Statistical Analysis Plan

A WhatsApp-Delivered AI-Enabled Support Programme versus Paper-Based Support for First-Time Hearing Aid Users in Low-Resource Communities: A Randomized Controlled Trial

Version 1.0

Trial Registration Number: NCT07690670

Date: 23 June 2026

Authors: Caitlin Frisby, Karina C. De Sousa, De Wet Swanepoel, Marien A. Graham, David R. Moore, Naveen Nagaraj

Corresponding author: De Wet Swanepoel
dewet.swanepoel@up.ac.za
University of Pretoria, Department of Speech-Language Pathology and Audiology

Contents

1. Administrative Information

1.1 Study identifiers

1.2 Contributors to the statistical analysis plan

2. Study site and investigators

2.1 Study site

2.2 Study investigators and administrative structure

3. Introduction

3.1 Overview of the study

3.2 Study design

3.3 Study interventions

3.4 Objectives and endpoints

4. Study population

4.1 Sample size determination and justification

4.2 Inclusion and exclusion criteria

4.3 Analysis populations

5. Statistical analysis

5.1 General principles

5.2 Withdrawals, dropouts and handling of missing data

5.3 Level of statistical significance

5.4 Statistical software

5.5 Participant characteristics and baseline comparability

5.6 Primary endpoint analysis

5.7 Secondary endpoint analyses

5.8 Exploratory and sensitivity analyses

5.9 Adverse events and AI-related adverse events

6. Data management

6.1 Overview of data collection methods

6.2 Data entry, cleaning and validation process

6.3 Software and tools used for data management and analysis

7. Reporting of results

7.1 Format for reporting outcomes

7.2 Graphical presentation of results

8. References

1. Administrative Information

1.1 Study identifiers

- Research ethics approval: University of Pretoria Humanities Research Ethics Committee; approval number to be inserted once available.
- Clinical trial registry: ClinicalTrials.gov; identifier to be inserted once registered.
- Protocol version: R33 RCT #2 Protocol, Version 11.0, dated 23 June 2026.

1.2 Contributors to the statistical analysis plan

Name	Primary affiliation	Role on the study	SAP contribution
Caitlin Frisby	University of Pretoria, Department of Speech-Language Pathology and Audiology, South Africa	Research coordinator & Investigator	Prepared initial draft and statistical analyses
Karina C. De Sousa	University of Pretoria, Department of Speech-Language Pathology and Audiology, South Africa	Research coordinator & Investigator	Research coordinator & Investigator
De Wet Swanepoel	University of Pretoria, Department of Speech-Language Pathology and Audiology, South Africa	Co - Primary Investigator	Co - Primary Investigator
David R. Moore	Cincinnati Children's Hospital Medical Center, United States	Co-Principal Investigator	Reviewed draft.
Naveen Nagaraj	University of Colorado / study collaborator	Co-Principal Investigator	Reviewed draft.
Marien A. Graham	University of South Africa, Department of Mathematics Education	Study Statistician	Reviewed and critically revised statistical analysis plan.

2. Study site and investigators

2.1 Study site

The clinical study will take place at community centres or participant homes, where required, across two sites across two provinces in South Africa:

- Atteridgeville District, Gauteng Province, South Africa.
- Khayelitsha District & Drakenstein Distrcit, Western Cape Province, South Africa.

2.2 Study investigators and administrative structure

Role	Name	Responsibility
Co-Principal Investigator	De Wet Swanepoel	Study planning, oversight, data analyses, and reporting.
Co-Principal Investigator	David R. Moore	Study planning, oversight, data analyses, and reporting.
Co-Principal Investigator	Naveen Nagaraj	Study planning, oversight, data analyses, and reporting.
Co-investigator	Herman Myburgh	Study planning, technical oversight, data analyses, and reporting.
Co-investigator and Statistician	Marien A. Graham	Study planning, technical oversight, statistical analyses, and reporting.
Site Coordinator	Tersia de Kock	Study planning, technical oversight, data collection, analysis and reporting; site coordination for Khayelitsha.
Research Coordinator and Site Coordinator	Caitlin Frisby	Study planning, technical oversight, data analyses and reporting; site coordination for Atteridgeville.
Research Coordinator	Karina De Sousa	Study planning, technical oversight, data collection, data analyses, and reporting

3. Introduction

3.1 Overview of the study

Hearing loss is a leading cause of disability worldwide, affecting over 1.5 billion people in 2020, with more than 430 million experiencing disabling hearing loss (World Health Organization, 2021). The burden is disproportionately concentrated in low- and middle-income countries (LMICs), where 90% of individuals with moderate to profound hearing loss reside, yet access to hearing health services is severely limited (Mulwafu et al., 2017; WHO, 2021). In sub- Saharan Africa, hearing aid uptake is below 3% among those in need, highlighting a significant treatment gap (Bisgaard et al., 2021).

The World Health Organization has emphasized the use of digital health solutions and task-shifting strategies, such as deploying trained community health workers (CHWs), to improve hearing care access. Over-the-counter (OTC) and pre-set hearing aids, especially when combined with mobile health (mHealth) technologies, are increasingly recommended as scalable interventions suitable for LMICs (WHO, 2021; 2024).

This study is a two-arm randomized controlled trial evaluating whether a WhatsApp-delivered AI-enabled support and acclimatization program improves hearing aid self-efficacy compared with paper-based support among first-time hearing aid users in low-resource community settings in South Africa.

All participants will receive preset Go Ultra hearing aids fitted by trained community health workers (CHWs). Participants randomized to the experimental arm will receive a four-week WhatsApp-delivered AI-enabled support programme consisting of structured push content and on-demand access to an AI conversational support agent. Participants randomized to the control arm will receive equivalent core support information in a printed pamphlet at the time of fitting. Outcome assessment will occur five weeks post-fitting, one week after completion of the four-week intervention period.

3.2 Study design

This is an open-label, two-arm randomized controlled trial with blinded outcome assessment. The trial will recruit 80 participants to allow for attrition, with a target of 66 evaluable participants, corresponding to 33 participants per study arm.

Randomization will be implemented after eligibility is confirmed and baseline assessments are complete. Participants will be randomized 1:1 to:

- Experimental arm: WhatsApp-delivered AI-enabled support and acclimatization programme only.
- Control arm: paper-based support programme only.

The primary endpoint is the five-week follow-up visit. The primary outcome measure is the Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids (MARS-HA) global score at five weeks post-fitting.

3.3 Study interventions

Feature	Experimental arm	Control arm
Hearing aid fitting	Go Ultra preset hearing aids fitted by trained CHWs.	Go Ultra preset hearing aids fitted by trained CHWs.
Support programme	WhatsApp-delivered AI-enabled support programme with 16 structured voice notes/images over four weeks and on-demand AI conversational support.	Printed paper-based pamphlet with equivalent core content provided at fitting.

3.4 Objectives and endpoints

The primary objective is to determine whether the WhatsApp-delivered AI-enabled support programme results in superior hearing aid self-efficacy compared with paper-based support.

The primary endpoint is the difference in mean MARS-HA global score between the experimental and control arms at five weeks post-fitting.

Secondary outcomes include self-reported average daily hearing aid wear time, IOI-HA global and item scores, RHHI-S score, HASKI score, SNI score, device retention, device-related issues, and participant perceptions of the support programme content and delivery.

Endpoint	Description	Timing
The Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids (MARS-HA)	The Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids (West & Smith, 2007) is a 24-item self-report instrument measuring four factors: (1) Basic Handling, (2) Advanced Handling, (3) Adjustment, and (4) Aided Listening. Each item is rated on a 0–100% confidence scale in 10% increments (0% = no confidence at all, 100% = full confidence). The global score is the arithmetic mean of the 24 item responses and ranges from 0 to 100, with higher scores indicating greater hearing aid self-efficacy. Subscale scores are calculated as the mean of the items contributing to each factor. A global score of 80% or higher is conventionally interpreted as high self-efficacy; scores below 80% indicate low to moderate self-efficacy. The MARS-HA has been translated into isiXhosa and Sepedi for use in this study; both translations underwent forward-backward translation and pilot testing for cultural equivalence (Hall et al., 2018). Participants may complete the MARS-HA in	Five weeks post-fitting.

	English, isiXhosa, or Sepedi.	
International Outcome Inventory for Hearing Aids (IOI-HA)	The IOI-HA is a validated seven-item questionnaire to measure the effectiveness of the hearing aid intervention (Cox and Alexander, 2002). It targets seven domains including, (i) daily use, (ii) benefit, (iii) residual activity limitations, (iii) satisfaction, (iv) residual participation restrictions, (v) impact on others, and (vi) quality of life. Each item has five response choices, from worst to best outcome. The IOI-HA has been translated to the languages most commonly used in these communities namely isiXhosa and Sepedi. The participants will have the option to complete the IOI-HA in either of these languages or in English.	Five weeks post-fitting.
Daily wear time	Open-ended self-report of average daily hearing aid use.	Five weeks post-fitting.
Revised Hearing Handicap Inventory-Screening (RHHI-S)	The RHHI-S is a validated 10-item questionnaire that is a strong unidimensional, clinically informative measure of self-perceived hearing handicap that can be used by adults of all ages (Cassarly, Matthews, Simpson, & Dubno, 2020). Each question has three possible responses, including yes, sometimes, or no. The RHHI-S has been translated into the languages most commonly used in these communities, namely isiXhosa and Sepedi. The participants will have the option to complete the RHHI-S in either of these languages or in English.	Baseline and five weeks.
Hearing Aid Skills and Knowledge Inventory (HASKI)	The HASKI (Hearing Aid Skills and Knowledge Inventory) is a clinician-administered questionnaire designed to evaluate the practical skills and knowledge of	Five weeks post-fitting.

	hearing aid users. It covers key areas such as basic handling (insertion, removal, cleaning, and battery management), operational knowledge (using settings, adjusting controls, and troubleshooting), communication strategies (effective use in different environments), self-management (accessing resources and scheduling appointments), and care and maintenance (routine care and recognizing when professional servicing is needed). This tool helps audiologists identify areas where users may need additional training or support, ultimately improving hearing aid outcomes and user satisfaction. Scoring of the HASKI-clin is calculated by summing the total number of items identified as competent (not problematic), dividing this by the number of items for which a response was given (i.e., excluding the “Not Applicable responses), and multiplying by 100 to give a percentage of competency. Items are equally weighted. A higher overall score indicates greater competency (i.e., a score of 100% represents full competency).	
Berkman-Syme Social Network Index (SNI)	The Berkman-Syme Social Network Index (SNI) is a well-established tool used to assess the extent of an individual's social connections. The SNI examines the relationship between social networks and health outcomes. The SNI evaluates social integration by assessing four key domains: Marital status – Whether the individual is married or in a committed partnership. Contacts with close friends and relatives – Frequency of social interactions with family	Baseline and five weeks.

	and friends. Religious group membership – Whether the individual is part of a church or religious group and attends regularly. Participation in voluntary or community organizations – Involvement in social or civic groups outside of family and work. The SNI assigns points based on the individual's responses to the above components. Scores range from 0 to 4, with higher scores indicating greater social integration.	
Device retention/issues and participant perceptions	A non-standardized questionnaire will be included to obtain information from the participants on their perceptions of the support programme, as well as self-reported daily use in hours. This includes Likert scale and open-ended questions	Five weeks post-fitting.

4. Study population

4.1 Sample size determination and justification

The trial is designed to include 66 evaluable participants, with 33 participants per study arm. To allow for attrition, 80 participants will be recruited. The study is powered for a superiority comparison on the MARS-HA global score at five weeks post-fitting.

No minimum clinically important difference has been established for the MARS-HA in the context of an mHealth or AI-enabled intervention among first-time hearing aid users in low-resource settings. The study is therefore powered to detect a moderate standardized effect size of Cohen's d = 0.5. Using a conservative standard deviation estimate of 15 points, this corresponds to a between-group difference of 7.5 points on the MARS-HA global score.

4.2 Inclusion and exclusion criteria

The following criteria are summarized from the clinical investigation protocol and will be confirmed prior to randomization.

Inclusion criteria include:

- Adults aged 18 years or older, with no upper age limit.
- Bilateral mild-to-severe hearing loss, defined as pure-tone average 20 to <80 dB HL in both ears.
- Living in the participating low-resource community settings.
- Access to personal WhatsApp on a personal mobile phone, as required for eligibility in this study.
- Able and willing to provide informed consent and complete study procedures.

Exclusion criteria include:

- Ear or medical findings that require referral before hearing aid fitting, as determined by CHW otoscopy and supervising clinical review.
- Hearing loss profile or ear canal status that makes use of the study hearing aids inappropriate.
- Any condition that, in the opinion of the study team, would prevent safe participation or valid completion of study outcomes.

4.3 Analysis populations

The primary analysis population will be the intention-to-treat (ITT) population, defined as all randomized participants with an assigned study arm, analysed according to their randomized allocation where outcome data are available.

A per-protocol (PP) population will be used for sensitivity analyses. The PP population will include participants who completed the allocated intervention sufficiently to support interpretation of the intervention effect and who completed the five-week primary outcome assessment without major protocol deviations likely to affect the primary endpoint.

A safety population will include all randomized participants who received hearing aids and/or study support and will be analysed according to the intervention actually received.

5. Statistical analysis

5.1 General principles

Analyses will be conducted after database cleaning and locking. The statistical analysis will follow this SAP unless documented amendments are made before database lock. The primary comparison will be a superiority test comparing the experimental arm with the control arm on the MARS-HA global score at five weeks post-fitting.

Continuous outcomes will be summarized using means and standard deviations, as well as medians and interquartile ranges where appropriate. Categorical variables will be summarized using frequencies and percentages. Assumptions for parametric models will be assessed using residual plots and descriptive diagnostics. Where assumptions are materially violated, appropriate transformations, robust approaches, or non-parametric analyses will be considered and documented.

5.2 Withdrawals, dropouts and handling of missing data

Participant flow will be reported from screening through randomization, intervention delivery, follow-up, and analysis. Reasons for withdrawal, loss to follow-up, and missing outcome data will be summarized by randomized group.

The primary analysis will use available outcome data in the ITT population. If the proportion of missing primary outcome data is low, a complete case analysis will be used as the primary approach. If missingness is substantial or differs meaningfully by group, sensitivity analyses will be conducted to assess robustness of findings under different missing data assumptions. These may include multiple imputation using baseline covariates associated with missingness and outcome values, and/or best-worst or pattern-mixture sensitivity analyses.

For questionnaire scales, scoring will follow the relevant instrument scoring guidance. If item-level missingness occurs, prorating or scale calculation will only be applied where consistent with the instrument's scoring rules. Otherwise, the scale score will be treated as missing.

5.3 Level of statistical significance

The overall level of statistical significance for the primary superiority hypothesis will be set at 5% (alpha = 0.05), using a two-sided test. The primary effect estimate will be reported with a 95% confidence interval. Secondary and exploratory analyses will also use a 5% significance level but will be interpreted as supportive and hypothesis-generating. No formal multiplicity adjustment will be applied to secondary outcomes; the pattern, magnitude and precision of effects will be emphasized.

5.4 Statistical software

Analyses will be performed primarily using IBM SPSS Statistics, version 30.0 or later, and/or R. The software version and packages used will be documented in the final statistical report.

5.5 Participant characteristics and baseline comparability

Baseline characteristics will be presented by randomized group and overall. No formal hypothesis testing will be used to compare baseline characteristics between groups, as any differences arise after randomization and should be interpreted descriptively. Variables to be summarized include:

- Age
- Sex
- Study site
- Preferred language
- Years of self-reported hearing difficulty
- Ability to use a phone/WhatsApp
- State of readiness

- Otoscopy findings
- Pure-tone average and hearing loss severity category
- Baseline RHHI-S score
- Baseline SNI score

5.6 Primary endpoint analysis

5.6.1 Primary hypothesis

Null hypothesis (H0): There is no difference in mean MARS-HA global score at five weeks post-fitting between the experimental and control arms.

Alternative hypothesis (H1): Participants randomized to the WhatsApp-delivered AI-enabled support programme will demonstrate a higher mean MARS-HA global score at five weeks post-fitting than participants randomized to paper-based support.

5.6.2 Primary model

The primary outcome is the MARS-HA global score at five weeks post-fitting. The primary analysis will compare mean MARS-HA global scores between the experimental and control arms using a linear regression model with randomized group as the main fixed effect. The model will estimate the adjusted mean difference between groups and its 95% confidence interval.

The primary model will adjust for study site and baseline hearing loss severity category (mild, moderate, moderately-severe to severe), as these factors are clinically relevant and may influence outcome. If model convergence or sample-size constraints limit adjusted modelling, an unadjusted independent-samples t-test and corresponding mean difference will be reported alongside the adjusted model.

The MARS-HA global score will be treated as a continuous variable. Model residuals will be examined for approximate normality and homoscedasticity. If assumptions are not adequately met, a robust linear model or non-parametric Mann-Whitney U test will be used as a sensitivity analysis, with the primary interpretation still focused on the estimated between-group difference and confidence interval.

5.6.3 Effect size and interpretation

The between-group effect size will be reported as Cohen's d, calculated using the pooled standard deviation. The study is powered to detect a 7.5-point difference on the MARS-HA global score, corresponding to Cohen's d = 0.5 using a standard deviation of 15 points. Because no established MCID exists for this context, results will be interpreted using the observed point estimate, 95% confidence interval, effect size, and the powered detectable difference.

5.7 Secondary endpoint analyses

5.7.1 IOI-HA

The IOI-HA global score at five weeks will be analysed as a continuous outcome using linear regression with randomized group as the main fixed effect, adjusted for study site and hearing loss severity category where feasible. Individual IOI-HA items will be summarized descriptively and may be compared between groups using ordinal or non-parametric methods as appropriate.

5.7.2 Daily hearing aid use

Self-reported average daily hearing aid wear time will be summarized as a continuous variable in hours per day. Between-group comparisons will be conducted using linear regression or an independent-samples t-test if distributional assumptions are reasonable. If the distribution is skewed, a Mann-Whitney U test or categorized descriptive summaries will be used.

5.7.3 RHHI-S

RHHI-S will be collected at baseline and five weeks. The primary secondary analysis will compare follow-up RHHI-S scores between groups using analysis of covariance (ANCOVA), adjusting for baseline RHHI-S score, study site, and hearing loss severity category. Within-group change from baseline to five weeks will be summarized descriptively. Higher RHHI-S scores indicate greater self-perceived hearing handicap.

5.7.4 HASKI

HASKI total and/or domain scores at five weeks will be summarized by group and compared using linear regression or non-parametric methods depending on the scale distribution. Results will be interpreted as supportive evidence regarding practical hearing aid skills and knowledge.

5.7.5 SNI

SNI scores will be collected at baseline and five weeks. Between-group differences at follow-up will be assessed using ANCOVA adjusted for baseline SNI score, study site, and hearing loss severity where feasible. Changes in SNI over the short follow-up period will be interpreted cautiously given that social network integration may change slowly.

5.7.6 Device retention and device-related issues

Device retention at five weeks will be summarized as the number and proportion of participants retaining and continuing to use the study hearing aids. Device-related issues will be summarized by type and group using counts and percentages. Exploratory group comparisons may use chi-square or Fisher’s exact tests where appropriate.

5.7.7 Participant perceptions of support programme content and delivery

Likert-scale items on participant perceptions will be summarized using medians, interquartile ranges, frequencies, and percentages. Open-ended responses will be analysed descriptively using simple content categories or thematic summaries to contextualize the quantitative findings. These analyses will be considered exploratory.

5.8 Exploratory and sensitivity analyses

Exploratory subgroup analyses may examine whether the intervention effect differs by site, language, hearing loss severity, age group, baseline RHHI-S, and baseline SNI. These analyses will include interaction terms where sample size permits and will be interpreted as hypothesis-generating.

Engagement with the WhatsApp-delivered intervention will be described for the experimental arm, including receipt of structured messages, participant-initiated interactions with the AI agent, number and type of queries, and escalation events. Associations between engagement metrics and outcomes may be explored within the experimental arm but will not be interpreted as randomized comparisons because engagement occurs after randomization.

Sensitivity analyses for the primary outcome may include:

- Unadjusted between-group comparison of MARS-HA global score at five weeks.
- Per-protocol analysis excluding major protocol deviations.
- Robust or non-parametric analysis if distributional assumptions are violated.
- Missing data sensitivity analysis if primary outcome missingness is substantial or imbalanced between groups.

5.9 Adverse events and AI-related adverse events

Adverse events, adverse device effects, device deficiencies, and AI-related adverse events will be summarized descriptively by study arm. Counts and percentages will be reported by event category, severity, seriousness, relatedness to the intervention, and resolution status.

AI-related adverse events will be categorized according to the protocol-defined categories: missed safety escalation, out-of-scope clinical advice, inaccurate device guidance, inappropriate response to psychological distress, and breach of language or cultural appropriateness. Confirmed missed safety escalations and AI-related events resulting in actual or reasonably foreseeable harm will be reported according to the protocol-defined timelines.

6. Data management

6.1 Overview of data collection methods

Study data will be collected using paper-based and/or electronic case report forms and transcribed into secure study databases. Outcome data will be collected by trained study staff using the participant's preferred language where available. WhatsApp intervention engagement data, AI interaction logs, escalation logs, and audit findings will be exported from the relevant digital platforms where applicable and linked to participant IDs using de-identified study codes.

6.2 Data entry, cleaning and validation process

Data quality checks will include range checks, logic checks, missing data checks, duplicate participant checks, and verification of key variables against source documents where required. Queries will be resolved before database lock. Any changes to the dataset will be documented through an audit trail, including the date, responsible person, and reason for change.

6.3 Software and tools used for data management and analysis

Data may be managed in secure REDCap, Excel, Google Sheets, or equivalent password-protected study databases, depending on the operational workflow approved for the trial. Final analysis datasets will be exported to SPSS and/or R for analysis. De-identified analysis datasets and syntax files will be stored securely in accordance with institutional and sponsor requirements.

7. Reporting of results

7.1 Format for reporting outcomes

Results will be reported according to CONSORT principles for randomized controlled trials. Participant flow will be presented in a flow diagram showing screening, eligibility, randomization, intervention allocation, follow-up, and analysis. Primary and secondary outcomes will be reported as group-specific summary statistics, mean or median differences, 95% confidence intervals, p-values where applicable, and effect sizes where appropriate.

The primary conclusion will be based on the between-group difference in MARS-HA global score at five weeks. Interpretation will emphasize the magnitude and precision of the treatment effect rather than statistical significance alone.

7.2 Graphical presentation of results

Graphical displays may include:

- CONSORT participant flow diagram.
- Baseline characteristics table by study arm.

- Boxplots or violin plots of MARS-HA global score by study arm at five weeks.
- Forest plot of primary, secondary and exploratory effect estimates with 95% confidence intervals.
- Bar charts for categorical outcomes such as device retention, device issues, and adverse event categories.
- Engagement summaries for the experimental arm, including message completion, AI-agent use and escalation counts.

8. References

Bisgaard, N., Zimmer, S., Laureyns, M., & Groth, J. (2021). A model for estimating hearing aid coverage worldwide using historical data on hearing aid sales. International Journal of Audiology, 60(10), 761-769.

Cassarly, C., Matthews, L. J., Simpson, A. N., & Dubno, J. R. (2020). The Revised Hearing Handicap Inventory and screening version. Ear and Hearing.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. Academic Press.

Cox, R. M., & Alexander, G. C. (2002). The International Outcome Inventory for Hearing Aids (IOI-HA): Psychometric properties of the English version. International Journal of Audiology, 41(1), 30-35.

Ferguson, M. A., Brandreth, M., Brassington, W., & Wharrad, H. (2016). Information retention and self-efficacy in first-time hearing aid users.

World Health Organization. (2021). World report on hearing. Geneva: World Health Organization.

World Health Organization. (2024). Hearing aid service delivery approaches for low- and middle-income countries.