Patient-Centered Hearing Aid Trial: Comparison of Direct-To-Consumer Delivery Models for Hearing Devices (HL-2019C1-16094)

Document Date – October 18, 2024

Analysis Plan:

Primary and Secondary Outcome Measures: The Profile of Hearing Aid Performance (PHAP, Cox & Gilmore, 1990) and the Hearing Handicap Inventory for the Elderly (HHIE, Ventry & Weinstein, 1982) served as the instruments for the primary and secondary outcome measures, respectively. These instruments were administered in the first in-person session before participants were fitted with hearing aids (S1) to establish baseline, approximately six weeks later (S2), and for a final time approximately six months after the initial fitting (S2). Benefit, computed as the difference in scores on these instruments, S1 score – S2 score and S1 score – S3 score, were the outcome measures of interest. The benefit measures are labeled HHIE-total benefit. The benefit for the PHAP is termed Profile of Hearing Aid Benefit (PHAB) which can be calculated as a global score (PHAB global) for five of the seven subscales pertaining to speech communication. These five subscales are: Familiar Talkers (FT) Ease of Communication (EC), Reverberation (RV), Reduced Cues (RC), and Background Noise (BN). Finally, these scores can be computed from the entire instrument (66 items) or from an abbreviated version, the Abbreviated Profile of Hearing Aid Benefit (APHAB), a 24-item subset.

Sample Size and Statistical Power: To determine the sample size needed, estimates of treatment effects, effect sizes, and common standard deviations for the primary and secondary outcome measures were derived from Humes et al. (2017). As in the present randomized controlled trial (RCT), Humes et al. (2017) obtained unaided or pre-fit measures for the PHAP and the HHIE total initially, with aided PHAP and HHIE total scores obtained at the conclusion of a 6-week trial. For the primary outcome measure, the PHAB-global, the mean treatment effect for AB service delivery relative to placebo control was 11 in Humes et al. (2017). Incidentally, the common standard deviation was also 11, yielding a Cohen's d effect size of 1.0 for the PHAB-global. As a non-inferiority RCT, a margin must also be specified a priori, which effectively represents a

boundary between clinically significant inferior and non-inferior outcomes. The most widely adopted convention is to set the margin to ½ the effect of the standard of care (Audiology based – AB) relative to placebo, or ½ the pooled standard deviation (D'Agostino et al., 2003; Norman et al., 2003). Following the latter convention, a non-inferiority margin of 5.5 was established for the PHAB global. It also translated, in the pertinent region of the PHAP rating scale, to a change of about ½ scale unit on the 7-point PHAP response scale, which would likely be considered a minimally clinically important change from unaided to aided performance. For the secondary outcome measure, HHIE-total benefit, the treatment effect for AB service delivery relative to placebo was 13 with a common standard deviation of 13 as well (Humes et al., 2017), consistent with observations in several prior studies using the HHIE (Chisolm et al., 2007). Accordingly, following the same rationale described for the PHAB-global, the non-inferiority margin for the HHIE-total benefit score was set to 6.5.

nQuery Advisor (V8.3.1) was used to estimate the sample size needed for the non-inferiority analyses of the comparative effectiveness of fitting methods. Two separate one-sided significance tests of group differences, AB vs Consumer Decides (CD) and AB vs Easy Fitting (EF), were planned for the outcome measures. As a result, the significance level was set to 0.025 (0.05/2). Power calculations estimated that when the sample size in each group was 64, a two group one-sided 0.025 significance level t-test for multiple (two) comparisons will have 80% power to reject the null hypothesis that the test (CD or EF) and standard (AB) are not non-inferior [the difference in means, μCD-μAB or μEF-μAB, is 5.5 (PHAB), or is 6.5 (HHIE benefit), or farther from zero in the same direction] in favor of the alternative hypothesis. The alternative hypothesis is that the means of the two groups are non-inferior, assuming the expected difference in means is 0 and the common standard deviation is 11 and 13 for the PHAB global and HHIE total benefit measures, respectively. Thus, with three groups and a required N of 64 per group, a total of 192 participants were needed.

Based on prior work of a similar nature (Humes et al., 2019; Humes et al., 2017), it was assumed that about 36% of those who enrolled would have audiometrically defined normal hearing (better-ear PTA4 < 20 dB HL), better than the sponsor-required range of mild to moderate audiometric hearing loss. As a result, a total of 389 individuals (130 per group) were targeted for study completion (Session 3) so that at least 249 individuals (83 per group) with mild or moderate hearing loss would be included in the analyses. Assuming, again based on experience with similar RCTs, that overall attrition from study enrollment to the 6-month follow-up would be about 35%, the target enrollment at Session 1 was established to be 591 (197 per group).

A total of 584 adults enrolled at Session 1, but both the percentage with normal hearing and the attrition rates were lower than anticipated. In the end, enrollment of 584 participants in Session 1 resulted in 172, 171 and 180 adults with mild or moderate audiometric hearing loss assigned to the AB, CD, and EF groups. Of these, 152, 133, and 138 completed all three sessions for groups AB, CD, and EF, respectively. Based on the desired sample size of 64 per group for the primary (PHAB global) and secondary (HHIE total benefit) outcome measures, as well as the sample size of 83 per group targeted for eventual APHAB global analyses, the intention-to-treat (ITT) and per-protocol (PP) analyses were sufficiently powered. In addition, when the numbers of those with mild or moderate audiometric hearing loss in each group were tabulated, ITT analyses were sufficiently powered (group sizes > 64) to perform sensitivity analyses for each hearing-loss severity separately using the primary and secondary outcome measures.

As noted, in addition to the PHAB-global and HHIE-total benefit primary and secondary outcome measures, a measure of quality of life, the PROMIS-10 Global (Hays et al., 2009), was obtained at Sessions 1 and 3. In particular, the raw PROMIS physical-health and mental-health scores were computed following PROMIS-10 guidelines, and these two subscale scores represented an additional long-term outcome measure obtained in this RCT. No non-inferiority

margin was established *apriori* for this measure because it was not the primary or secondary outcome and because few data were available from RCTs of hearing-aid interventions that had used the PROMIS-10 Global.

Analyses: Some participants in each branch decided to withdraw from the study prior to completion. All participants remained in the branch to which they were randomly assigned at study outset. SAS Version 9.4 was used to impute missing data using the Markov Chain Monte Carlo (MCMC) method (n=20 imputations) and the multivariate-normal approach. Both ITT and PP analyses were performed, as has been recommended (Rehal et al., 2016). After addressing the missing data, initial analysis of the effects of fitting method followed the recommendations by the International Conference on Harmonization (2001; 1998) and CONSORT (Piaggio et al., 2001; Piaggio et al., 2006). The primary focus of the analyses is the comparative effectiveness of each self-fitting method (CD, EF) to best practices (AB). Finally, sensitivity analyses were performed by examining the outcomes separately for those with mild versus moderate audiometric hearing loss. All statistical analyses used SPSS Version 29.0.1.0. This included boot-strapped (n=1000; Bias-corrected and accelerated, BCa) General Linear Model analyses, t-tests, means, and 95% confidence intervals.

References

Chisolm, T. H., Johnson, C. E., Danhauer, J. L., Portz, L. J., Abrams, H. B., Lesner, S., McCarthy, P. A., & Newman, C. W. (2007). A systematic review of health-related quality of life and hearing aids: final report of the American Academy of Audiology Task Force On the Health-Related Quality of Life Benefits of Amplification in Adults. *J Am Acad Audiol*, *18*(2), 151-183. https://www.ncbi.nlm.nih.gov/pubmed/17402301

Cox, R. M., & Gilmore, C. (1990). Development of the Profile of Hearing Aid Performance (PHAP). *J Speech Hear Res*, *33*(2), 343-357. https://doi.org/10.1044/jshr.3302.343

D'Agostino, R. B., Sr., Massaro, J. M., & Sullivan, L. M. (2003). Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med*, *22*(2), 169-186. https://doi.org/10.1002/sim.1425

Hays, R. D., Bjorner, J. B., Revicki, D. A., Spritzer, K. L., & Cella, D. (2009). Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res*, *18*(7), 873-880. https://doi.org/10.1007/s11136-009-9496-9

Humes, L. E., Kinney, D. L., Main, A. K., & Rogers, S. E. (2019). A Follow-Up Clinical Trial Evaluating the Consumer-Decides Service Delivery Model. *Am J Audiol*, *28*(1), 69-84. https://doi.org/10.1044/2018_AJA-18-0082

Humes, L. E., Rogers, S. E., Quigley, T. M., Main, A. K., Kinney, D. L., & Herring, C. (2017). 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*, *26*(1), 53-79. https://doi.org/10.1044/2017_AJA-16-0111

International Conference no Harmonisation. (2001). *Good Clinical Practice*.

International Conference on Harmonisation. (1998). *Statistical Principles for Clinical Trials*.

Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*, *41*(5), 582-592. https://doi.org/10.1097/01.MLR.0000062554.74615.4C

Piaggio, G., Carroli, G., Villar, J., Pinol, A., Bakketeig, L., Lumbiganon, P., Bergsjo, P., Al-Mazrou, Y., Ba'aqeel, H., Belizan, J. M., Farnot, U., Berendes, H., & Group, W. H. O. A. C. T. R. (2001). Methodological considerations on the design and analysis of an equivalence stratified cluster randomization trial. *Stat Med*, *20*(3), 401-416. https://doi.org/10.1002/1097-0258(20010215)20:3<401::aid-sim801>3.0.co;2-1

Piaggio, G., Elbourne, D. R., Altman, D. G., Pocock, S. J., Evans, S. J., & Group, C. (2006). Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*, *295*(10), 1152-1160. https://doi.org/10.1001/jama.295.10.1152

Rehal, S., Morris, T. P., Fielding, K., Carpenter, J. R., & Phillips, P. P. (2016). Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open*, *6*(10), e012594. https://doi.org/10.1136/bmjopen-2016-012594

Ventry, I. M., & Weinstein, B. E. (1982). The hearing handicap inventory for the elderly: a new tool. *Ear Hear*, *3*(3), 128-134. https://doi.org/10.1097/00003446-198205000-00006