*D*igital th*E*rapy *F*or *I*mproved ti*N*nitus car*E* Study (DEFINE)

Statistical Analysis Plan

Version 1.0 from 2023-07-21

Sponsor: Oto Health Ltd
Author: Dr. Jan Multmeier, Dipl.-Psych.

Responsibilities

Sponsor	Oto Health Ltd 4th floor Silverstream House 45 Fitzroy Street London England W1T 6EB
Independent statistician	Dr. Jan Multmeier Bachstelzenweg 13A 16515 Oranienburg Germany

Content

Abbreviations

AE	Adverse event
CBT	Cognitive behavioural therapy
CC	Complete-cases
CG	Control group
CI	Confidence interval
HUI3	Health Utility Index Mark 3
IG	Intervention group
ITT	Intention-to-treat
MICE	Multiple imputation by chained equations
MRMM	Mixed-effects model for repeated measurements
PP	Per-protocol
Q-Q plot	Quantile-quantile plot
SAE	Serious adverse event
SD	Standard deviation
SUS	System usability scale
TFI	Tinnitus Functional Index

1 Background

1.1 Study objective

Tinnitus is a common condition, affecting approximately 15% of UK adults. It is usually perceived as a buzzing or ringing in the ears, without a stimulus in the outside world. There is currently no cure, but cognitive behavioural therapy (CBT) has been shown to be effective in reducing self reported distress and illness severity. Unfortunately, access to CBT is limited, with significant healthcare costs associated with provision and additional costs to people with tinnitus including time away from work, travel and lost productivity.

Oto is a novel multimodal smartphone app-delivered approach to tinnitus. It combines patient education, CBT, relaxation, mindfulness and sound therapy in a customisable package.

The primary objective of this study is to assess whether Oto's digital tinnitus programme is as effective at reducing self-reported tinnitus severity as therapist-delivered CBT.

1.2 Study design

The study will be conducted as a two-arm, randomized, controlled, open-label study. After inclusion in the study, patients will be randomly assigned in a 1:1 ratio to the intervention (IG) or active control (CG) group. Assignment will be stratified by age groups (≤40, 41-60, >60), gender (male, female), pure tone average threshold (None, slight, mild, moderate, moderately severe, severe, profound). IG participants will receive the Oto tinnitus programme (smartphone app delivered CBT). This includes an evaluation of symptoms and tinnitus severity, CBT, mindfulness and other interventions. CG participants will receive one to one therapist-delivered tinnitus therapy from an audiologist/hearing therapist.

In addition to the baseline assessment and the final follow-up after 12 months, further follow-ups are planned after 1, 3 and 6 months (see Figure 1).

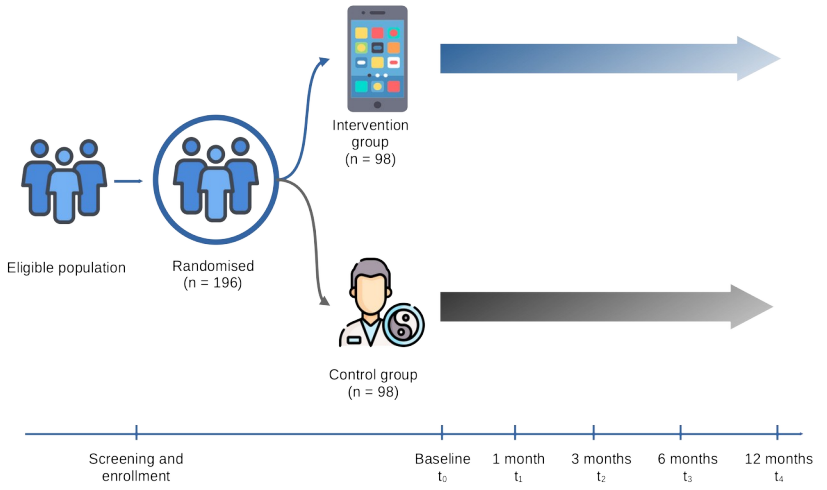

Figure 1: Study flow and follow-ups

2 Statistical methods and analytical strategy

The statistical analysis is based on the primary data collected in this study. When evaluating the primary and secondary endpoints, a significance level (α) of 5% is used. For multiple comparisons, an adjusted significance level according to the Bonferoni-Holm method applies. For inferential tests, an appropriate effect size measure is calculated in addition to the test statistic and *p*-value.

For inference tests, continuous dependent variables are examined for deviations from the normal distribution using Q-Q plots. If no clear deviations are apparent, parametric procedures are used to test for group differences (*t*-tests, linear regression models, linear mixed effect models for repeated measurements). If significant deviations from the normal distribution are apparent, the appropriate non-parametric procedures are used (Mann-Whitney U tests when comparing independent samples, Wilcoxon signed-rank tests when comparing paired samples, robust regression models, generalised estimation equations).

2.1 Study populations

Three study populations are constructed for statistical analyses:

1) the *Intention-to-treat* population (ITT),
2) the *Per-protocol* population (PP), as well as
3) the *Complete-cases* population (CC)

2.1.1 ITT

The ITT population includes all patients who were enrolled into the study, randomized into groups, and for whom baseline measurements are available. Missing data will be imputed for the analysis of primary and secondary endpoints.

2.1.2 PP

The PP population includes all patients of the ITT population for whom data on the primary endpoint are available at the measurement times t_0 and t_3, and who used Oto for at least 60 minutes in total (IG), or attended at least one session of therapist-delivered tinnitus therapy (CG).

2.1.3 CC

The CC population includes patients in the PP population who have complete data of the primary and secondary endpoints, patient characteristics, and covariates and instrumental variables at all measurement time points. Patients with missing values in these variables are not included.

2.2 Variables and Endpoints

Table 1 displays the measurement timepoints for each variable and describes how each endpoint is operationalised.

Table 1: Variables, endpoints and measurement timepoints

		Measurement time point

		Baseline	**1 month**	**3 months**	**6 months**	**12 months**
		t_0	t_1	t_2	t_3	t_4
Primary endpoints	**Measure/ Operationalistion**					
Non-inferiority of Oto compared to therapist-delivered tinnitus therapy regarding the change in TFI between baseline (t_0) and 6 months (t_3)	Difference in TFI score ($TFI_{t3} – TFI_{t0}$)	X			X	
Secondary endpoints						
Superiority of Oto compared to therapist-delivered tinnitus therapy regarding the change in TFI between baseline (t_0) and 6 months (t_3)	Difference in TFI score ($TFI_{t3} – TFI_{t0}$)	X			X	
Assess the self-reported impact of Oto on aspects of participants tinnitus experience compared to therapist-delivered tinnitus therapy	8 subscale scores of TFI: Intrusive, Sense of Control, Cognitive, Sleep, Auditory, Relaxation, Quality of Life and Emotional	X	X	X	X	X
Assess the impact on overall health-related quality of life of Oto compared to therapist-delivered tinnitus therapy	EuroQol EQ-5D-5L and Health Utilities Index Mark 3 (HUI3) scores will be converted into utility scores using value sets	X		X	X	X
Assess the usability of the Oto	System Usability Scale value			X		
Compare the number of adverse events between Oto and therapist-delivered tinnitus therapy	(S)AE collected over the course of the study	X	X	X		

Assess intervention Adherence	Number of participants completing the intervention	X		X	X	
Covariates and derived variables						
Pure tone average threshold category	Categorised pure tone average threshold (‘None’ <=15dB, ‘Slight’ 16-25dB, ‘Mild’ 26-40dB, ‘Moderate’ 41-55dB, ‘Moderately severe’ 56-70dB, ‘Severe’ 71-90dB, ‘Profound‘ 91dB+)	X				
Patient characteristics						
Age	Age categories (≤40, 41-60, >60)	X				
Sex		X				
Use of hearing aid					X	

2.3 Missing data

The number of missing outcome values for the primary endpoint (i.e., TFI scores at t_0 and t_3) is assessed. If more than 10% of TFI change scores cannot be determined due to missing data, missing values in outcome variables, covariates, or derived variables are imputed using the *Multiple Imputation by Chained Equations* (MICE) technique (Buuren and Groothuis-Oudshoorn 2011). The imputed values will then be used in sensitivity analyses to assess the effect of missing data on the primary endpoint.

The number of missing values in the outcome variables, covariates, or instrumental variables is reported at each follow-up point.

2.4 Statistical Methods

2.4.1 Descriptive Statistics

Descriptive statistical measures are used to summarize the endpoints, covariates, and instrumental variables at the follow-up points. The frequency of missing values per variable and study time point is reported, but missing values are not replaced.

For categorical variables, absolute and relative frequencies are reported. For ordinal variables, median and quartiles are used, for metric variables, arithmetic means and standard deviations are

reported additionally. The distribution of the outcome variables is also displayed via boxplots or equivalent graphical procedures for each group and follow-up point. Measures of central tendency are displayed with variability parameters (e.g., arithmetic means with standard deviation/standard error/confidence interval).

2.4.2 Primary endpoint

The primary endpoint is **non-inferiority** of Oto compared to therapist-delivered tinnitus therapy in the change of TFI score from t_0 to $t_{3:}$

$$TFI\ change\ score = TFI_{t_3} - TFI_{t_0}$$

Non-inferiority will be assessed by constructing a 95% confidence interval (CI) around the difference in TFI change scores between the IG and CG group. If the lower bound of the CI does not include the pre-specified non-inferiority margin of 13 points on the TFI scale, Oto will be determined to be non-inferior to therapist-delivered tinnitus therapy.

The analysis of the primary endpoints is carried out using the ITT population, using only observed outcome variables. In case the TFI change score cannot be determined for more than 10% of subjects, a sensitivity analysis will examine the effect of missing data on the primary endpoint.

2.4.3 Secondary endpoints

The study contains 6 secondary endpoints, which are described below.

Superiority of Oto compared to therapist-delivered tinnitus therapy regarding the change in TFI between baseline (t_0) and 6 months (t_3)

Like for the analysis of the primary endpoint, the TFI change score between baseline (t_0) and 6 months (t_3) is compared between the IG and CG, this time using a linear regression model, controlling for the stratification variables:

Population:	ITT
Dependent variable:	TFI change score
Independent variables:	• Group (IG vs. CG) • Age group (≤40, 41-60, >60) • Sex (male, female) • Pure tone average threshold category (None, slight, mild, moderate, moderately severe, severe, profound)

Assess the self-reported impact of Oto on aspects of participants tinnitus experience compared to therapist-delivered tinnitus therapy

Each of the 8 subscale scores of the TFI assessing one aspect of tinnitus experience (Intrusive, Sense of Control, Cognitive, Sleep, Auditory, Relaxation, Quality of Life and Emotional) will be modelled using a linear mixed-effects regression model for repeated measurements (MRMM) with a random effect for the subject, controlling for the stratification variables:

Population:	ITT
Dependent variable:	TFI subscale score (Intrusive, Sense of Control, Cognitive, Sleep, Auditory, Relaxation, Quality of Life and Emotional)
Independent variables:	• Group (IG vs. CG) • Age group (≤40, 41-60, >60) • Sex (male, female) • Pure tone average threshold category (None, slight, mild, moderate, moderately severe, severe, profound)

Assess the impact on overall health-related quality of life of Oto compared to therapist-delivered tinnitus therapy

Health-related quality of life is assessed through the EUROQOL EQ-5D-5L and Health Utilities Index Mark 3 (HUI3) instruments. Scores will be converted into utility scores using respective value sets. For each instrument, the utility score value will be modelled using a linear generalised linear mixed-effects model with a random effect for the subject, controlling for the stratification variables.

Population:	ITT
Dependent variable:	TFI subscale score (Intrusive, Sense of Control, Cognitive, Sleep, Auditory, Relaxation, Quality of Life and Emotional)
Independent variables:	• Group (IG vs. CG) • Age group (≤40, 41-60, >60) • Sex (male, female) • Pure tone average threshold category (None, slight, mild, moderate, moderately severe, severe, profound)

Assess the usability of the Oto

The usability of Oto will be assessed through the System Usability Scale (SUS) value, assessed 3 months after starting the intervention. The SUS assesses usability using 10 items, each answered on a 5-point rating scale ranging from “Strongly agree” to “Strongly disagree”, scored as 4 through 0. The final score is computed by summing all responses and multiplying by 2.5, which normalises the scale from 0 to 100. As the SUS will only be used in the IG, and only collected once, descriptive analyses are conducted to assess the distribution of values and their central tendency and dispersion. Mean and median, as well as standard deviation, minimum, maximum, and quantiles are reported.

Compare the number of adverse events between Oto and therapist-delivered tinnitus therapy

The number of (S)AEs reported within the first three months after randomisation are reported for each group, per (S)AE category and in total. The total number of (S)AEs will be compared between groups using a poisson regression model, controlling for the stratification variables:

Population:	ITT
Dependent variable:	Number of SAE and AE reported within 3 months
Independent variables:	• Group (IG vs. CG) • Age group (≤40, 41-60, >60) • Sex (male, female) • Pure tone average threshold category (None, slight, mild, moderate, moderately severe, severe, profound)

As we assume that all (S)AE are observed within the study follow-up, the observed data are used without imputation.

Assess intervention adherence

Adherence to the intervention will be assessed by comparing the proportion of participants completing the intervention between baseline (t_0) and 1 month (t_1), 1 month (t_1) and 3 months (t_2), and 3 months (t_2) and 6 months (t_3) between IG and CG. Completion is defined as observing no further interaction with the Oto app (IG), or no further in-person sessions with a therapist (CG). Cumulative completion rates for each interval are compared between groups using χ^2-tests.

2.4.4 Sensitivity analyses

Sensitivity analyses are performed to determine whether assumptions or specifications made in the preparation and analysis of the data had an influence on the effects of the primary and secondary endpoints. In this study, this concerns the choice of the analysis population and the choice of covariates. Accordingly, the analyses for the primary and secondary endpoints will be repeated with different input parameters. The planned sensitivity analyses are summarized in Table 2.

Table 2: Summary of planned sensitivity analyses

No.	Endpoint	Variable	Population	Covariates
1.1	PEP	Difference in TFI score, including imputed values ($TFI_{t3} – TFI_{t0}$) [non-inferiority]	ITT	See main analysis
1.2	PEP	Difference in TFI score ($TFI_{t3} – TFI_{t0}$) [non-inferiority]	PP	See main analysis
1.3	PEP	Difference in TFI score ($TFI_{t3} – TFI_{t0}$) [non-inferiority]	CC	See main analysis
2.1	SEP	Difference in TFI score ($TFI_{t3} – TFI_{t0}$) [superiority]	PP	See main analysis
2.2	SEP	Difference in TFI score ($TFI_{t3} – TFI_{t0}$) [superiority]	CC	See main analysis
3.1	SEP	8 subscale scores of TFI: Intrusive, Sense of Control, Cognitive, Sleep, Auditory,	PP	See main analysis

		Relaxation, Quality of Life and Emotional		
3.2	SEP	8 subscale scores of TFI: Intrusive, Sense of Control, Cognitive, Sleep, Auditory, Relaxation, Quality of Life and Emotional	CC	See main analysis
4.1	SEP	EuroQol EQ-5D-5L and Health Utilities Index Mark 3 (HUI3) scores will be converted into utility scores using value sets	PP	See main analysis
4.2	SEP	EuroQol EQ-5D-5L and Health Utilities Index Mark 3 (HUI3) scores will be converted into utility scores using value sets	CC	See main analysis

2.4.5 Subgroup analyses

Subgroup analyses will examine whether effects differ descriptively between predefined groups. In this study, analyses of the primary and secondary endpoints are planned separately for the following groups:

- Sex, according to stratification

2.5 Sample size determination

2.5.1 Approach

The mean TFI score after 6 months will be compared between the groups. Oto will be judged as equally effective as therapist-delivered tinnitus therapy if the difference between the groups is smaller than a pre-defined non-inferiority margin (δ). Specifically, CBT and app-delivered therapy will be considered equally effective if the lower bound of the 95%-confidence-interval (two-sided) around the difference in means is larger than δ.

Hypotheses

The following hypotheses are tested:

H_0: $TFI_{CG} - TFI_{IG} > \delta$

H_1: $TFI_{CG} - TFI_{IG} \leq \delta$,

where TFI_{CG} is the mean TFI score of the CG, TFI_{IG} is the mean TFI score of the IG, and δ is the non-inferiority margin (see below).

Assumptions

Medical and statistical judgement is needed to define the appropriate non-inferiority margin (δ). Two previous studies (Beukes et al. 2017; Rademaker et al. 2020) have used a non-inferiority margin of 13 points on the TFI scale to demonstrate non-inferiority of different interventions and CBT. This value corresponds to the minimum clinically important difference defined by Meikle et al. (2012) and will also be used for the sample size calculation of this trial.

Beukes et al. (2018) reported a standard deviation (SD) of 18.4 for the TFI score overall, but larger SDs of up to 31.5 points have been reported for the individual subscales (Wrzosek et al. 2016). To account for a potentially larger variation in TFI scores in the present trial, a conservative SD of 31.0 will be used.

Furthermore, the significance level (α) will be set to 5%, and the desired power (1-β) is set to 90%.

2.5.2 Sample size calculation

The sample size is calculated using the package `epiR`, version 2.0.58 in `R` version 4.2.2.
The resulting total sample size is **196**, with 96 subjects in the IG and 96 in the CG.

3 Protocol deviations

Deviations from the statistical analysis plan for the primary and secondary endpoints are not anticipated. All unanticipated changes will be documented as amendments to this document.

4 Data management

The data collected as part of the clinical study are pseudonymized and stored in a database that is technically secured against unauthorized access. The chief investigator, clinical team and other authorised members of the trial team will have access to records.

5 Software

Data analysis is performed using the software *R* in Version 4.2.2 or later.

6 References

Beukes, Eldré W., David M. Baguley, Peter M. Allen, Vinaya Manchaiah, and Gerhard Andersson. 2017. ‘Guided Internet-Based versus Face-to-Face Clinical Care in the Management of Tinnitus: Study Protocol for a Multi-Centre Randomised Controlled Trial’. *Trials* 18 (1): 186. https://doi.org/10.1186/s13063-017-1931-6.

———. 2018. ‘Audiologist-Guided Internet-Based Cognitive Behavior Therapy for Adults With Tinnitus in the United Kingdom: A Randomized Controlled Trial’. *Ear and Hearing* 39 (3): 423–33. https://doi.org/10.1097/AUD.0000000000000505.

Buuren, Stef van, and Karin Groothuis-Oudshoorn. 2011. ‘Mice: Multivariate Imputation by Chained Equations in R’. *Journal of Statistical Software* 45 (December): 1–67. https://doi.org/10.18637/jss.v045.i03.

Meikle, Mary B., James A. Henry, Susan E. Griest, Barbara J. Stewart, Harvey B. Abrams, Rachel McArdle, Paula J. Myers, et al. 2012. ‘The Tinnitus Functional Index: Development of a

New Clinical Measure for Chronic, Intrusive Tinnitus'. *Ear and Hearing* 33 (2): 153–76. https://doi.org/10.1097/AUD.0b013e31822f67c0.

Rademaker, Maaike Maartje, Inge Stegeman, Arno Lieftink, Metten Somers, Robert Stokroos, and Adriana L. Smit. 2020. 'MinT-Trial: Mindfulness versus Cognitive Behavioural Therapy in Tinnitus Patients: Protocol for a Randomised Controlled, Non-Inferiority Trial'. *BMJ Open* 10 (2): e033210. https://doi.org/10.1136/bmjopen-2019-033210.

Wrzosek, Małgorzata, Eugeniusz Szymiec, Wiesława Klemens, Piotr Kotyło, Winfried Schlee, Małgorzata Modrzyńska, Agnieszka Lang-Małecka, Anna Preis, and Jan Bulla. 2016. 'Polish Translation and Validation of the Tinnitus Handicap Inventory and the Tinnitus Functional Index'. *Frontiers in Psychology* 7. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01871.