - English Fluency: Number of people who did not meet English fluency eligibility
    - Study commitment: Number of people who did not meet study commitment eligibility
    - Hearing Aid History: Number of people who did not meet hearing aid history eligibility
    - Otologic Screen Score: Number of people who did not meet otologic screen eligibility
    - PTA <25dB: number of people who did not meet minimum PTA threshold eligibility in their better ear
    - Screening Reliability: number of people who did not meet screening reliability eligibility

    - HHIE<8: number of people who did not meet minimum HHIE score eligibility
- Consented: Number of eligible individuals who consented to be randomized
- Randomized: Number of people who were randomized
- Allocation to Immediate Intervention: Number of people allocated to receive the intervention immediately
    - Received Intervention: number of people who successfully completed the intervention within delivery window (1 month from baseline visit)
    - Delayed intervention: number of people who successfully completed the intervention, but outside of delivery window (over 1 month from baseline visit)
    - Did not receive Intervention: Number of people who did not successfully complete the intervention due to refusal, withdrawal, or loss to follow up
- Allocation to Delayed intervention: Number of participants allocated to receive a delay in intervention delivery. The primary analysis paper will focus on the primary contrast of interest, i.e., those immediately treated in Wave 1 vs. those waitlisted in Wave 2 at 3-months post-randomization. Presentation and analysis of longer-term outcomes, including post-intervention outcomes for Wave 2, will be covered under a different statistical analysis plan.
    - Received Intervention: number of people who successfully completed the intervention within delivery window (1 month from repeat baseline visit)
    - Delayed intervention: number of people who successfully completed the intervention, but outside of delivery window (over 1 month from repeat baseline visit)
    - Did not receive Intervention: Number of people who did not successfully complete the intervention due to refusal, withdrawal, or loss to follow up
- Loss to Follow up: Number of people who were lost to follow up (defined as no successful contact with participant more than 6 months from intervention delivery)
- Discontinued Study: Number of participants who discontinued study early
    - Obtained Hearing Aids: Participant discontinued study participation due to obtaining hearing aids outside of the study before intervention or follow up meeting
    - Study related adverse Event: Participant discontinued study participation due to study related adverse event
    - Non-study related adverse event: Participant discontinued study participation due to a non-study related adverse event
    - Withdrawn consent: Participant withdrew consent to continue study
    - Withdrew due to adverse event not resulting in death: Participant withdrew consent to continue study due to non-fatal adverse event
    - Death: participant passed before receiving intervention or completing follow up meeting
    - Other: participant discontinued study due to other reason not listed above
- Analyzed: Number of participants who completed primary outcome
    - Timing of primary outcome collection:
        - Primary outcome collected within of study window.
        - Primary outcome collected outside of study window: number of people who were unable to complete primary outcome within window of 6 months after intervention delivery (immediate group) or 6 months after baseline (delayed group)
    - Excluded from analysis: number of participants who are excluded from analysis due to varying reasons

- Missing Primary outcome: number of participants who did not complete primary outcome

Withdrawals, missed visits, and loss to follow up will be reported in consort diagram by study visit (3, 6, and 12-months post-enrollment). In the primary analysis, only 3-month post-randomization information will be reported, with other attrition reported in subsequent publications.

Scoring of Outcomes: Instruments with any missing items (e.g. "Refused", "Don't Know") will be coded as missing unless there are published criteria for handling missing items. For the Cohen Social Network Index, social network size (SNS) will be scored using published criteria that involve top-coding groups of size greater than 6 at 7 individuals in item 12 – this keeps the scoring of item 12 consistent with items 5-11 and limits the range of SNS between 0-103.

Baseline patient characteristics: Baseline characteristics will be summarized in aggregate and by treatment arm as specified in Table 3 below.

Measure	Domain	Type	Summary
Age	Demographics	Continuous	Mean (SD)
Sex	Demographics	Categorical	Frequency (%)
Race/ethnicity	Demographics	Categorical	Frequency (%)
Living arrangement	Demographics	Categorical	Frequency (%)
Education	Demographics	Categorical	Frequency (%)
Monthly household income	Demographics	Categorical	Frequency (%)
MoCA	Cognition	Continuous	Mean (SD)
Speech-Frequency Pure Tone Average	Hearing	Continuous	Mean (SD)
Hearing loss category	Hearing	Categorical	Frequency (%)
Prior hearing screening	Hearing	Categorical	Frequency (%)
Prior hearing aid use	Hearing	Categorical	Frequency (%)
HHIE-S	Hearing & Communication	Continuous	Mean (SD)
UCLA Loneliness Scale	Social Isolation & Loneliness	Continuous	Mean (SD)
PHQ-9	Depression	Continuous	Mean (SD)
Cohen SNI: Network Diversity	Social Network	Continuous	Mean (SD)
Cohen SNI: Social Network Size	Social Network	Continuous	Median (IQR)
SF-12 Mental Component Score	Health Related Quality of Life	Continuous	Mean (SD)
SF-12 Physical Component Score	Health Related Quality of Life	Continuous	Mean (SD)
REALM-SF	Health Literacy	Continuous	Mean (SD)
AATCQ Computer Self Efficacy	Technology Self Efficacy	Continuous	Mean (SD)
AATCQ Listening Device Interest	Technology Self Efficacy	Continuous	Mean (SD)
AATCQ Listening Device Self Efficacy	Technology Self Efficacy	Continuous	Mean (SD)
ALSEQ Listening Self Efficacy	Listening Self Efficacy	Continuous	Mean (SD)
Own smartphone	Technology Self Efficacy	Categorical	Frequency (%)
Used a computer within past month	Technology Self Efficacy	Categorical	Frequency (%)
Used text or email within past month	Technology Self Efficacy	Categorical	Frequency (%)

Table 3: Baseline characteristics of participants to be summarized. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted. Hearing loss in the better hearing ear will be categorized using the speech frequency pure tone

average (PTA in db HL) as Normal (≤ 25) Mild (25 < PTA ≤ 40), Moderate or greater (>40). Abbreviations – MoCA: Montreal Cognitive Assessment; SNI: Social Network Index; SF-12: Short Form 12; AATCQ: Adapted Attitudes Towards Computers Questionnaire; REALM-SF: Rapid Estimate of Adult Literacy in Medicine – Short Form; ALSEQ: Adapted Listening Self-Efficacy Questionnaire

Section 6: Analysis

Study Outcomes: All study outcomes are listed in Table 4 below.

Outcome Measure	Outcome Type	Transformation	Units [Range]
HHIE-S	Primary	Change from Baseline	Points [-40, 40]
UCLA Loneliness Scale	Secondary	Change from Baseline	Points [-60, 60]
PHQ-9	Secondary	Change from Baseline	Points [-27, 27]
SF-12 Mental Component Score (MCS)	Secondary	Change from Baseline	Points [-41.7, 41.7]
SF-12 Physical Component Score (PCS)	Secondary	Change from Baseline	Points [-32.6, 32.6]
SNI: Network Diversity	Secondary	Change from Baseline	Points [-12, 12]
SNI: Social Network Size	Secondary	Change from Baseline	Points [-103, 103]
Valuation of Life (VoL)	Secondary	Change from Baseline	Points [-52, 52]
ALSEQ Total Score	Secondary	Change from Baseline	Points [-50, 50]
AATCQ Computer Self Efficacy	Secondary	Change from Baseline	Points [-20, 20]
AATCQ Device Self Efficacy	Secondary	Change from Baseline	Points [-20, 20]
AATCQ Device Interest	Secondary	Change from Baseline	Points [-20, 20]
ALSEQ Subscale: Dialogue in Quiet (DQ)	Exploratory	Change from Baseline	Points [-10, 10]
ALSEQ Subscale: Directed Listening (DL)	Exploratory	Change from Baseline	Points [-20, 20]
ALSEQ Subscale: Complex Listening (CL)	Exploratory	Change from Baseline	Points [-20, 20]
Hearing Related Knowledge	Exploratory	Untransformed Score	Points [0, 8]

Table 4: A list of primary and secondary outcomes for the HEARS trial, including the instruments used, transformation, and the scale and possible range of the measurement. Each outcome is assessed at baseline, and at each of the study visits (3-month post randomization and 12-month post-intervention in both immediate and delayed treatment groups; 3-month post intervention (6-month post-randomization in the delayed treatment group). Abbreviations - SNI: Social Network Index; SF-12: Short Form 12; AATCQ: Adapted Attitudes Towards Computers Questionnaire; ALSEQ: Adapted Listening Self-Efficacy Questionnaire

Primary Analysis: The primary analysis will use a doubly-robust weighted least squares (DR-WLS) estimator of the average treatment effect (ATE) of the HEARS intervention on the change in HHIE-S score at 3-months post randomization on the sample of complete cases under the intention to treat principle. This class of estimators consist of an outcome model component and a missingness model component, which can give consistent estimates of the treatment when the missing at random (MAR) assumption holds and either the outcome model or the missingness model is correctly specified (Robins, Sued, Lei-Gomez, & Rotnizky, 2007). This estimator will also mitigate the effects of chance imbalances between randomized groups in factors that are known to be associated with the outcome and improve precision of estimates. Adjusted analyses will include age and the 4-frequency pure tone average (PTA) in the better hearing ear (the average of the individual's hearing thresholds at 0.5, 1, 2, and 4 kHz, the band of frequencies in which most speech information is encoded) as linear terms.

For participant i, let R_i indicate the event that their outcome 3 months post-randomization, denoted Y_{i1}, is observed at the 3-month post-randomization follow-up ($R_i = 1$ if $Y_{i1}$is observed, and

$R_i = 0$ otherwise), let Y_{i0} denote their baseline outcome score, and let T_i indicate assignment to the immediate treatment arm ($T_i = 1$ for Wave 1, the immediate treatment arm, and $T_i = 0$ for Wave 2, the delayed treatment arm).

First, any missing baseline covariate values (baseline outcome, age, PTA) are imputed using mean imputation. Next, a logistic regression model will be used to model the probability of receiving the intervention (immediate treatment) according to baseline characteristics.

$$logit(\Pr\{T_i = 1\}) = logit(\pi_i^T) = \alpha_0 + \alpha_1 Y_{i0} + \alpha_2 Age + \alpha_3 PTA$$

The fitted values from this model for participant i is $\hat{\pi}_i^T = expit\{\boldsymbol{X}_i'\hat{\boldsymbol{\alpha}}\}$, where $expit(\cdot)$ is the inverse of the logistic function. The propensity score for participant i is $\hat{\pi}_i^{PS} = T_i\hat{\pi}_i^T + (1 - T_i)(1 - \hat{\pi}_i^T)$.

After the propensity score is created, a logistic regression model will be used to model the probability of being observed at follow-up according to baseline characteristics (the outcome measured at baseline, age, PTA, and treatment assignment). The propensity score model will be defined as follows:

$$logit(\Pr\{R_i = 1\}) = logit\left(\pi_i^R\right) = \gamma_0 + \gamma_1 Y_{i0} + \gamma_2 Age + \gamma_3 PTA + \gamma_T T_i$$

The fitted values from this model for participant i is $\hat{\pi}_i^R = expit\{\boldsymbol{X}_i'\hat{\boldsymbol{\gamma}} + \gamma_T T_i\}$. An inverse probability weight for each participant is then created: $w_i = 1/(\hat{\pi}_i^R\hat{\pi}_i^{PS})$. Inverse probability weights above 20 will be set to 20 to mitigate the influence of individuals with large inverse probability weights.

After the inverse weight is created, this weight is used in a g-computation step. The model used in g-computation involves transforming the outcome to the range [0, 1], and performing a weighted logistic regression (Gruber & van der Laan, 2010). If an outcome has no maximum value, no transformation is performed. The outcome for individual i, Y_{i1}, is mapped to the range [0, 1]: if y_{MAX} and y_{MIN} respectively denote the maximum and minimum possible values of the outcome, this transformation and its inverse are given by:

$$g(y) = (y - y_{MIN})/(y_{MAX} - y_{MIN})$$

$$g^{-1}(y) = y(y_{MAX} - y_{MIN}) + y_{MIN}$$

For example, the primary outcome is the change in HHIE from baseline, this ranges from $y_{MIN} = -40$ to $y_{MAX} = 40$. The transformed outcome on the [0, 1] interval is then modeled using a logistic regression model, weighted by the inverse probability weights described above:

$$logit(g(Y_{i1})) = \beta_0 + \beta_1 Y_{i0} + \beta_2 Age_i + \beta_3 PTA_i + \beta_T T_i$$

This model can be fit using an overdispersed binomial or 'quasibinomial' family of generalized linear model. In R, this involves using the `glm` command with `family = "quasibinomial` argument, and in Stata, this can be fit using the `fracreg` command with a logit link. For secondary outcomes that do not have a defined range, a weighted linear regression model is fit, weighted by the inverse probability weights described above:

$$Y_{i1} = \beta_0 + \beta_1 Y_{i0} + \beta_2 Age_i + \beta_3 PTA_i + \beta_T T_i + \varepsilon_i$$

Using the appropriate weighted regression model, a predicted outcome is generated for each participant under assignment to treatment and control. For the logistic regression model:

$$\hat{\eta}_i^{(1)} = expit\left(\hat{\beta}_0 + \hat{\beta}_1 Y_{i0} + \hat{\beta}_2 Age_i + \hat{\beta}_3 PTA_i + (\hat{\beta}_T \cdot 1)\right)$$

$$\hat{\eta}_i^{(0)} = expit\left(\hat{\beta}_0 + \hat{\beta}_1 Y_{i0} + \hat{\beta}_2 Age_i + \hat{\beta}_3 PTA_i + (\hat{\beta}_T \cdot 0)\right)$$

For the linear regression model:

$$\hat{\eta}_i^{(1)} = \hat{\beta}_0 + \hat{\beta}_1 Y_{i0} + \hat{\beta}_2 Age_i + \hat{\beta}_3 PTA_i + (\hat{\beta}_T \cdot 1)$$

$$\hat{\eta}_i^{(0)} = \hat{\beta}_0 + \hat{\beta}_1 Y_{i0} + \hat{\beta}_2 Age_i + \hat{\beta}_3 PTA_i + (\hat{\beta}_T \cdot 0)$$

The doubly robust estimator for the average treatment effect is calculated as the average difference between counterfactual outcomes. If the outcome was originally transformed, this average is taken after they are transformed back to the original scale of the outcome:

$$\hat{\delta}_{DR} = \frac{1}{n}\sum_{i=1}^{n}\left(g^{-1}\left(\eta_i^{(1)}\right) - g^{-1}\left(\eta_i^{(0)}\right)\right)$$

For untransformed outcomes, no inverse transformation is required:

$$\hat{\delta}_{DR} = \frac{1}{n}\sum_{i=1}^{n}\left(\eta_i^{(1)} - \eta_i^{(0)}\right)$$

The advantage of the transformation is that the fitted values $g^{-1}\left(\eta_i^{(t)}\right)$ are constrained to the range of their respective outcomes. While Gruber and van der Laan (2010) use the Targeted Maximum Likelihood Estimator rather than the DR-WLS estimator proposed here, the same principles apply.

Confidence intervals will be obtained using bias corrected and accelerated (BCa) bootstrap resampling procedure with 10,000 replicates. Imputation of any missing baseline covariates will be performed within each bootstrap sample.

The primary analysis will utilize the HHIE-S score obtained in-person whenever possible irrespective of study window and will use HHIE-S Scores obtained over the phone when an in-person visit was not obtained. Sensitivity analyses will include repeating the analysis with exclusion of outcomes obtained beyond 30 days outside of the study window: this will involve a new missing data indicator $R_i^S = 1$ if $Y_{i1}$is observed within the study window, and $R_i^S = 0$ otherwise.

Secondary Analyses: The secondary analyses will follow similarly from the primary analysis, using the same doubly robust estimator as described above. Secondary outcomes are listed in table 4: secondary outcome models will have the same specification for the outcome model (treatment indicator and linear terms for the baseline value of the outcome, age, and PTA) and the propensity score model (treatment indicator and linear terms for the baseline value of the outcome, age, and PTA).

Planned secondary analyses will be performed on the primary outcome, communication function as measured by HHIE-S, using the complier average causal effect (CACE) of the intervention, following a methodology similar to Stuart and Jo (2015). In this analysis, individuals are conceptualized as belonging to four latent classes, known as principal strata, which are unaffected by randomization. These strata include ‘compliers,’ who comply with whichever intervention assigned to them, ‘always takers,’ who will always take the treatment regardless of their treatment assignment, ‘never takers,’ who will never take

the treatment regardless of their treatment assignment, and 'defiers,' who will take whichever treatment they are not assigned. Adherence with the intervention will be defined as attending both scheduled intervention meetings and using the amplification device 1 or more hours per day at 3-months post-intervention. Since the HEARS intervention is not available outside of the study, we assume that there are no 'always takers,' and similarly, we assume that there are no 'defiers.'

This analysis will be similar to the analysis described above, with an additional step added in creating the weights for the weighted regression. We will fit a logistic regression model for the probability of adherence or compliance, denoted A_i, among those in the immediate treatment group (wave 1):

$$logit(\Pr\{A_i = 1|T_i = 1\}) = logit\left(\pi_i^A\right) = \theta_0 + \theta_1 Y_{i0} + \theta_2 Age + \theta_3 PTA$$

This model will be used to predict the probability of adherence in the wait list control group (wave 2) based on these characteristics. A propensity score will be calculated as follows:

$$1/\hat{\pi}_i^A = \begin{cases} 1: T_i = 1, A_i = 1 \\ 0: T_i = 1, A_i = 0 \\ 1/expit\{\boldsymbol{X}_i'\widehat{\boldsymbol{\theta}}\}: T_i = 0 \end{cases}$$

Since compliance status (A_i) can only be observed among those in the treatment arm, those in the control arm ($T_i = 0$) will have their propensity to comply predicted using the regression model fit among individuals in the treatment arm ($T_i = 1$). An additional weight will be created as follows:

$$w_i^C = 1/(\hat{\pi}_i^R \hat{\pi}_i^{PS} \hat{\pi}_i^A)$$

This weight w_i^C, which incorporates the compliance propensity score $\hat{\pi}_i^A$, will be used in the estimation procedure described above instead of w_i. The goal of this procedure is to re-weight the data to not only account for differences between arms at baseline and differences due to attrition, but also to account for differences between observed compliers in Wave 1 and waitlisted controls in Wave 2.

This method uses the assumption of principal ignorability (PI), which assumes that principal stratum membership ('compliers' vs. 'never takers') is conditionally independent of the potential outcomes given the observed covariates. Additional exploratory analyses may be conducted under different causal inference assumptions or utilizing different predictors of treatment adherence.

Exploratory subgroup analyses will be performed to examine the potential influence of factors on participant outcomes, including 1) cognition (normal cognition MoCA total score ≥ 26 vs. Cognitive impairment MoCA total score < 26; including adjustment of 1 additional point for individuals with 12 years or fewer of education), 2) severity of hearing loss (mild hearing loss ≤ 40 dB better ear speech PTA vs. More severe hearing loss > 40 dB better ear speech PTA) and 3) education (High school or less vs. Greater than high school). When subgroups contain fewer than 30 individuals with observed outcome data or rates of missingness that pose convergence problems for inverse probability weight models, alternative methods may be considered.

Missing Data: As stated above, instruments with any missing items (e.g. "Refused", "Don't Know") will be coded as missing unless there are published criteria for handling missing items. Frequency and patterns of missing data will be tabulated in aggregate and by treatment arm. The doubly robust estimator is valid under the assumption that data are missing at random: that the degree of missingness does not depend on unobserved variables or the values of the missing data. Additionally, a

supplementary table will compare baseline characteristics between those with missing 3-month post-randomization outcomes to those with complete data both in aggregate and by treatment arm.

If an individual's baseline outcome is missing, and the outcome of interest is a baseline change score, imputed values will only be used to calculate propensity scores and inverse probability weights: since the original value was not observed, the change from baseline will be considered missing.

Risks and Potential Harm: The risks involved in participation in the study are believed to be low. There is a theoretical risk of noise exposure related to use of an amplification device, primarily if the device malfunctions. The risk of noise exposure is minimized with volume limitations as part of the device design. All products used in the study are commercially available and all models used in the study are considered over-the-counter devices.

Other than the intervention failing to produce an effect, there are no known possible adverse effects documented in the literature or experienced in prior pilot studies involving over-the-counter amplification devices and the HEARS intervention. All devices and intervention materials have been either previously used in pilot studies or are currently used by consumers in the United States with no known harmful effects.

During the study, survey questions regarding depression, social isolation, and loneliness will be posed to participants. All participants receive a brochure detailing information regarding both audiological and otological care as well as behavioral health resources. During survey administration, depression severity is assessed using PHQ-9 (score 0-27). Participants who score of 15 or above will be counseled to speak with their primary care physician. An offer to help contact their primary care physician is given and effort made to relay survey score results to their primary care office is made if permission is given. The occurrence is noted and resources provided to the participant. In addition, if the participant mentions any level of intention to harm self or others, verbal contract not to harm self is collected and adverse event protocol is followed if immediate threat to self of others is apparent. In the case of participants who score 5-14 on the PHQ-9, the occurrence is also noted and resources provided to the participant. Occurrences where the PHQ-9 score is 5 or above are noted, but the occurrence is not considered an adverse event.

Given the low risk related to the HEARS intervention, no to few adverse events are expected. All adverse events are recorded by the research coordinator and immediately reported to the Principal Investigator. The research coordinator will collect information regarding the adverse event from the participant and/or informant, including details surrounding the event to inform assessment of the severity of the event, causality, and the relatedness to the intervention and trial involvement.

Information on adverse events will be summarized in terms of severity, causality, and relatedness to the study intervention and information presented in the final trial results.

Statistical Software: Data management and analyses will be conducted using the R environment for Statistical Computing and Stata as needed. In R, 95% BCa bootstrap confidence intervals will be calculated using the boot package.

References:

Gruber, S., & van der Laan, M. J. (2010). A Targeted Maximum Likelihood Estimator of a Causal Effect on a Bounded Continuous Outcome. The International Journal of Biostatistics, 6(1):1-14. https://doi.org/10.2202/1557-4679.1260

Robins, J.M., Sued, M., Lei-Gomez, Q., Rotnizky, A. (2007). Comment: Performance of Double-Robust Estimators when Inverse Probability Weights are Highly Variable. Statistical Science, 22(4):544–559.

Stuart, E.A., & Jo, B. (2015). Assessing the sensitivity of methods for estimating principal causal effects. Stat Methods Med Res, 24(6):657-74. doi: 10.1177/0962280211421840. Epub 2011 Oct 3. PMID: 21971481; PMCID: PMC3253203.

Social Integration and Network Size - The Common Cold Project - Carnegie Mellon University, www.cmu.edu/common-cold-project/measures-by-study/psychological-and-social-constructs/social-relationships-loneliness-measures/social-integration-network-size.html