

Towards Understanding Tinnitus Symptom Dynamics During Treatment



Dissertation
zur Erlangung des Doktorgrades
der Humanwissenschaften
(Dr. sc. hum.)

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Milena Engelke
aus
Augsburg

im Jahr
2025

Towards Understanding Tinnitus Symptom Dynamics During Treatment



Dissertation
zur Erlangung des Doktorgrades
der Humanwissenschaften
(Dr. sc. hum.)

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Milena Engelke
aus
Augsburg

im Jahr
2025

Dekan: Prof. Dr. Dirk Hellwig

Betreuer: Prof. Dr. Winfried Schlee

Erstgutachter: Prof. Dr. Winfried Schlee

Zweitgutachter: PD Dr. Stefan Schoisswohl

Drittgutachter: Prof. Dr. Dirk De Ridder

Datum der mündlichen Prüfung: 06.10.2026

Table of Contents

ABSTRACT	5
ZUSAMMENFASSUNG	6
GLOSSARY OF TERMS.....	7
INTRODUCTION.....	8
TINNITUS.....	8
<i>Definition</i>	<i>8</i>
<i>Phenotypes</i>	<i>9</i>
<i>Prevalence</i>	<i>9</i>
<i>Pathophysiology</i>	<i>9</i>
<i>Diagnostics and Measurement</i>	<i>10</i>
<i>Treatment.....</i>	<i>11</i>
ECOLOGICAL MOMENTARY ASSESSMENT.....	11
<i>Development and Definition</i>	<i>12</i>
<i>Current Foci in EMA health research</i>	<i>13</i>
<i>Limitations.....</i>	<i>14</i>
ECOLOGICAL MOMENTARY ASSESSMENT IN TINNITUS	15
<i>Feasibility.....</i>	<i>16</i>
<i>Retrospective Comparison</i>	<i>16</i>
<i>Symptom Fluctuations and Interactions</i>	<i>16</i>
<i>Machine Learning</i>	<i>17</i>
<i>Software.....</i>	<i>18</i>
<i>Outcome Research.....</i>	<i>18</i>
SCOPE OF THIS THESIS.....	19
STUDY I: PILOT STUDY OF A SMARTPHONE-BASED TINNITUS THERAPY USING STRUCTURED COUNSELING AND SOUND THERAPY: A MULTIPLE-BASELINE DESIGN WITH ECOLOGICAL MOMENTARY ASSESSMENT.....	20
INTRODUCTION	21
METHODS	23
RESULTS.....	28
DISCUSSION.....	37
STUDY II: UNDERSTANDING TINNITUS SYMPTOM DYNAMICS AND CLINICAL IMPROVEMENT THROUGH INTENSIVE LONGITUDINAL DATA.....	43
INTRODUCTION	44
METHODS	48
RESULTS.....	51
DISCUSSION.....	55

STUDY III: UNCOVERING LATENT TRAJECTORIES OF DAILY TINNITUS SYMPTOMS THROUGH APP-BASED MONITORING DURING TREATMENT	59
INTRODUCTION	60
METHODS	62
RESULTS.....	65
DISCUSSION	72
DISCUSSION	76
DECOUPLING BETWEEN TINNITUS DISTRESS AND LOUDNESS: INDICATOR OF TREATMENT RESPONSE?	76
TINNITUS SYMPTOM NETWORK	77
EMA VS. PROMS: INTERCHANGEABLE OR COMPLEMENTARY?.....	79
DOES COMPLIANCE LIMIT FEASIBILITY?	80
ADDRESSING WITHIN- AND BETWEEN-TINNITUS HETEROGENEITY THROUGH METHODOLOGICAL APPROACHES	81
FINDING THE SWEET SPOT BETWEEN IDIOGRAPHIC AND NOMOTHETIC APPROACHES.....	82
LIMITATIONS.....	83
FUTURE DIRECTIONS	83
CONCLUSION.....	85
REFERENCES	86
LIST OF FIGURES.....	106
LIST OF TABLES	106
SUPPLEMENTARY MATERIAL	107
ACKNOWLEDGEMENTS.....	120
AFFIDAVIT	121

Note 1: The author's own publications and co-authored works are indicated in the footnotes.

Note 2: Chapters *Ecological Momentary Assessment* and *Ecological Momentary Assessment in Tinnitus* draw upon material previously published (1).

Abstract

Tinnitus is defined as the conscious perception of a tonal or composite noise for which there is no identifiable corresponding external sound source. In its severe form, also referred to as tinnitus disorder, it is accompanied by emotional distress, cognitive dysfunction, and autonomic arousal. While the percept itself affects approximately 14.4% of adults, severe tinnitus is present in about 2.3% of adults and is associated with reduced quality of life, fear, constant awareness, annoyance, concentration difficulties, and mental diseases. Yet, the substantial heterogeneity in tinnitus phenotypes and aetiologies hinders the development of effective treatments. Promising treatment outcomes often lack replication, and in many clinical trials, only a subset of patients respond.

Current clinical tinnitus research largely relies on aggregated data collected before and after treatment providing little insight into the temporal dynamics of symptom change. The ubiquitous use of smartphones makes them well suited for regular symptom monitoring during treatment, which might shed light on unseen symptomatic patterns or dynamics during treatment, which could help to explain treatment heterogeneity. To date, the potential of methods such as Ecological Momentary Assessment (EMA) remains underexplored in clinical tinnitus research.

In this thesis, I explore the potential of daily symptom monitoring across different tinnitus treatments to investigate symptom changes and dynamics. The introduction first provides a clinical overview of tinnitus, followed by an introduction to EMA, its methodological foundations, and its current applications in health research and tinnitus studies. The first study examines different analytical methods for quantifying treatment effects using EMA data and compares them with established patient-reported outcome measures. The second study investigates in greater detail how clinical improvement manifests in EMA-derived symptoms and whether patterns of symptom fluctuation can predict treatment effects. The third study addresses heterogeneity in symptom trajectories using latent class analysis. In the final section the results are discussed across shared conceptual and methodological themes such as symptom dynamics, methodological innovations in EMA analysis, and the balance between idiographic and nomothetic approaches.

Zusammenfassung

Tinnitus wird definiert als bewusste Wahrnehmung eines tonalen oder zusammengesetzten Geräuschs ohne identifizierbare äußere Schallquelle. Während etwa 14,4% der Erwachsenen einen Tinnitus wahrnehmen, leiden rund 2,3% an einer schweren Form, die mit einer verminderten Lebensqualität, Ängsten, Konzentrationsschwierigkeiten und psychischen Erkrankungen einhergehen kann. Die Heterogenität der Tinnitus-Phänotypen und Ursachen erschwert die Entwicklung wirksamer Behandlungen; in vielen klinischen Studien spricht nur ein Teil der PatientInnen auf die Behandlung an.

Die aktuelle klinische Tinnitusforschung stützt sich weitgehend auf aggregierte Daten, die vor und nach einer Behandlung erhoben werden, und liefert daher nur begrenzt Einblicke in die zeitliche Dynamik von Symptomveränderungen. Die weit verbreitete Nutzung von Smartphones ermöglicht die regelmäßige Erfassung von Symptomen mittels Ecological Momentary Assessment (EMA), dessen Potenzial in der klinischen Tinnitusforschung bislang weitgehend ungenutzt bleibt.

Die vorliegende Dissertation untersucht das Potenzial täglicher Datenerhebungen zur Erfassung von Veränderungen und Dynamiken von Tinnitussymptomen während einer Behandlung. Die erste Studie pilotiert verschiedene analytische Methoden zur Quantifizierung von Behandlungseffekten anhand von EMA-Daten und vergleicht diese mit den Ergebnissen etablierter Messinstrumente. Die zweite Studie entwickelt einen methodischen Ansatz zur Vorhersage von Behandlungserfolg auf Basis von EMA-Daten. Mithilfe einer Feature-basierten Zeitreihenanalyse und regularisierter Regressionsmodelle werden Merkmale identifiziert, die klinische Verbesserung vorhersagen. Die dritte Studie befasst sich mit der Heterogenität von Symptomverläufen mithilfe einer latenten Klassenanalyse längsschnittlicher Daten. Abschließend werden die Ergebnisse entlang gemeinsamer konzeptueller und methodischer Themen wie Symptodynamiken, methodische Innovationen, personalisierter Tinnitusbehandlung und dem Spannungsfeld zwischen idiographischen und nomothetischen Forschungsansätzen diskutiert.

Glossary of terms

AIC	Akaike Information Criterion
BFI-2	Big Five Inventory-2
BIC	Bayesian Information Criterion
CBT	Cognitive Behavioural Therapy
CDSS	Clinical Decision-Support System
CGI-I	Clinical Global Impression Scale – Improvement
CI	Confidence Interval
EDD	End-of-Day Diary
EMA	Ecological Momentary Assessment
ESIT-SQ	European School of Interdisciplinary Tinnitus Research Screening Questionnaire
GMM	Growth Mixture Model
GUEF	German Hyperacusis Questionnaire
HA	Hearing Aid
HRV	Heart Rate Variability
JITAI	Just-in-Time Adaptive Intervention
MAR	Missing at random
MNAR	Missing not at random
MCID	Minimal Clinically Important Difference
PHQ-9	Patient Health Questionnaire
PPM	Posterior Probability of Membership
PROM	Patient-Reported Outcome Measure
RCT	Randomised Clinical Trial
SC	Structured Counseling
SCED	Single-Case Experimental Design
SoN	Shades of Noise
ST	Sound Therapy
TFI	Tinnitus Functional Index
THI	Tinnitus Handicap Inventory
TQ	Tinnitus Questionnaire
UNITI	Unification of Treatments and Interventions for Tinnitus Patients
VAS	Visual Analogue Scale

Introduction

“I know I should be more idiographic in my research, but I just don’t know how to be,” psychologist Edward Tolman once remarked to his colleague Gordon Allport (2). This statement remains strikingly relevant today, as psychological science continues to rely predominantly on aggregated data from groups of individuals (3). The proportion of studies employing repeated-measures designs even decreased in recent years (4), a trend that may reflect the strong emphasis on group-comparison designs in the training of empirical researchers.

While nomothetic approaches aim to derive generalisable principles from group-level analyses, idiographic approaches focus on understanding individual trajectories and within-subject variability (3). Although nomothetic methods are valuable for identifying general patterns, they often overlook the dynamic fluctuations and interdependencies that occur within individuals. It has been shown that variance estimates derived from interindividual designs tend to underestimate intraindividual variability (5).

Tinnitus, a condition characterised by substantial heterogeneity, is particularly well suited for longitudinal investigation; yet such designs remain scarce. This thesis analyses intensive longitudinal data to explore the balance between idiographic and nomothetic perspectives: acknowledging intraindividual variability while still enabling inferences at the (sub)group level. In doing so, this work aims to broaden the methodological perspective in tinnitus research and to encourage researchers to move beyond traditional cross-sectional designs towards innovative approaches for capturing and analysing intensive longitudinal data.

Tinnitus

Definition

The term *tinnitus* is derived from the Latin word *tinnire* which means *to ring*. Current definitions vary from simplistic (“ringing of the ears”) to complex descriptions (“the conscious perception of an auditory sensation in the absence of a corresponding external stimulus”), that are either omitting important aspects such as perception in the head or include distinct phenomena, e.g. auditory hallucinations (6). De Ridder et al. distinguish between the *tinnitus percept* defined as “conscious awareness of a tonal or

composite noise for which there is no identifiable corresponding external sound source” and *tinnitus disorder*¹ defined as tinnitus “associated with emotional and/or cognitive dysfunction, and/or autonomic arousal, leading to behavioural changes and functional disability” (7).

Phenotypes

The challenge in defining tinnitus stems at least partly from its heterogeneity. It can manifest as ringing, buzzing, or hissing, which can be tonal or noise-like of varying frequencies and intensities. Tinnitus can be perceived in one ear, both ears, or inside the head. Temporal patterns also vary; tinnitus can be constant, intermittent, occasional, rhythmical, or pulsatile, as well as fluctuating or changing over time. The degree to which individuals are affected by their tinnitus ranges from not at all to severe (8). Patients report reduced quality of life, fear, constant awareness, annoyance, and concentration difficulties as the most common reasons for their suffering (9). Severe tinnitus is often accompanied by psychiatric disorders such as anxiety, depression, and insomnia (10). Objective tinnitus can be delineated from subjective tinnitus by the existence of an internal sound source of muscular or vascular origin (11). Acute and chronic tinnitus are typically separated by a 3 or 6 month threshold, though empirical support for a cut-off is limited (6).

Prevalence

Tinnitus is a common condition affecting 14.4% (95% CI: 12.6% – 16.5%) of adults and 13.6% (95% CI: 8.5% – 21.0%) of children and adolescents according to a recent meta-analysis (12). Severe tinnitus occurs in about 2.3% (95% CI: 1.7% – 3.1%) of adults and 2.7% (95% CI: 0.8% – 8.4%) of children and adolescents. Within the adult population, prevalence of tinnitus and tinnitus disorder increases with age (12).

Pathophysiology

There are many pathophysiological models of tinnitus, none of which fully account for all cases, reflecting its heterogeneity (13). Yet, tinnitus is often linked to sensorineural hearing loss, which can be triggered by aging processes or noise trauma (14). It is hypothesized that sensory deprivation associated with damage to the auditory periphery leads to maladaptive plasticity, characterised by increased spontaneous firing rates and enhanced neural synchrony, which might generate the tinnitus percept.

¹ In this thesis, tinnitus disorder and severe tinnitus are used synonymously.

Introduction

Abnormal neuronal activity has been observed throughout the auditory pathway including the cochlear nucleus, inferior colliculus, auditory thalamus, and primary auditory cortex (15,16), as well as in non-auditory regions like anterior cingulate cortex, anterior insula, amygdala, hippocampus, parahippocampus, posterior cingulate cortex, precuneus, parietal cortex, and prefrontal cortex forming networks responsible for salience, emotional distress, memory, and attention (17,18). More recently, an imbalance between excitatory and inhibitory pathways has been postulated (19). In patients with tinnitus but normal audiograms, reduced cochlear activity has been observed, providing evidence for hidden hearing loss, a condition in which conventional audiometry fails to detect underlying auditory dysfunction (20). Next to otological factors, temporo-mandibular joint disorder is considered a risk factor for developing tinnitus via somatosensory-auditory integration (14,16).

Diagnostics and Measurement

Tinnitus diagnostic is mainly based on patient self-report. Although neurophysiological research has identified candidate biomarkers, none have been established for clinical use (21,22). In most clinical cases, the tinnitus is considered idiopathic with unknown aetiology (23). However, due to some comorbidities requiring urgent treatment, such as severe neurological and psychiatric conditions, comprehensive individual assessment is essential (24). Standard diagnostic procedures include a detailed examination of tinnitus history, tinnitus severity and potential psychiatric comorbidities, otological examinations, audiometric assessments and psychoacoustic measurements of tinnitus' perceptual characteristics (23–25). Further diagnostic steps should be taken depending on clinical indications such as brain imaging in case of asymmetric hearing loss, unilateral or pulsatile tinnitus (23).

Tinnitus severity, the impact of tinnitus on the individual, is typically quantified using Patient-Reported Outcome Measures (PROM). These tools are employed both in clinical settings to determine treatment necessity and in research to assess intervention efficacy (26). There is a variety of psychometrically validated PROMs (27), with the most commonly used being the Tinnitus Handicap Inventory (THI), the Tinnitus Questionnaire (TQ) and the Tinnitus Functional Index (TFI) (26,28–31). Minimal Clinically Important Difference (MCID) scores have been defined for these instruments

to determine whether score changes are clinically meaningful² (1,32). They are considered the gold standard for proof of efficacy in clinical trials due to their wide use, strong psychometric properties, and availability in validated translations (26,31). However, limitations include limited comparability across instruments³, potential bias due to retrospective assessments, and insufficient sensitivity to longitudinal symptom fluctuations during treatment (33,34).

Treatment

From the patients' perspective, current treatment options are often unsatisfactory as no treatment consistently removes the tinnitus percept or reduces its loudness (35). Consequently, treatment generally targets associated distress or comorbid conditions. For tinnitus disorder, Cognitive Behavioural Therapy (CBT), which aims to modify maladaptive thoughts and behaviours related to tinnitus, has demonstrated the strongest evidence in reducing tinnitus-associated distress (23,36,37). Other therapies, including hearing aids (HA), sound therapy (ST), neuromodulation, or drug treatment, have produced mixed results (36,38). A multicentre trial comparing various single and combined treatment modalities in 461 tinnitus patients found that CBT (Cohens $d = 0.93$, 95%-CI: 0.70 - 1.21), structured counseling (SC; Cohens $d = 0.83$, 95%-CI: 0.51 - 1.27) and HA (Cohens $d = 1.00$, 95%-CI: 0.78 - 1.28) were similarly effective in reducing tinnitus distress, while ST (Cohens $d = 0.24$, 95%-CI: -0.02 – 0.53) had a smaller effect⁴ (39). The variability in outcomes may reflect differential treatment needs among patients, yet predictive factors remain largely unknown (40). Importantly, all current tinnitus treatments are considered off-label (41).

Ecological Momentary Assessment

Given the heterogeneity of tinnitus and the variability in treatment outcomes, understanding how symptoms and distress evolve during daily life and treatment becomes crucial. Capturing these dynamics requires assessment methods that go

² Engelke M, Basso L, Langguth B, Zeman F, Schlee W, Schoisswohl S, et al. Estimation of minimal clinically important difference for Tinnitus Handicap Inventory and Tinnitus Functional Index. *Otolaryngol Head Neck Surg.* 2025;173(1):69–79.

³ Engelke M, Simões JP, Langguth B, Schlee W, Basso L. The 83 symptoms of tinnitus: content overlap of commonly used scales for tinnitus burden. *Under review.*

⁴ Schoisswohl S, Basso L, Simões J, Engelke M, Langguth B, Mazurek B, et al. Single versus combination treatment in tinnitus: an international, multicentre, parallel-arm, superiority, randomised controlled trial. *Nat Commun.* In press.

beyond retrospective questionnaires, which can be offered by Ecological Momentary Assessment (EMA).

Development and Definition

The momentary assessment of human behaviour and emotion outside the laboratories has a long history. One of the first pioneers was J.C. Flügel, who studied the ratio of positive and negative experience in human life in the 1920s (42). In his study, participants kept frequent records of their mental state in daily life, resulting in more pleasant than unpleasant experiences. In 1977, Csikszentmihalyi et al. were interested in the interaction of adolescents' daily behaviour and feelings (43). Adolescents reported their activities and emotions in response to random prompts of an electronic paging device. Talking to peers was associated with happiness, whereas watching TV was a result of a lack of alternatives.

This early study helped establish the Experience Sampling Method (ESM), later refined by Csikszentmihalyi (44). It captures individuals' thoughts, feelings, and behaviour in real-time through random prompts and laid the groundwork for Csikszentmihalyi's well-known research on the flow concept. ESM data of daily activities showed that engaging in intrinsically motivated activities, well-balanced between being challenging but not overwhelming, increased the probability of experiencing flow, a state of deep engagement (45).

In the 1990s, Stone and Shiffman established the concept of EMA, which they defined as "monitoring or sampling strategies to assess phenomena at the moment they occur in natural settings" (46). The ESM methodology was extended by allowing for different sampling schedules in addition to random prompts. Also, while ESM was hitherto used in psychological research to study the dynamics of daily life experiences and subjective well-being in the general population, Stone and Shiffman adapted it to the clinical setting to study health-related behaviour and symptoms (47). Today, ESM and EMA are mostly used synonymously.

EMA aims to minimise recall bias by capturing experiences in the moment and involves repeated assessment of individuals in their natural environment, increasing ecological validity (46,47). Over time, EMA has developed into a flexible method, varying in timing, technology, subjects, and topics, yet always grounded in repeated, near real-time data collection in subjects' daily life circumstances (34). With the widespread integration of smartphones into everyday life, EMA research has significantly expanded in reach, usability, and feasibility (48). Today, EMA is widely

applied in eHealth research to study symptom fluctuations and can be combined with passive forms of ambulatory assessment, such as the continuous monitoring of objective data via device sensors (49).

Current Foci in EMA health research

Even if current EMA health research foci are widespread, the following section aims to provide insights into the potentials and challenges, yet without being an exhaustive compilation.

EMA data enables fine-grained evaluation of fluctuations by frequent and/or long-term repeated sampling which allows to study both short-term as well as long-term patterns of symptoms. The “Blue Sunday Effect”, suggesting that people feel worse on Sundays, was specified in that positive mood increased until noon, just like on the other days, yet in the afternoon, mood stagnated on Sundays and increased further on the other days (50). A meta-analysis investigated the association of emotion dynamics on different timescales with well-being and found that high well-being was linked to lower emotional reactivity and more intact emotional regulation operationalized by variability, instability, and inertia (51).

EMA research intends to catch people in their immediate environment ranging from office hours to hiking trips. Supported by GPS data, a large-scale study found that feeling better was related to being in natural surroundings after controlling for confounders such as weather, daylight, activity and demographics (52). This effect was found to be long-lasting, as ‘seeing trees and the sky’ predicted well-being 2.5 hours later and ‘feeling of being in contact with nature’ predicted well-being 5 hours later (53).

Emotional vulnerability to extreme life events is well-known for some mental diseases. But what about emotional reactivity to daily life stress? EMA research found that depressed and psychotic patients were more reactive to daily social stress in negative affect compared to bipolar patients and healthy controls. Psychotic and bipolar patients showed higher daily social stress reactivity in positive affect than depressed patients and healthy controls (54). Another study investigated whether anhedonia in schizophrenic patients is related to the lack of hedonic capacities or positive experiences. Patients reported higher intensity in negative affect, lower intensity in positive affect, and fewer positive events compared to healthy controls. However, when experiencing a positive event, hedonic experience was equally strong for patients and controls ($\chi^2(3) = 1.69, p = .64$). The same was found for social hedonic capacity; patients were less often in company of others, but expressed the same level

of positive affect when in company compared to controls ($\beta = 0.01$, $p = .76$). Thus, patients lacked positive and social experiences, not hedonic capacities (55).

Hen or egg – does poor sleep lead to bad mood or vice versa? Even if repeated EMA sampling doesn't allow for causal conclusions, the effect of sleep on mood ($\beta = .344$; $p < .001$) was much larger than vice versa ($\beta = .132$; $p < .001$) after standardization accounted for individual differences and potential confounders were controlled. If sleep quality had an impact on next-day mood clinical implications emerged (56).

Some psychiatric disorders are faced with temporary events such as paranoid episodes. EMA research sharpened our knowledge of the episode's onset which is preceded by increased anxiety and decreased self-esteem, the episode's character which is marked by high levels of anxiety, depression, and anger and low self-esteem, as well as the episode's length which correlates with initial paranoia intensity, depression and anger (57).

EMA can facilitate to understand the effect of treatment response on daily life outcomes. Antidepressant treatment responders spent more time on household chores and less time in passive leisure activities, experienced more positive and less negative affect and reported better sleep quality (58). Further, change in EMA of positive affect during the first week predicts score change, response and remission rates at the end of treatment ($OR = -0.6/4.3/9.3$, $p < 0.001$) (59). Reporting side effects via EMA might be more feasible than retrospective assessment, as it was correlated with lower momentary quality of life and higher dropout risk (60).

Limitations

In its current form, EMA requires vision, literacy, self-awareness, smartphone skills and internet access (at least from time to time), and the ability to respond regularly. This excludes certain populations such as infants, persons with severe forms of dementia or disabilities. Even for healthy individuals, EMA can be demanding and fatigue may lower compliance and data quality.

Despite measuring in real-world contexts which is claimed to achieve high ecological validity, EMA still relies on subjective reporting, which can be biased by individual interpretation of items or careless responding. Frequent assessments may also influence participants' behaviour, reducing ecological validity (61). Comparisons with retrospective ratings show that depressed patients often recall symptoms more negatively, while nonclinical individuals show mixed recall patterns (62). Momentary

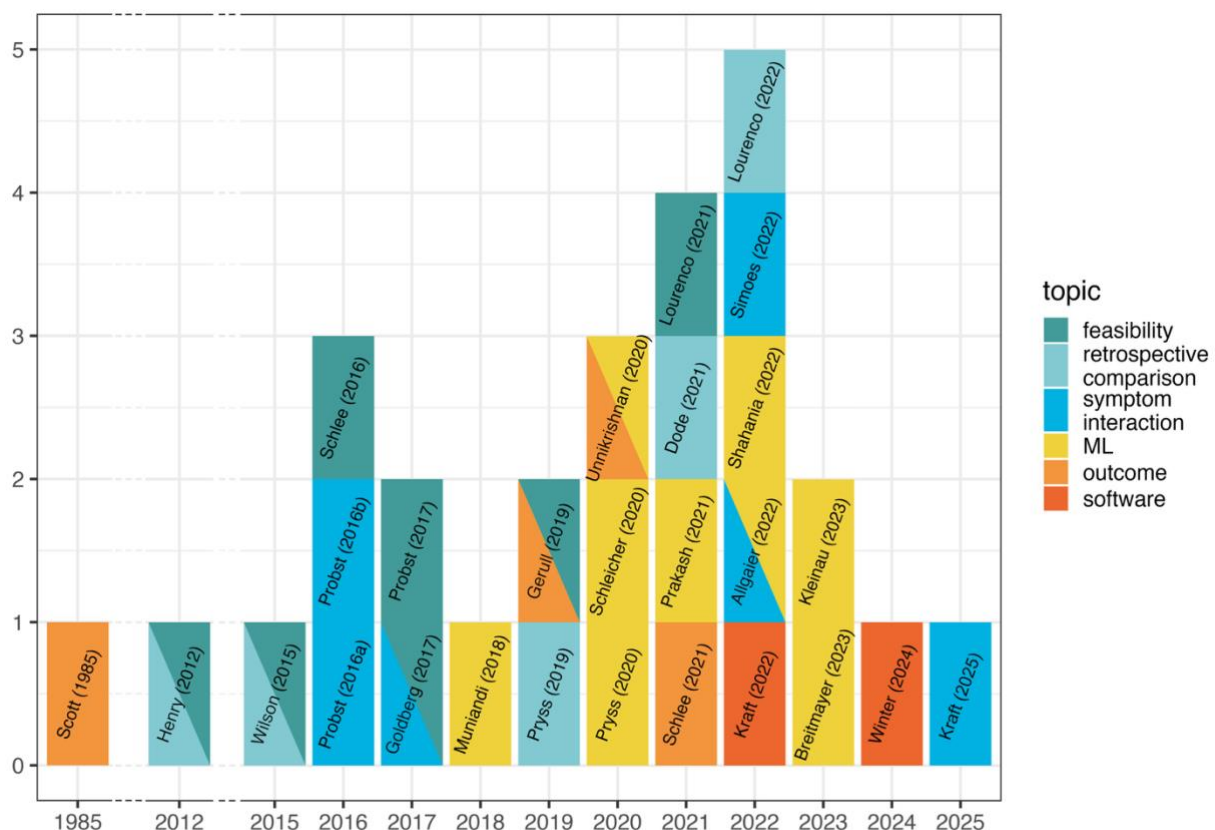
assessments reduce recall bias but may still reflect cognitive distortions, especially in patients with comorbid conditions like depression or hyperacusis (63).

Though EMA supports temporal analysis, it remains observational. Even when confounders are controlled, causal conclusions cannot be firmly drawn. EMA is promising for understanding treatment effects, but missing data can be non-random if patients drop out due to poor response (63). Also, frequent monitoring may increase self-awareness and reduce symptoms, potentially inflating perceived treatment effects (61).

Ecological Momentary Assessment in Tinnitus

Despite some pioneering work in earlier days (64,65), the EMA method entered tinnitus research 10 years ago. An overview of EMA studies published over the years is depicted in Figure 1⁵. Findings from those studies are highlighted in the following.

Figure 1. Scientific Articles on EMA in Tinnitus



Note. ML = Machine learning. Some studies could be allocated to more than one category (bicoloured). The scale breaks indicate greater time intervals.

⁵ Figure in adapted form taken from: Engelke M, Müller S, Langguth B, Pryss R, Schlee W. Tinnitus measured in everyday life: a literature review of ecological momentary assessment studies. *J Assoc Res Otolaryngol.* 2025;26:369–88.

Feasibility

Feasibility research was driven by the question if EMA might worsen or improve tinnitus through increased awareness (65). To date, there is no empirical evidence for that. EMA does not affect tinnitus distress (33,65–69) nor loudness (33,66), while the former was evaluated more often. Importantly, several EMA schedules of different lengths and intensities have been tested. It should be noted that EMA might increase tinnitus awareness (65,66). Nevertheless, EMA is suggested to be a safe and non-reactive measurement tool.

Retrospective Comparison

EMA research revealed insights into considerable differences between prospective and retrospective assessment of tinnitus variation and influence of stress on tinnitus, indicating that some data might be hardly accessible from retrospective assessment (70). A study compared one week of EMA measured tinnitus distress with the Mini-TQ, showing that correlations between both declined over the study period (1st day corr: 0.52) (71). However, both assessments do not capture the same period, as the Mini-TQ was assessed prior to the EMA timeframe. An individual-level comparison between retrospective and prospective measures of tinnitus severity showed that more patients had more severe values in the retrospective than in the prospective assessments (65,72). It is likely that both methods cover different aspects of the same phenomenon.

Symptom Fluctuations and Interactions

EMA research empirically confirmed the suspected high within- and between-subject variability in tinnitus distress among participants, including high fluctuations within a single day (33,65,72). Wilson et al. suggested a high moment-to-moment variability with unpredictable and strong changes in tinnitus symptoms as a potential cause of tinnitus bother.

A direct relationship between tinnitus loudness and tinnitus distress, with a mediating role of stress and its components, emotional arousal, and valence was found (73). A follow-up investigation showed that variability of emotional arousal and valence was positively correlated with the Mini-TQ score and moderated the effect between tinnitus loudness and tinnitus distress. Variability in emotional valence showed an effect on the time course of tinnitus loudness; the higher the variability, the more tinnitus loudness increased over time (74). Goldberg et al. showed that tinnitus bother,

loudness, and general stress correlated within individuals over time. Thus, if stress increased, tinnitus bother and loudness increased as well (66). Circadian rhythm analysis revealed an increase in tinnitus loudness and distress during the night and early in the morning compared to other times of the day and controlling for general stress (75). Autocorrelation is a well-known phenomenon in EMA research, which means that the value of one assessment has an effect on the following. Yet, an EMA study could not find cross- (same day) or auto-correlations (following day) between tinnitus symptoms, behavioural and emotional processes on a group level. In individual models, tinnitus loudness and distress were cross-correlated in 65% of participants, while auto-correlated only in 4%. Cross-correlation was generally more frequent than auto-correlation, thus, the condition of one day is independent of the condition on the next day for many tinnitus patients (76).

A very recent study investigated the relationship of within-subject momentary tinnitus loudness with sensor-based ambient sound level and identified three groups with differing effects of sound on tinnitus; 18% showed a reduction in tinnitus loudness with increasing environmental sound level, 77% did not change, and 5% showed an increase in tinnitus loudness (77).

Machine Learning

Machine learning algorithms are able to detect patterns beyond observation and basic statistics, yet they require a huge amount of data, which EMA is able to provide. Muniandi et al. predicted EMA tinnitus symptoms from cross-sectional baseline data, though they note the preliminary nature of their results (78). Pryss et al. identified systematic differences in EMA responses between iOS and Android answers with a classification accuracy of 79% (79). A study on user adherence found a steady decline in app usage over time, predicted readmission of regular reporting by stress and tinnitus perception during the first days of usage, and showed that female participants had higher adherence levels (80). Further, it was found that non-personalised tips introduced later in the study increased long-term adherence to the protocol (81). Prakash et al. examined if neighbour-hood based models were able to predict future time courses of EMA data with the result that reliability decreases if time distance increases (82). Another study found that prediction of symptom time courses by individual instead of group-level data is only superior if a certain amount of individual data is available (83). Incorporating neighbourhood data into idiographic models

improved their accuracy⁶ (84). Allgaier et al. found varying patterns of tinnitus perception depending on country and season (85). Breitmayer et al. predicted the presence of tinnitus based on tinnitus-unspecific dimensions such as mood, stress, arousal, and concentration with accuracies up to 78% (86).

Software

Comparability of ambulatory assessment studies is hampered by different usage of soft- and hardware, Kraft et al. identified that those differences could be identified and corrected for by the help of EMA data (87). Winter et al. suggested an interdisciplinary modelling approach of EMA data based on process mining, a technique which uses information technology data logs to predict and improve business processes (88).

Outcome Research

Scott et al. compared the pooled momentary and end-of-day data of a treatment group against a waiting-list control group to incorporate variations over time and different situations (64). Unnikrishnan et al. examined distributions of EMA data before and after treatment to detect changes, however, the findings were not supported by changes in the THI (81). Another study analysed longitudinal trends in tinnitus symptoms during treatment and found that the most pronounced improvements occurred in tinnitus loudness and general stress (89).

Despite such examples, the potential of EMA in clinical tinnitus research remains underexplored. In most clinical studies, information on a patient's condition is obtained solely via PROMs administered before and after treatment. This leaves a substantial gap in our understanding of the intervening period, which informs us about how patients improve or deteriorate, the extent to which symptoms fluctuate, as well as potential interactive mechanisms with daily life behaviour. Through repeated, ecologically valid assessments, EMA can provide rich insights into individual improvement trajectories, setbacks, and potential needs for treatment duration in real-world contexts.

⁶ Kleinau A, Flügel S, Pryss R, Vogel C, Engelke M, Schlee W, et al. Predicting patient-based time-dependent mobile health data. In: *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*. 2023; p. 79–84. IEEE.

Scope of this Thesis

The overarching aim of this thesis was to investigate the clinical utility of intensive longitudinal assessments conducted during tinnitus treatment. The scope was twofold: (i) to characterise and understand the fluctuations in tinnitus symptoms during treatment, and (ii) to explore analytical approaches suitable for longitudinal, real-world data.

Therefore, the first study explored several methods for quantifying treatment effects using EMA data and compared those with established PROMs. The second study examined in greater detail how clinical improvement manifests in EMA-derived symptoms and focused on whether patterns of symptom fluctuation can predict treatment effects. The third study explored heterogeneity in symptom trajectories using latent class analysis. All studies also examined coherence between EMA measures and PROMs.

This thesis' research was conducted within the UNITI-RCT (Unification of Treatments and Interventions for Tinnitus Patients – Randomised Clinical Trial), which compared the effect of several single and combined tinnitus treatments with the ultimate goal of developing a decision-support system supporting clinicians and patients in selecting the most optimal treatment in a data-driven way. A dedicated smartphone application delivered two treatments and hosted a tinnitus diary for EMA data collection. The first study drew on a pilot trial evaluating the app-based treatments in a small cohort. The second and third study were based on data from the full clinical trial comparing all treatment arms. All three studies employed the same EMA items (Table S1).

Study I: Pilot Study of a Smartphone-based Tinnitus Therapy using Structured Counseling and Sound Therapy: A Multiple-Baseline Design with Ecological Momentary Assessment

Milena Engelke^{1*}, Jorge Simões¹, Carsten Vogel², Stefan Schoisswohl¹, Martin Schecklmann¹, Stella Wölflick¹, Rüdiger Pryss², Thomas Probst³, Berthold Langguth¹, & Winfried Schlee^{1,4}

¹ Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany

² Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

³ Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems, Austria

⁴ Institute for Information and Process Management, Eastern Switzerland University of Applied Sciences, St. Gallen, Switzerland

The following text, tables and figures are copy-edited versions of the final article published on the 17th of January 2023 in PLOS Digital Health:

Engelke M, Simões J, Vogel C, Schoisswohl S, Schecklmann M, Wölflick S, et al. Pilot study of a smartphone-based tinnitus therapy using structured counseling and sound therapy: A multiple-baseline design with ecological momentary assessment. PLOS Digit Health 2023;2(1);e0000183. <https://doi.org/10.1371/journal.pdig.0000183>

The final publication is available at:

<https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000183>

Copyright: © 2023 Engelke et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Tinnitus affects a considerable part of the population and develops into a severe disorder in some sufferers. App-based interventions are able to provide low-threshold, cost-effective, and location-independent care for tinnitus patients. Therefore, we developed a smartphone app combining structured counseling with sound therapy and conducted a pilot study to evaluate treatment compliance and symptom improvement (trial registration: DRKS00030007). Outcome variables were Ecological Momentary Assessment (EMA) measured tinnitus distress and loudness and Tinnitus Handicap Inventory (THI) at baseline and final visit. A multiple-baseline design with a baseline phase (only EMA) followed by an intervention phase (EMA and intervention) was applied. 21 patients with chronic tinnitus (≥ 6 months) were included. Overall compliance differed between modules (EMA usage: 79% of days, structured counseling: 72%, sound therapy: 32%). The THI score improved from baseline to final visit indicating a large effect (Cohens $d = 1.1$). Tinnitus distress and loudness did not improve significantly from baseline phase to the end of intervention phase. However, 5 of 14 (36%) improved clinically meaningful in tinnitus distress ($\Delta\text{Distress} \geq 10$) and 13 of 18 (72%) in THI score ($\Delta\text{THI} \geq 7$). The positive relationship between tinnitus distress and loudness weakened over the course of the study. A trend but no level effect for tinnitus distress could be demonstrated by a mixed effect model. The improvement in THI was strongly associated with the improvement scores in EMA of tinnitus distress ($r = -0.75; 0.86$). These results indicate that app-based structured counseling combined with sound therapy is feasible, has an impact on tinnitus symptoms and reduces distress for several patients. In addition, our data suggest that EMA could be used as a measurement tool to detect changes in tinnitus symptoms in clinical trials as has already been shown in other areas of mental health research.

Introduction

Experiencing subjective tinnitus implies an involuntary perception of a phantom sound that is generated by abnormal neural activity in the auditory system (7,90). Around 14-15% of the population in European countries are affected by tinnitus, while 1-2% are severely impaired by tinnitus (91,92). Tinnitus is a heterogeneous condition, not only in terms of level of impairment, but also regarding its perceptual characteristics, etiological factors, pathological markers, clinical comorbidities, and its treatment response (7,93,94). Psychological, auditory, and neurological treatment approaches

exist that aim towards reduction of tinnitus symptoms and improvement of life quality, but up to now, there is no universal cure for tinnitus available for the clinical routine (90,95).

Patients who seek health care usually suffer from severe tinnitus which is accompanied by emotional distress, cognitive dysfunction and increased arousal ultimately leading to functional disability and reduced life quality (7,91). Among others, patients complain sleep disturbances, problems in concentration, attention, communication, and memory and some are severely affected by mental disorders (96,97). Tinnitus severity is mainly influenced by psychological factors. High life stress, poor coping resources, co-morbid mental diseases and personality dimensions shape negative emotional reactions, selective attention, and cognitive appraisal towards the tinnitus. Involving numerous feedback loops, tinnitus distress is likely to persist which in turn prevents habituation to the new stimulus (98–100). Tinnitus management strategies aim towards stimulation of the habituation process by reducing stress, anxiety, negative emotional responses, and dysfunctional cognitions on one hand, and by enhancing coping resources on the other hand (36,101).

Despite scientific efforts, the health care situation for tinnitus patients is deficient, since the access to available treatment options for tinnitus is limited (102). In addition to the lack of trained therapists, environmental factors such as the COVID-19 outbreak in 2020 can further reduce access to face-to-face treatments (103–105). Moreover, tinnitus has a substantial economic burden in terms of health care and indirect costs with increasing trend (91,106,107).

In response to the deficient health care situation, smartphone-based therapies are capable to provide low-threshold (easily accessible), cost-effective and location-independent healthcare and can potentially increase patients treatment compliance (102,108,109). They are particularly beneficial for the management of chronic diseases as they strengthen patient empowerment and are able to induce behaviour modification and lifestyle changes (108,109). Furthermore, smartphone-based therapies can be used to provide information about tinnitus to a large number of affected people, as well as to track the course of tinnitus symptoms in a longitudinal, ecologically valid, and momentary fashion (33,72,102). Currently, there are many available smartphone applications for tinnitus, but evidence for their efficacy is limited (102,110). Preliminary findings indicate that app-based treatments have the potential to improve tinnitus symptoms (110–115). Most evaluated apps use CBT approaches

and auditory treatments; to the best of our knowledge, there is currently no evaluation of an app-based SC combined with ST.

Therefore, we developed a tinnitus management app that provides SC supported by ST. SC has been shown to improve tinnitus distress and quality of life in face-to-face settings or through written manuals (116,117) and may have better effects when combined with ST (117–119). Additionally, the app contains an EMA module that tracks changes in tinnitus symptoms on a daily basis. The feasibility of EMA to measure the course of tinnitus has already been investigated in several studies (33,67,72) and has also been used to evaluate tinnitus treatments (89).

Hence, we conducted a pilot-study to evaluate the app-based treatment. The objective was to investigate treatment adherence and improvement of tinnitus symptoms individually and on a group level. A trend effect (improvement over time) for daily measured tinnitus distress was hypothesized. Furthermore, we analysed the association of the two measurement methods, namely EMA and clinical questionnaires. To account for tinnitus heterogeneity and individual treatment effects, a multiple-baseline across groups design was chosen. Multiple-baseline designs belong to the group of single-case experimental designs (SCED) where the outcome variable is measured longitudinally before (baseline phase) and during the intervention (intervention phase) in single cases. Hence, with the baseline phase, each case has its own control and is considered as replication (120). As SCEDs can be performed with small sample sizes, they are especially recommended for piloting novel interventions (121). In multiple-baseline designs, the introduction of the intervention phase is staggered in time to separate treatment effects from random time effects and therefore enhance internal validity (120).

Methods

Study design

A randomised multiple-baseline design across groups was used. Participants first underwent a baseline phase where tinnitus symptoms were measured daily using app-based EMA, followed by an intervention phase where the intervention (SC and ST) was additionally introduced. Participants were allocated to 4 groups with different baseline lengths to realize staggered introduction of the intervention and to fulfil the recommendations of at least 3 replications (121). The baseline phase had a minimum duration of 7 days (i.e. 7 EMA data points) to exceed the recommended standard of 5

Study I

data points per phase (122). Furthermore, the groups entered the intervention phase with a distance of 3-5 days to each other to keep baseline phase as short as possible. As randomization is recommended for SCEDs to minimize threats to internal validity, the Koehler-Levin randomization procedure was used to randomize the intervention start (123,124). Therefore, 2 potential starting points for each group were defined considering the 3 – 5 days distance to the other groups. First, the participants were randomly assigned to the 4 groups. Then, the order of the groups was randomly assigned and finally, one out of two potential starting points was randomly selected for each group. The intervention phase had the same duration of 12 weeks for all participants. At the beginning of the baseline phase and at final visit (after intervention phase), participants completed online versions of clinical questionnaires about tinnitus and other health-related aspects using an electronic database. The study was designed in line with SCED recommendations (122).

The study is approved under the ethic vote 20-1936_2-101 „UNification of treatments and Interventions for Tinnitus patients – Randomised Clinical Trial (UNITI-RCT)“ (Amendment No. 2) by the ethics committee of the University of Regensburg, Germany (125). All persons were informed about the objectives and methodologies of the present study, and gave written informed consent prior to study start. Participants received neither compensation nor incentives for adherence to the study protocol. The study was conducted according to the guidelines in the Declaration of Helsinki. It was retrospectively registered in the German Clinical Trials Register (DRKS-ID: DRKS00030007) because of lack of awareness that pilot trials need to be registered.

Participants

Inclusion criteria were: 1) adult (age ≥ 18 years), 2) chronic tinnitus (duration ≥ 6 months), and 3) THI-score ≥ 18 (indicating at least a mild tinnitus handicap) (28). We included patients with any tinnitus type (e.g., subjective, objective, somatosensory) regardless of comorbidities (e.g., hearing loss, hyperacusis). Exclusion criteria were: 1) tinnitus-related therapy within the last 3 months and 2) similar therapy with long term effects e.g., noiser. Further, potential participants had to own a smartphone with the operating systems iOS 12.0 (or newer) or Android 5.0 (or newer) in order to use the mobile application as part of the study.

Procedures

The whole study was conducted online during the COVID-19 pandemic (May 2021 – October 2021). Patient communication was realized by mail. By giving their consent, participants were included into the study and randomised to one of the four groups. Also, they were provided online via a database with the clinical questionnaires before the baseline phase. With the start of the baseline phase on the same day for all participants, information about app installation and login data were provided. During baseline phase, participants were instructed to answer the diary (EMA module) every evening. With the start of the intervention phase, participants received instructions and new login data that enabled access to the treatment modules. All participants immediately started the intervention phase on the defined day. Apart from answering the diary, participants were instructed to complete daily sessions of SC and ST during a 12-week intervention phase. At halftime (after 6 weeks of intervention), a reminder to fulfil the study protocol (daily usage of intervention and diary) was sent out. The intervention phase ended after 12 weeks with the final visit assessment.

App-based intervention

Both EMA recordings and the intervention modules (SC and ST) were provided via the smartphone-based application UNITI running on the operating systems iOS 12.0 or newer and Android 5.0 or newer. It is available in the Apple App Store and the Google Play Store. App architecture is described elsewhere (126).

Structured counseling

The SC module *TinEdu* consists of 12 chapters, each chapter includes 7 sections and a quiz at the end. The sections contain informative texts, related links, illustrations, and audio files imparting knowledge about tinnitus ranging from basics about hearing and tinnitus to more advanced topics such as therapeutic options and current tinnitus research. In addition, practical tips are provided on how to deal with tinnitus on a daily basis. The last section of each chapter gives a short summary of the provided content. After that, participants can test themselves in a subsequent quiz. The quiz consists of single- and multiple-choice questions related to the respective chapter. Participants were provided with feedback on their test performance and were advised to repeat the relevant sections in case of a wrong answer. The Feedback section is a separate module of the UNITI app that functionally belongs to *TinEdu* and saves the feedback of the quizzes. As the intervention is scheduled for 12 weeks, one section should be

Study I

completed each day so that a chapter is completed every week. To ensure the processing order of *TinEdu*, each section is unlocked only when the previous one has been completed.

Sound therapy

The ST *Shades of Noise* (*SoN*) forms the second treatment module of the UNITI application. It provides a sound library of 64 sounds from which 29 are naturalistic and 35 are digitally produced sounds. In the first category, sounds as beach waves or waterfall noise are presented. The second category consists of pure tones and broad band noises of different frequencies with optional filters and modulation rates representing the different sound therapeutic approaches (e.g., masking, residual inhibition). At the first use, subjects could optionally set their tinnitus frequency, if known, so that some sounds could be adjusted to the pitch. To increase the benefit of the therapy, it was recommended to listen to the sounds with headphones to achieve a more even distribution to both ears. When participants started a session, they had to indicate their current tinnitus loudness and distress on a visual analogue scale (VAS). After that, they could select any sound from the library to listen to. Different play times could be set, between 1 minute and infinite play time. Also, sounds could be marked as favourites, making them appear at the top of the sound list for the upcoming sessions. Furthermore, each sound could optionally be rated on how much the tinnitus loudness changed after listening to it. Participants were encouraged to use *SoN* in a self-guided manner with a recommendation of at least 15 minutes per day.

Measures

Outcomes were measured with clinical questionnaires at the beginning of the baseline phase and at the final visit assessment after the end of the intervention provided via an electronic database, as well as with active EMA data using the diary via the smartphone-application. Data protection and data transfer was handled according to the highest security standards and has been approved by the Data Protection Officer (DPO) of the University of Regensburg, Germany.

Clinical questionnaires included the THI (28), the German Hyperacusis Questionnaire (GUEF) (127), the Patient Health Questionnaire for Depression (PHQ-9) (128), and the European School of Interdisciplinary Tinnitus Research Screening Questionnaire (ESIT-SQ) (129). The THI consists of 25 items with 3 answer options addressing the impact of tinnitus on daily life. Higher THI score indicates higher tinnitus handicap (28). The remaining questionnaires provide information on demographics,

Study I

tinnitus characteristics, co-morbid hyperacusis and mental health. The questionnaire data were stored anonymized in an electronic database (EU Tinnitus Database). The database is hosted by database experts from the University of Regensburg at the Bezirksklinikum Regensburg, where the server is located.

The EMA module (tinnitus diary) consisted of a collection of 10 VAS questions regarding current tinnitus perception and thoughts, neck and jaw tension, as well as psychological wellbeing. In an additional empty field, subjects were allowed to make notes on special occurrences. The EMA module was designed as an end-of-day diary (EDD) to track changes in tinnitus symptoms evoked by the intervention, therefore, participants received daily notifications at 19:30. The authors are aware that the notifications might influence the time of app usage. For this study, results of two questions about tinnitus perception have been included in the analysis (see Table S1). These questions assess the current tinnitus distress (*How burdensome do you find your tinnitus at the moment?*) and the current tinnitus loudness (*How loud is your tinnitus at the moment?*) on a continuous rating scale (0 – 100; *not burdensome/inaudible – very burdensome/very loud*). The app data are stored in a relational database using Maria DB 11, which runs in a Linux environment (Debian Buster) and on a LAMP technology stack. No personal information is stored there.

Statistical analysis

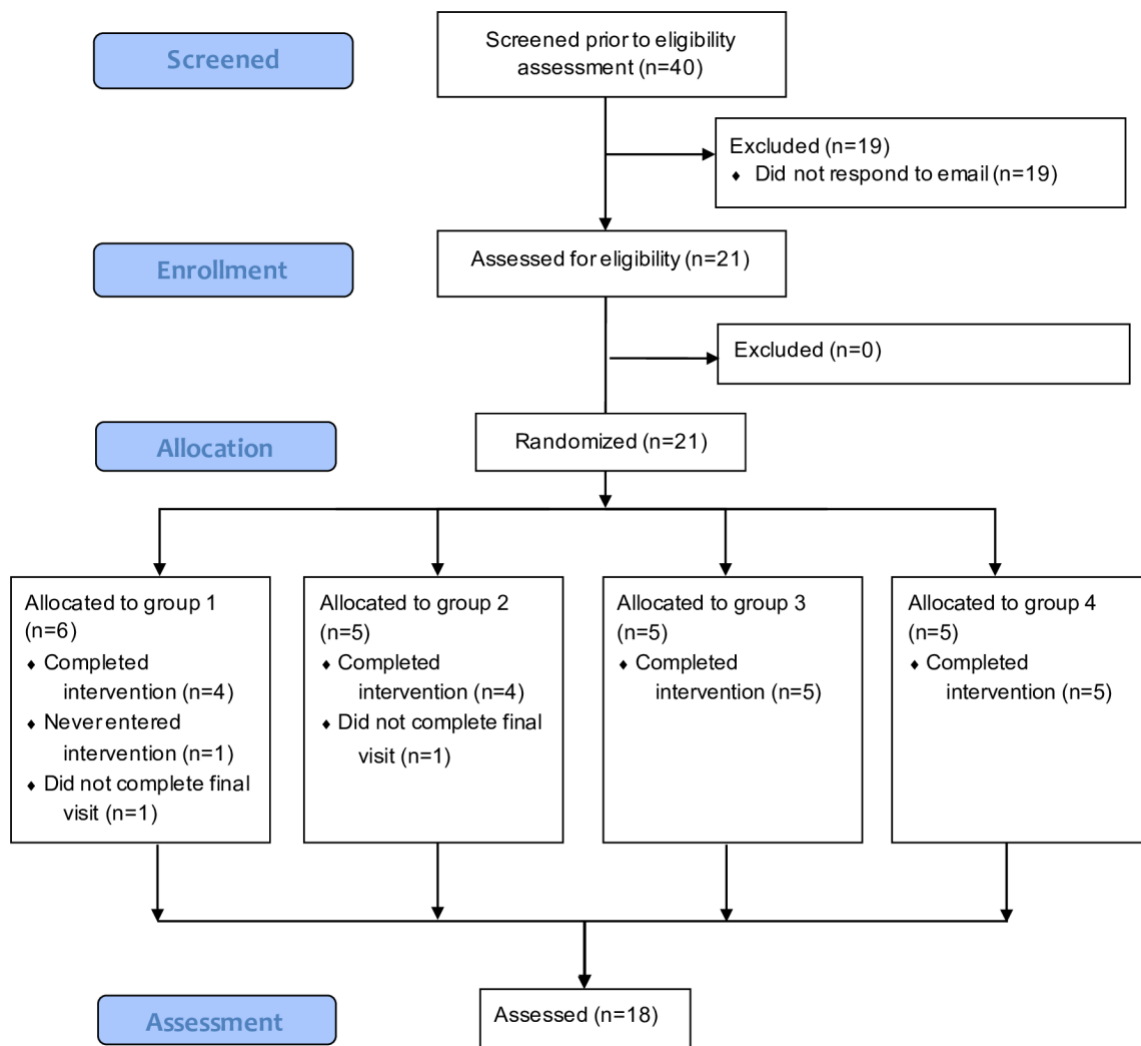
All participants who filled in the 'begin of baseline' and 'final visit' questionnaires and used the app at least once during baseline and intervention phase were included in the data analysis. Descriptive statistics for demographics, tinnitus characteristics (retrieved from the clinical questionnaires) and app use were calculated to characterize the sample. A paired t-test was used to determine the sample difference in THI score from begin of baseline to final visit. The main analysis focus was on the EMA data provided by the app diary. Therefore, data were additionally prepared and missing values were imputed using the `aregImpute` function of the `Hmisc` R package with its default settings (130). To support visualization of tinnitus symptom courses, individual means and standard deviation were calculated for baseline phase and end of intervention phase and statistically contrasted on a group level using a paired t-test. The change in relation between tinnitus distress and loudness was described with Pearson's product-moment correlation for each week. A mixed effect model was computed to predict tinnitus distress including demographics (age, gender), psychological burden (THI, PHQ-9) and intervention variables (trend, level) as fixed effects and participants as random

intercepts. To compare the different measurement methods, Pearson's product-moment correlation between improvement in THI and improvement in tinnitus distress was calculated. Data preparation and analysis was conducted in R (version 4.1.2).

Results

Participants

A minimum of 20 participants were defined for inclusion to allocate at least 5 participants to each group. 40 potentially eligible patients were preselected from a patient pool based on inclusion criteria and contacted per mail (see Figure 2 for patient flow). 21 declared their consent, were included into the study and randomised to one of the 4 groups. The Koehler-Levin method was applied (123,124). Group 1 was assigned to a baseline length of 11 days, group 2 of 19 days, group 3 of 16 days and group 4 of 8 days. Double-checking of the eligibility at begin of the baseline phase confirmed that all 21 participants fulfilled the inclusion criteria. For data analysis, 3 cases (3/21, 14%) had to be excluded, because of missing final visit questionnaires (case 02, case 09) and missing app data from the intervention phase (case 01). Hence, 18 patients are considered for data analysis.

Figure 2. CONSORT Flow Diagram

Note. Flow diagram of patient recruitment and dropouts. Study design: Multiple-baseline design across groups. The four groups introduced the treatment staggered in time to separate treatment effects from potential time effects and therefore enhance internal validity. Baseline length for each group was randomly selected following the Koehler-Levin procedure and patients were randomly allocated to groups. Intervention phase was the same length for all groups. Before baseline and after intervention phase, clinical questionnaires were applied.

Demographics, Tinnitus Characteristics and App Use

Table 1 summarizes demographics, tinnitus characteristics and app use. The sample was balanced in terms of gender, age and education level. PHQ-9 score at baseline indicated mild depressive symptoms on average (131). The sample was moderately handicapped (THI) by their tinnitus, slightly to moderately impaired by hyperacusis (GUEF) and has had tinnitus for an average of 183 months (127,132). 11 patients had seen a physician in the past year due to their tinnitus complaints. Patients had good overall compliance using the app. On average, the diary was used on 75.9 days (about 79% of days), SC was used on 60.7 days (72% of days), and ST on 26.6 days (32%

of days). A daily usage was recommended (the intervention phase lasted for 84 days and the baseline phase for 8 – 19 days, depending on group affiliation).

Table 1. Participant Demographics, Tinnitus Characteristics and App Use

Participant demographics	
Participants	<i>N</i> = 18
Female, <i>n</i> (%)	8 (44.4%)
Age in years, mean (range)	49.8 (24 – 68)
Education, <i>n</i>	2 (elementary school), 5 (secondary school), 3 (A-levels), 8 (university)
PHQ-9, mean (SD)	5.7 (3.3)
Tinnitus characteristics	
THI, mean (SD)	38.4 (15.4)
Tinnitus duration in months, mean (range)	183 (20 – 350)
GUEF Hyperacusis, mean (SD)	10.3 (6.7)
Vertigo, <i>n</i>	4 (never), 6 (less than once a year), 8 (more than once a year)
Number of perceived sounds, <i>n</i>	12 (1 sound), 6 (more than 1 sounds)
Tinnitus onset, <i>n</i>	6 (slowly increasing), 8 (sudden), 4 (don't know)
Loudness changes, <i>n</i>	5 (never), 11 (sometimes), 2 (always)
Tinnitus quality, <i>n</i>	10 (tone), 5 (noise), 3 (other)
Tinnitus pitch, <i>n</i>	12 (high), 5 (medium), 1 (deep)
Tinnitus lateralization, <i>n</i>	6 (right), 3 (left), 8 (same for both ears), 1 (inside the head)
Tinnitus healthcare visits during the past year, <i>n</i>	7 (0 visits), 7 (1 visit), 4 (2-4 visits)
App use (in sessions)	
Diary, mean (SD)	75.9 (25.6)
TinEdu, mean (SD)	60.7 (29.5)
Shades of Noise, mean (SD)	26.6 (33.3)

Note. Description of the sample at baseline. Tinnitus characteristics were assessed with the European School of Interdisciplinary Tinnitus Research Screening Questionnaire (ESIT-SQ). PHQ-9: Patient Health Questionnaire for Depression; THI: Tinnitus Handicap Inventory; GUEF: German Hyperacusis Questionnaire. App use: Mean and standard deviation of the number of sessions.

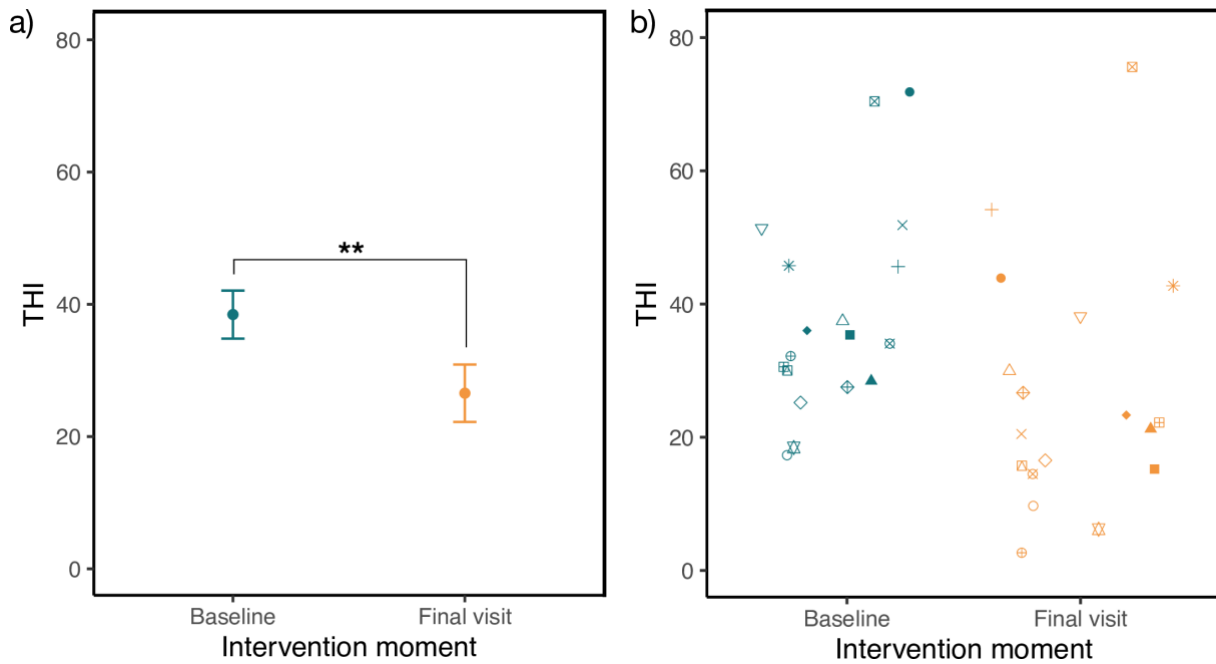
Change in THI

Change of THI score from the begin of baseline phase to final visit is depicted in Table 2 and visualized in Figure 3. The mean THI was improved by 11.8 points from begin of baseline to final visit indicating a large effect (Cohens $d = 1.1$). 13 of 18 cases (72%) improved clinically meaningful on the THI score ($\Delta\text{THI} \geq 7$). A THI improvement of 7 points has been empirically determined as the MCID (133).

Table 2. Improvement of THI from Begin of Baseline Phase to Final Visit

Begin of baseline		Final visit		Paired t-test		
Mean	SD	Mean	SD	$t(17)$	p	Cohen's d [95% CI]
38.4	15.4	26.6	18.4	4.55	< .001	1.1 [0.49, 1.69]

Note. Mean and standard deviation of the Tinnitus Handicap Inventory (THI) score at begin of baseline phase and final visit. Paired t-test reveals a statistically significant improvement from baseline to final visit with a strong effect size. $N = 18$.

Figure 3. Improvement of THI from Begin of Baseline Phase to Final Visit

Note. a) Significant improvement of Tinnitus Handicap Inventory (THI) score from the begin of baseline phase to final visit of the sample. Higher scores indicate higher handicap. Dots represent means, error bars are standard errors; ** $p < .001$; $N = 18$. b) Individual THI values at baseline and final visit. The same shape represents the same individual.

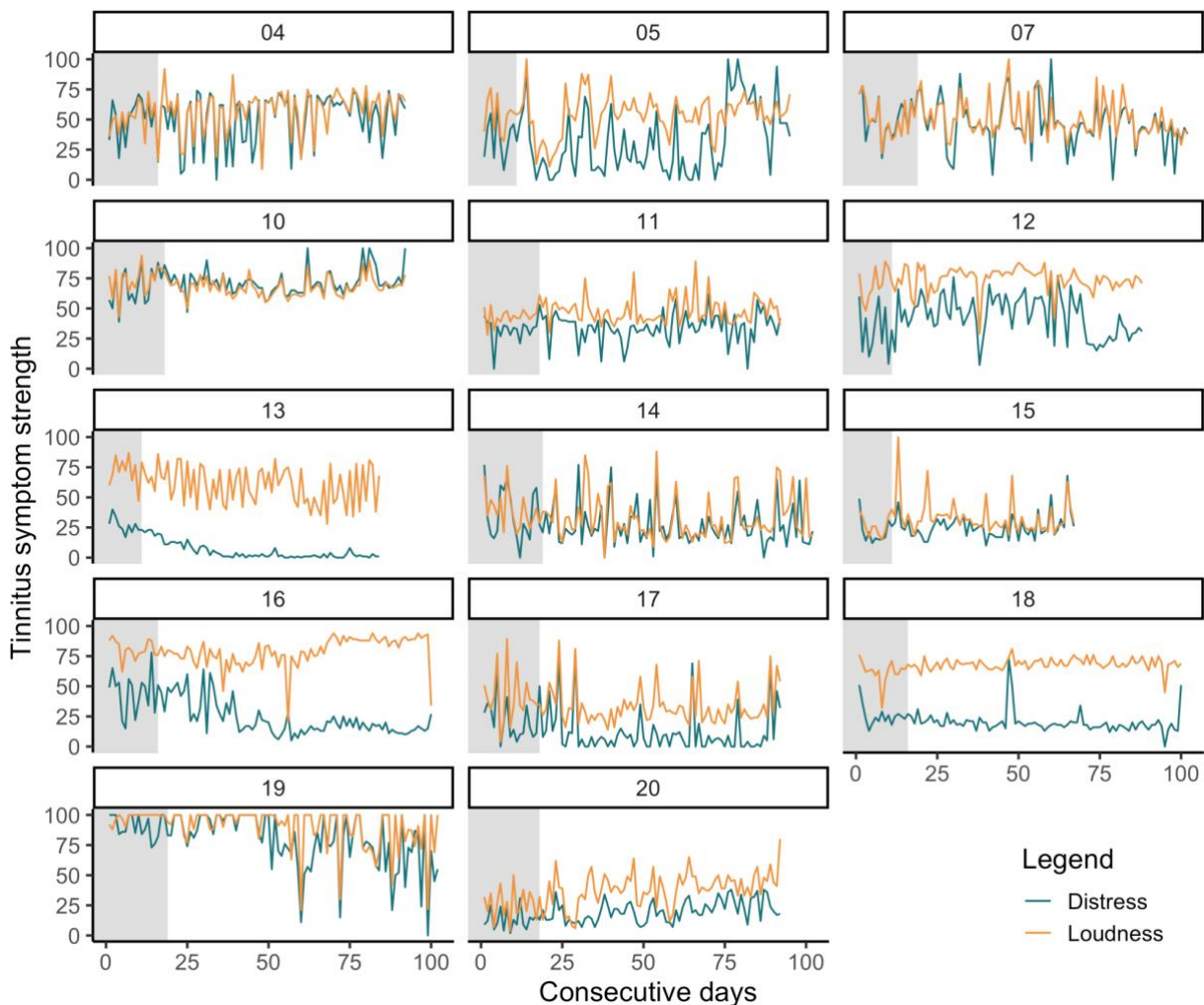
Change in EMA of Tinnitus Symptoms

EMA data from the tinnitus diary were prepared for visualization and analysis creating a structure of one data point per case per day. In case of multiple entries per day, the first one was kept ($n = 79$ duplicates removed). Cases were visually inspected regarding randomness of missing values, 4 cases had to be excluded because of data missing not at random (MNAR) (cases 03, 06, 08, 21). For the remaining 14 cases, missing data points were imputed with multiple imputation using bootstrapping to simulate the process of drawing predicted values from a full Bayesian predictive distribution ($n = 1.510$ imputed data points). The EMA subset ($N = 14$) is comparable with the full sample ($N = 18$) regarding participant demographics and tinnitus characteristics (see Table S2).

Visualization of Tinnitus Symptom Courses

Figure 4 displays the course of daily measured tinnitus distress and loudness for each case over the study duration. The grey area represents the baseline phase (only EMA), the white area represents the intervention phase (EMA, SC, ST). The heterogeneity between cases and the fluctuation of symptom courses hamper identification of symptom improvement from visualization, therefore, descriptive summaries and statistical analysis are performed in the next step.

Figure 4. Course of Tinnitus Distress and Loudness for each Case (EMA)



Note. Individual courses of tinnitus symptoms that were assessed daily with EMA over the study period. The grey area represents the baseline phase (EMA), the white area represents the intervention phase (EMA, SC, ST). Distress: How burdensome do you find your tinnitus at the moment? (0 not burdensome – 100 very burdensome); Loudness: How loud is your tinnitus at the moment? (0 inaudible – 100 very loud). $N = 14$.

Improvement from Baseline Phase to the End of Intervention Phase

Next, we were interested if tinnitus symptoms improved from baseline phase to the end of the intervention phase. The end of the intervention phase is defined by the last days

Study I

of the intervention phase using the same time window than the baseline phase (Group 1: 11 days, group 2: 19 days, group 3: 16 days, group 4: 8 days). Mean and standard deviation were calculated for the baseline phase and the end of the intervention phase. Individual values are depicted in Table 3 for tinnitus distress and Table 4 for tinnitus loudness. On a group level, no significant differences between baseline phase and end of intervention phase have been found neither for tinnitus distress (mean of the difference = 4.5; $t(13) = 1.08$, $p = 0.30$, 95% CI [-4.5;13.6]), nor for tinnitus loudness (mean of the difference = -2.2; $t(13) = -0.79$, $p = 0.45$, 95% CI [(-8,1;3.8)]. On an individual level, 5 of 14 cases (36%) improved clinically meaningful on distress ($\Delta\text{Distress} \geq 10$). Those patients improved clinically meaningful on the THI score as well ($\Delta\text{THI} \geq 7$, see Table 3). Regarding tinnitus loudness, only 2 of 14 cases (14%) improved clinically meaningful from baseline phase to the end of intervention phase ($\Delta\text{Loudness} \geq 10$, see Table 4). An improvement of 10 points has been empirically determined as the MCID for tinnitus distress and loudness measured on a 100 point VAS (134).

Table 3. Improvement in Tinnitus Distress from Baseline to End of Intervention Phase

Cases	Baseline phase		End of intervention phase		$\Delta\text{Distress}$ (EMA)	ΔTHI
	Mean	SD	Mean	SD		
04	52.0	15.5	53.8	16.0	-1.8	8
05	35.0	17.5	48.1	22.0	-13.1	-8
07	50.6	15.4	38.5	13.3	12.1	10
10	63.3	16.9	74.1	10.8	-10.8	-6
11	31.1	14.6	42.9	9.4	-11.8	4
12	32.3	21.0	30.6	6.5	1.7	2
13	26.8	6.5	2.2	2.2	24.6	30
14	36.1	21.5	25.4	18.1	10.7	12
15	21.3	11.2	31.7	15.4	-10.4	8
16	44.1	17.5	14.3	4.3	29.8	20
17	28.4	20.8	21.1	25.3	7.3	14
18	25.4	8.8	18.2	10.1	7.2	20
19	91.2	9.4	60.7	27.9	30.5	28
20	12.7	8.5	25.3	10.3	-12.6	6

Note. Individual distribution of tinnitus distress (EMA) during baseline phase and end of intervention phase (last days according to baseline length). Higher values indicate higher distress. $\Delta\text{Distress}$: Baseline phase – End of intervention phase; ΔTHI : THI begin baseline – THI final visit.

Table 4. Improvement in Tinnitus Loudness from Baseline to End of Intervention Phase

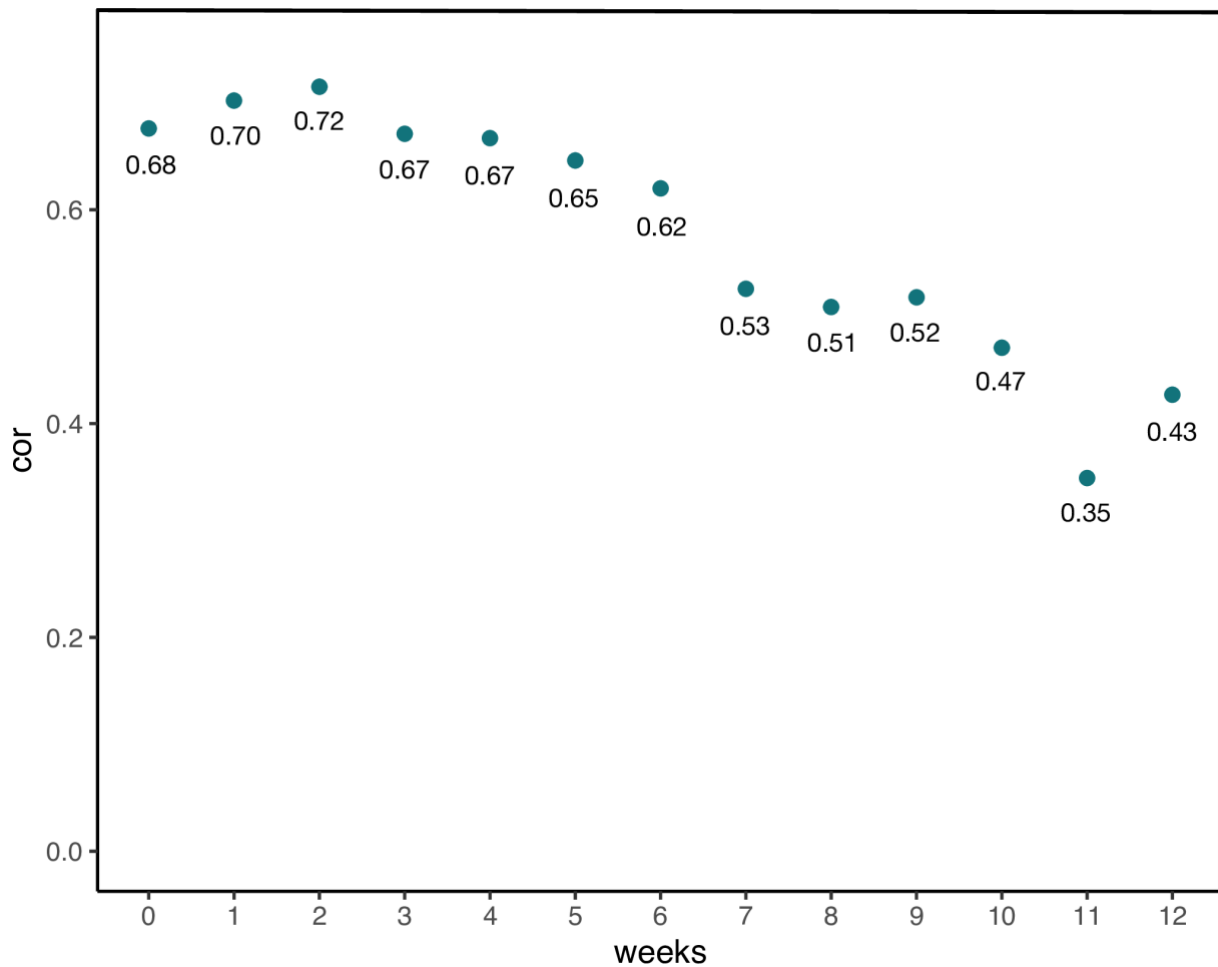
Cases	Baseline phase		Intervention phase		Δ Loudness (EMA)
	Mean	SD	Mean	SD	
04	52.3	12.1	62.6	10.8	-10.3
05	53.2	15.5	56.5	7.9	-3.3
07	50.8	15.3	42.3	8.4	8.5
10	68.9	14.5	69.8	3.9	-0.9
11	41.6	9.3	51.0	8.0	-9.4
12	70.2	13.7	71.5	3.6	-1.3
13	71.2	12.1	58.6	19.9	12.6
14	39.5	15.2	35.7	19.5	3.8
15	25.8	8.0	31.5	12.7	-5.7
16	80.9	7.8	85.6	14.1	-4.7
17	37.9	23.0	41.9	20.5	-4.0
18	61.9	9.7	68.4	7.1	-6.5
19	97.9	3.8	81.9	22.1	16.0
20	25.9	12.1	51.0	14.3	-25.1

Note. Individual distribution of tinnitus loudness (EMA) during baseline phase and end of intervention phase (last days according to baseline length). Higher values indicate higher loudness. Δ Loudness: Baseline phase – End of intervention phase.

Change in Relationship between Tinnitus Distress and Loudness

Visualization of the individual symptom courses have led to the observation that the courses of tinnitus distress and loudness seem to diverge for some individuals (Case 13 & 16, see Figure 4). To quantify that observation, we calculated means for tinnitus distress and loudness for the baseline phase and for each of the 12 weeks of the intervention phase and correlated the weekly distress and loudness means with each other (see Figure 5). All correlations are positive and reach statistical significance ($p < .001$). The strongest correlation was measured in week 2 of the intervention phase ($r = 0.72$) and the weakest in week 11 ($r = 0.35$). On a descriptive level, the correlation coefficients between tinnitus distress and loudness tended to decline over the course of the study.

Figure 5. Change of Correlation between Distress and Loudness over the Study Period (EMA)



Note. Pearson's product-moment correlations between tinnitus distress and loudness during baseline phase (week 0) and intervention phase (week 1-12). All correlations are statistically significant ($p < .001$). $N = 14$.

Trend and Level Effect of Tinnitus Distress

A mixed effect model was only conducted for tinnitus distress, as prior findings indicate no effect for tinnitus loudness. In Table 5, the results are summarized. To evaluate the treatment effect for tinnitus distress, a trend and level effect was modelled (135). The trend effect models the change in tinnitus distress over time (in days), while the level effect models the change in distress between the phases (from baseline to intervention phase). Hence, trend and level were included as fixed effects. Besides from the intervention variables, demographics (age, gender) and psychological burden variables (THI, PHQ-9) were additionally included as fixed effects to enhance tinnitus distress prediction. Participants were included as random intercepts. The outcome variable was daily measured tinnitus distress.

Study I

The model explained 62% of the variance of daily tinnitus distress, mainly explained by the fixed effects (52% explained variance). On average, tinnitus distress was significantly reduced by 0.06 points per day (within the whole study period; trend effect) and descriptively reduced by 2.38 points from baseline to intervention phase (level effect). Moreover, cases with higher daily distress tended to be male with higher baseline THI and PHQ-9 scores.

To test the model for overfitting, the train/test split method was applied. The mixed effect model was fitted to the training data (80% of the data set) and tested on the test data. The explained variance of the test data was $R^2 = 0.67$ (conditional).

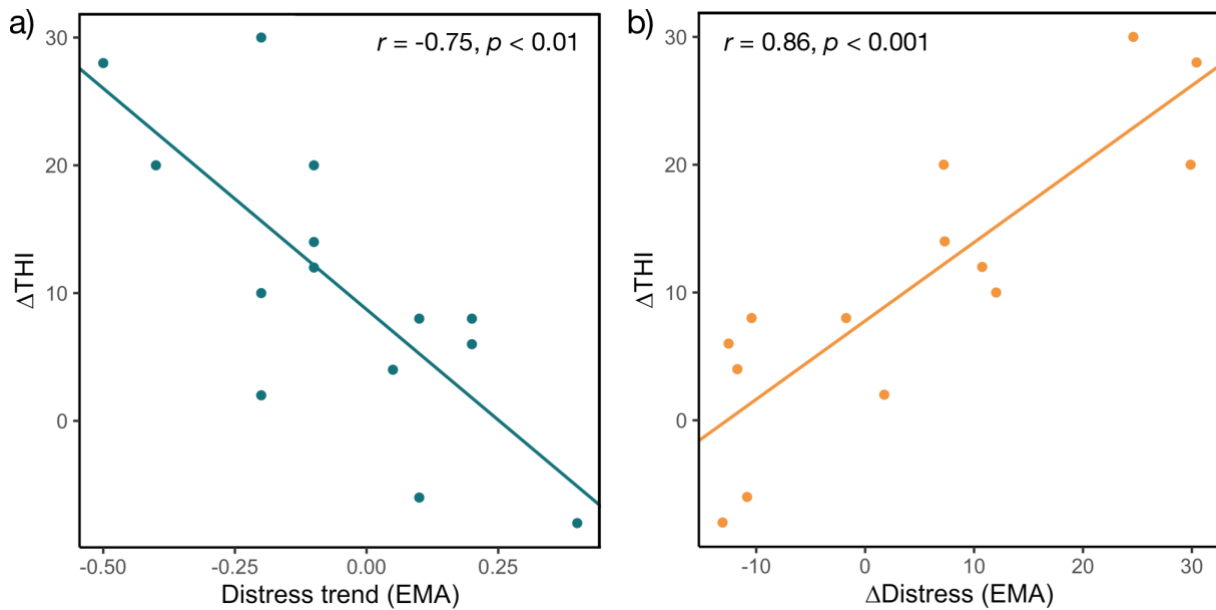
Table 5. Mixed Effect Model

Fixed effects	Coefficient	Se	95% CI	p
Intercept	-24.64	16.48	-53.43 – 4.20	0.17
Trend	-0.06	0.02	-0.10 – -0.02	< 0.01
Level (int.)	-2.38	1.70	-5.78 – 0.90	0.16
Age	0.38	0.27	-0.09 – 0.85	0.18
Gender (m)	17.25	5.13	8.29 – 26.24	< 0.01
THI	0.72	0.19	0.38 – 1.06	< 0.01
PHQ-9	1.52	0.56	0.45 – 2.53	< 0.01
Model fit	R² con	R² marg		
	0.623	0.515		

Note. Prediction of tinnitus distress using intervention variables (trend as time in days, level as difference from intervention to baseline phase), demographics (gender, age) and psychological burden (baseline THI, PHQ-9) as fixed effects and subjects as random effects. Model equation: Tinnitus distress ~ trend + level + age + gender + THI + PHQ-9 + (1|case). Dummy coding: Level (0 baseline phase, 1 intervention phase), Gender (0 female, 1 male). $N = 14$.

Comparison of Improvement in THI and Tinnitus Distress

As the THI score is a well-accepted clinical validation tool for tinnitus treatments, we were furthermore interested if improvement in EMA of tinnitus distress was associated with THI improvement and which EMA improvement score represents THI improvement best. Therefore, we defined two EMA improvement scores. The first score is the Distress trend (slope of tinnitus distress during intervention phase) and the second score is the Δ Distress used before (see Table 3, baseline phase – end of intervention phase). We compared the association of both EMA improvement scores with Δ THI (THI begin baseline – THI final visit), the results are depicted in Figure 6. Δ THI is strongly associated with both EMA improvement scores (Distress trend and Δ Distress), with a descriptively stronger association with Δ Distress (see Fig 5b; $r = 0.86$, $p < 0.001$).

Figure 6. Correlation of Δ THI with Improvement in EMA Tinnitus Distress

Note. Pearson's product-moment correlation of Δ THI with two EMA improvement scores: Distress trend (a) and Δ Distress (b). Δ THI = THI begin baseline – THI final visit; Distress trend = slope of tinnitus distress during intervention phase; Δ Distress = Baseline phase – end of intervention phase. $N = 14$.

Discussion

We conducted a pilot study to investigate the feasibility and efficacy of an app-based tinnitus intervention combining SC with acoustic therapy using a multiple-baseline across groups design with an initial sample of 21 participants. Outcome variables were EMA of tinnitus distress and loudness and retrospectively measured THI at the begin of baseline phase and final visit after intervention phase.

A good compliance using the app was observed, particularly regarding the diary (about 79% daily usage) and SC (72% daily usage). 14% (3/21) dropped out. This high adherence was unexpected since app-based interventions in chronic diseases are normally characterized by high attrition and dropout rates (136). This is especially true for the EMA module, because in a study with its predecessor (TrackYourTinnitus), only about 68% had more than one diary entry (80). The authors assumed that incentives could improve adherence rates noticeably. In fact, this could have been supported empirically in other trials (67,81). Therefore, in this study, although not intended as incentives, the intervention modules may have contributed to good user engagement, because of potential symptom improvement occurring through the usage. The finding of a more favourable compliance could have also been driven by the chosen study design with a small sample size that enabled a frequent mail contact (regarding study procedure, concerns and questions) and possibly increased the involvement of the

Study I

participants. Moreover, the significance of the EMA data for research was re-emphasized before and during the study.

The THI changed statistically and clinically significant from baseline to final visit with a large effect size (Cohens $d = 1.1$). Meta-analysis of CBT applied to tinnitus report effect sizes between 0.54 and 0.91 (36,37). However, caution should be taken in the interpretation of this effect. First the THI change occurred during the baseline and the intervention phase and can thus not be solely attributed to the intervention. Second, data come from a pilot study with a small sample size and no control group. Larger, controlled samples are needed to validate the effect on THI. Regarding the individual level, 13 of 18 (72%) improved clinically meaningful ($\Delta\text{THI} \geq 7$) (133). This reflects the well-known heterogeneity of treatment response in tinnitus patients, i.e. tinnitus distress decreases only for some individuals, but not for the whole sample. This could also be observed in visualization of individual symptom courses and in comparisons of baseline phase with end of intervention phase. Future research should aim to identify predictors of response to treatment, especially for app-based interventions (137). Previous evidence showed that the positive correlation between working alliance and treatment outcome known from face-to-face psychological therapy remains stable for internet-based treatments (138). Moreover, the acceptability of internet-based interventions is a considerable predictor of treatment outcomes (139,140). Hence, these factors should be considered in subsequent studies testing app-based tinnitus treatments.

The main analysis focus was on the EMA data (tinnitus distress and loudness). The heterogeneity between the cases and the fluctuation of symptom courses hampered identification of symptom improvement from visualization. High fluctuation has already been observed in prior EMA studies measuring tinnitus symptoms using mobile devices and reflects their usual time course (33,65,72). Fluctuations in tinnitus symptoms are probably influenced by a combination of many factors, including neurobiological markers, circadian rhythm, behavioural and emotional processes, and stress (66,73,75,76,141).

To overcome the obstacle of heterogeneity and symptom fluctuation, individual means and standard deviations were calculated for the baseline phase and the end of the intervention phase (same time window than baseline phase). On a group level, neither tinnitus distress nor loudness showed a significant change from baseline to the end of intervention phase. On an individual level, however, 5 of 14 cases improved

Study I

clinically meaningful in tinnitus distress ($\Delta\text{Distress} \geq 10$) as well as in the THI score ($\Delta\text{THI} \geq 7$). Thus, 36% of the sample experienced a reduction in tinnitus burden during the intervention. Interestingly, more patients improved clinically meaningful in the THI than in EMA of tinnitus distress (72% vs. 36%). This might indicate that the THI is more sensitive to change than EMA of tinnitus distress. Regarding tinnitus loudness, only 2 of 14 cases (14%) improved clinically meaningful from baseline to the end of intervention phase ($\Delta\text{Loudness} \geq 10$).

As visual inspection of the EMA data suggested a dissociation of tinnitus distress and loudness scores over the course of the study (see Figure 4), correlations scores between distress and loudness were calculated for the baseline phase and for each week of the intervention phase. Figure 5 depicts a descriptive decline in the correlation between tinnitus distress and loudness over the study period. This could reflect a psychological process that was triggered by the information about tinnitus received from SC and the provided tips that helped the patients to better cope with their tinnitus. In other words, the patients have learned to evaluate their tinnitus in a more differentiated way, i.e. more independent of acoustic features and as a result, tinnitus distress became more independent of tinnitus loudness. Prior studies identified tinnitus acceptance and emotional valence as partial mediators of the relationship between tinnitus distress and loudness (73,142). If the app-based intervention induces a change in those mediators could be of interest for subsequent studies.

To predict daily measured tinnitus distress, a mixed effect model was set up including demographics (age, gender), psychological burden (baseline THI and PHQ-9), intervention effects (trend, level) as fixed effects and participants as random intercepts. The model could explain 62% of the variance of the daily tinnitus distress and was checked for overfitting. Cases with higher daily distress tended to be male with higher baseline THI and PHQ-9 scores. Whereas the influence of THI and PHQ-9 on predicting the tinnitus distress score is not surprising, the gender effect was somehow unexpected, as female patients tend to report higher levels of distress than male in other studies (143,144). This discrepancy could be due to our small sample size which may not be representative for the tinnitus population. Regarding the treatment effect, tinnitus distress is significantly reduced by 0.06 points per day (trend effect). Thus, considering a study length of about 100 days (84 days intervention phase, 8-19 days baseline phase), tinnitus distress is reduced by 6 units in total (scale:

Study I

0-100). This effect is rather weak when compared with the THI which was reduced by almost 12 units (scale: 0-100). The level effect did not reach statistical significance.

Importantly, both the trend effect in tinnitus distress and the THI effect consider the whole study period, i.e. baseline and intervention phase. Thus, presumably in some patients, tinnitus distress improved already to some extent during the baseline phase. This may indeed explain why no group effect was found for tinnitus distress (EMA), as the whole baseline phase was summarized as a contrast to the end of the intervention phase. An improvement during baseline phase could describe an anticipation effect, that is participants experience symptom improvement before the intervention has started because they expect relief from the upcoming intervention (145). Another possibility is that the daily reflection of one's own emotional state through EMA somehow contributed to the improvement in tinnitus distress, although prior findings didn't assume any influence (33,67,72).

Finally, we compared the improvement in THI (pre-post clinical questionnaire) with improvement in tinnitus distress (EMA). Δ THI was strongly associated with the Distress trend ($r = -0.75$, $p < 0.01$) and Δ Distress ($r = 0.86$, $p < 0.001$). Despite high correlations, previous findings suggested that the THI could be more sensitive to change than EMA of tinnitus distress. Different thresholds for retrospective questionnaires and EMA might be appropriate. Even though, further validation research is needed to examine whether EMA is feasible to detect relevant changes in tinnitus symptoms and to validate clinical trials. In depression research, EMA has already been compared to standard assessment instruments (146,147). In those studies, it was shown that EMA did not perform worse than paper-pencil questionnaires using the same items with respect to sensitivity to change. The authors even argued that assessing symptoms with EMA methodology may be more accurate than pre-post questionnaires that reflect just a random time point overlooking the complex clinical picture. The high symptom fluctuation observed here underlines this point. Note that for the Distress trend only the intervention phase was considered, because we did not expect any relevant symptom changes during baseline phase. Thus, an improvement during baseline phase could explain the slightly weaker correlation of the Distress trend with Δ THI than Δ Distress revealed.

The results in this study should be interpreted alongside considerations of its limitations. First of all, it is a pilot study with a small sample size testing the app-based intervention for the first time, thus, further studies with larger samples sizes are

Study I

required to confirm its efficacy. Furthermore, due to the online and app-based study environment, data assessment and app usage could not be controlled with respect to correct app usage, user identity, and circumstances of EMA recordings. The latter should be paid special attention, because the given situation might influence the perception of tinnitus, at least its loudness. Considering future research, an implemented sensor detecting the ambient sound level might reduce this bias. Although the natural study setting leads to reduced internal validity, it increases ecological validity on the other hand. Moreover, there was no THI assessment after the baseline phase before the intervention phase started to distinguish effects between the phases. Furthermore, we have no data on long-term outcomes nor from severely affected patients. Nonetheless, based on available evidence one could speculate, that long-term outcomes and the inclusion of more severely affected patients would have increased the observed effect: counseling takes some time to evolve its whole effect and higher baseline scores in tinnitus questionnaires predict better treatment outcomes (112,114,118). Regarding the EMA subset, we had to exclude 4 more patients because of data MNAR. This should be considered when interpreting and comparing group effects on THI and tinnitus distress, although the data sets are comparable regarding participant demographics and tinnitus characteristics.

In conclusion, we found good treatment compliance, a meaningful improvement in tinnitus distress and tinnitus handicap for part of the sample and a weakening of the relationship between tinnitus distress and loudness over the course of the intervention. Also, there is a strong correlation between improvement in THI and improvement in EMA of tinnitus distress, however, sensitivity to change could differ and separate thresholds might be appropriate. Importantly, these results should be interpreted alongside the fact that the treatment is unguided, resource-efficient, location-independent and immediately available. On a large scale, it could provide relief for a considerable number of tinnitus patients without any additional cost or effort.

Acknowledgments

We want to thank Susanne Staudinger and Jan Musman who helped implement and conduct the study. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme, Grant Agreement Number 848261 (ME, JS, CV, SS, MS, SW, RP, BL, WS). TP didn't receive any funding. The

Study I

funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors have declared that no competing interests exist.

Study II: Understanding Tinnitus Symptom Dynamics and Clinical Improvement through Intensive Longitudinal Data

Milena Engelke¹, Jorge Piano Simões², Laura Basso¹, Nina Wunder¹, Berthold Langguth¹, Thomas Probst³, Rüdiger Pryss⁴, & Winfried Schlee^{1,5}

¹ Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany

² Department of Psychology, Health and Technology, University of Twente, Enschede, The Netherlands

³ Division of Psychotherapy, Department of Psychology, Paris Lodron University Salzburg, Salzburg, Austria

⁴ Institute of Clinical Epidemiology and Biometry, University of Würzburg, Germany

⁵ Institute for Information and Process Management, Eastern Switzerland University of Applied Sciences, St. Gallen, Switzerland

The following text, tables and figures are copy-edited versions of the final article published on the 14th of January 2025 in npj Digital Medicine:

Engelke M, Simões JP, Basso L, Wunder N, Langguth B, Probst T, et al. Understanding tinnitus symptom dynamics and clinical improvement through intensive longitudinal data. npj Digit Med 2025;8;27. <https://doi.org/10.1038/s41746-024-01425-w>

The final publication is available at:

<https://www.nature.com/articles/s41746-024-01425-w>

Copyright: © 2025 Engelke et al. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

Abstract

Intensive longitudinal sampling enhances subjective data collection by capturing real-time, dynamic inputs in natural settings, complementing traditional methods. This study evaluates the feasibility of using daily self-reported app data to assess clinical improvement among tinnitus patients undergoing treatment. App data from a multi-center randomised clinical trial were analysed using time-series feature extraction and nested cross-validated ordinal regression with elastic net regulation to predict clinical improvement based on the Clinical Global Impression – Improvement scale (CGI-I). With 50% app compliance (N = 129, 8480 entries), the model demonstrated good fit to the test data (McFadden $R^2 = 0.82$) suggesting its generalizability. Clinical improvement was associated with linear declines in tinnitus-related thoughts, jaw tension, tinnitus loudness, increases in happiness, and variability changes in tinnitus loudness and distress. These findings suggest that daily self-reported data on tinnitus symptoms is sensitive to treatment response and provides insights into specific symptom changes that occur during treatment.

Introduction

Tinnitus, characterized by the perception of an unspecified sound, is a complex and common condition affecting over 740 million adults globally according to recent estimates (12). In its severe form it can be accompanied by sleep disturbances, concentration difficulties, memory problems, anxiety and depression, leading to a considerable impairment in quality of life (7,90). Severe tinnitus is estimated to afflict more than 120 million individuals globally (12). In European countries, severe tinnitus imposes profound individual burden with around 1500€ in annual out-of-pocket expenses, as well as socioeconomic burden with approximately 4800€ in annual socioeconomic costs per patient (148). Efforts have been made to alleviate both health-related and economic burden by the design and implementation of various treatment strategies such as psychological, auditory and pharmacological treatments (18). Yet, the efficacy of many treatment options remains unclear, which is often attributed to the diverse etiological and clinical manifestation of tinnitus, often referred to as tinnitus heterogeneity (18,90,93).

However, methodological aspects are often overlooked. In psychosomatic and psychiatric research, treatment outcomes are usually assessed with clinical interviews or self-report measures at selected points in time (149). Such assessments may fail to

Study II

capture the dynamic nature of symptoms as they fluctuate over time and across different contexts. Further, patients may struggle to accurately recall the intensity, frequency, and impact of symptomatic episodes, leading to potential underreporting or overreporting of symptoms (46,150).

Intensive longitudinal sampling of symptoms has thus emerged as a complement to traditional measurement methods (141,151,152). A wide range of methods have been framed under the term Experience Sampling or EMA (34). EMA involves (close to) real-time, repeated collection of data in a patient's natural environment, providing a more granular and immediate perspective on symptomatology (34,46). Daily diaries or EDDs involve a single assessment per day at fixed intervals, typically in the evening (34). Although not capturing momentary experiences, EDDs are considered part of the EMA framework due to their repeated administration, which enables a dynamic investigation of the variables of interest (34). A key advantage of EDDs is their reduced participant burden, however, their retrospective component potentially causes the patient's end-of-day condition to influence their evaluation of the entire day. Empirical research comparing momentary EMAs throughout the day with EDD data indicated that EDD depicted a more negative picture of certain tinnitus symptoms (69). However, these effects were small and of uncertain clinical relevance. The authors argue that the slight differences favor the use of EDD over EMA, considering the additional burden imposed by EMA (69). While EMA is better suited for capturing within-day fluctuations, EDD may be a valid alternative for long-term monitoring. In this manuscript, we adopt the term EMA to encompass a range of designs, including EDD approaches, as outlined in work (34).

Already 15 years ago, EMA has been discussed as a powerful tool to investigate symptom dynamics and their environmental interactions to enhance understanding of psychopathological mechanisms and treatment response (153). During the last years, research efforts have strongly increased (49). With respect to the evaluation of treatment effects, EMA was found to be sensitive to change (147), to enable a detailed picture of symptom progression (154), to detect treatment and side effects at an early stage (155,156), to provide predictive baseline information (157,158), to reveal interactive effects between symptoms (159), and to match clinical criterion measures (146,160–162).

Despite growing research interest, EMA data should nevertheless be used with caution. Even if momentary or daily self-reports reduce memory biases, EMA answers

Study II

are likely to be influenced by other cognitive heuristics driven by the current context, assumptions, judgements, comparison standards and interpretation of items (63,163). Stone and colleagues proposed that EMA and retrospective measurements should not be inherently opposed, but rather, EMA should be viewed as complementary. The choice between these methods should be guided by the theoretical framework underlying the research construct (63).

Compared to the widespread use of EMA in other health disciplines, the audiological research field is lagging behind (164,165). Among tinnitus experts, there have been concerns that symptoms could potentially worsen with repeated questioning. Conversely, regular recording of symptom severity might empower patients, giving them more control over their symptoms by better understanding influencing factors (49,166). Empirical findings have not shown any influence of long-term EMA on tinnitus distress or loudness but revealed the fluctuating nature of tinnitus symptoms and its influence by emotional and environmental factors (33,65,66,72,73,75,76). Further, it has been used to predict the fluctuations of tinnitus by tinnitus-unspecific dimensions such as mood or concentration and the progression of symptoms using neighborhood data (84,86). Preliminary results regarding clinical utility showed that EMA data unveiled a descriptive decrease in the correlation between tinnitus distress and loudness throughout the duration of an app-based treatment (Study I). Further, changes in questionnaire-measured tinnitus distress were associated with trends in EMA-measured tinnitus distress (Study I).

Thus, while EMA excels in capturing the ups and downs of daily experiences and has shown initial indications of clinical utility, it remains unclear how clinical improvement manifests within EMA data. To address this question, clinical improvement is operationalized using the patient-reported CGI-I (167). The CGI-I is a commonly used 7-point ordinal scale for assessing the subjective degree of change after treatment. It ranges from 1 (very much improved) via 4 (no change) to 7 (very much worse) (168). In this study, we define clinical improvement as achieving CGI-I scores from 1 to 3, reflecting “very much improved” (1), “much improved” (2) or “minimally improved” (3). Further, feature-based time-series analysis is applied to daily self-reported app data (EMA and EDD questions) which involves extracting meaningful features or characteristics from time-series data (169). Feature-based time-series analysis has been applied in various domains, including finance, healthcare, environmental monitoring, and manufacturing. In clinical research, features have been

extracted from passively generated time-series data such as movement patterns and vital parameters to predict diagnostic status, change in symptom severity and treatment demands.

The primary objective of this study is to identify indicators of daily self-reported tinnitus (Table 1) that characterize clinical improvement. More specifically, we aim to analyze which symptoms change and how these changes manifest through time-series features (Table 6) in patients who respond to treatment. Additionally, we seek to assess the impact of missing app data on the model fit.

Table 6. Description of Time-Series Features

Time-series feature	Description
crossing-points	The number of times a time series crosses its median (high values indicate frequent median crossing).
curvature	The curvature of the trend component (STL decomposition). It is based on the coefficient from an orthogonal quadratic regression applied to the trend component (negative values indicate a concave curve, 0 indicates no curvature; positive values indicate a convex curve).
entropy	Forecastability of a time series (Shannon entropy; high values indicate difficulty to forecast).
flat-spots	Number of sections of the data where the series is relatively unchanging (high values indicate constancy).
hurst	Measure of long-term memory of a time series (long-term memory = significant autocorrelations for many lags; high values indicate many autocorrelations).
linearity	Linearity of the trend component (STL decomposition). It is based on the coefficient of a linear regression applied to the trend component (negative values indicate a negative linear trend; 0 indicates no trend; positive values indicate a positive linear trend).
lumpiness	Variance of the variances of non-overlapping windows (length of windows: 10 data points; high values indicate changes in variance).
nonlinearity	Modification of the statistic used in Teräsvirta's nonlinearity test (large values = nonlinear, values around 0 = linear).
stability	Variance of the means of non-overlapping windows (length of windows: 10 data points; high values indicate changes in mean).
trend	Strength of the trend component (STL decomposition): $1 - \frac{Var(remainder)}{Var(remainder) + Var(trend)}$ (high values indicate strength trends).

Note. Feature extraction was performed with the *theft* R package based on the *tsfeatures* R package. STL decomposition (Seasonal and Trend decomposition using Loess): decomposes a time series in a smoothed trend component, a seasonal component and a remainder component based on a Loess function. Further description of the features: <https://cran.r-project.org/web/packages/tsfeatures/vignettes/tsfeatures.html> (retrieved on the 29.04.2024)

Methods

Data were drawn from a multi-centre parallel-arm RCT that sought to examine the effect of single and combination treatments in patients with chronic subjective tinnitus (39,173). The study was registered at <https://clinicaltrials.gov/study/NCT04663828>. All participants gave written informed consent and ethics approval was obtained from the ethical committees of all five trial sites (University of Regensburg, Regensburg, Germany [combined ethical approval for clinical sites Berlin and Regensburg]; Katholieke Universiteit Leuven, Leuven, Belgium; Ethniko Kai Kapodistriako Panepistimo Athinon, Athens, Greece; Hospital Universitario Virgen de las Nieves/ Hospital Clinico Universitario San Cecilio, Granada, Spain).

Participants

Patients were included if they reported having chronic tinnitus (≥ 6 months) as their primary complaint, had at least mild tinnitus handicap ($\text{THI} \geq 18$), were aged between 18 and 80 years, and had not started any other tinnitus-related treatments in the last 3 months before study start. Patients with objective tinnitus, otosclerosis, acoustic neuroma, acute ear infections, Meniere's disease, severe hearing loss, as well as serious internal, neurological, or psychiatric conditions were excluded.

Study Design and Procedure

The data were collected between April 2021 and December 2022 at five clinical sites across Europe. Participants were randomised to one out of ten treatment arms, which were either single treatments (CBT, HA, SC, ST) or combinations thereof (CBT + HA, CBT + SC, CBT + ST, HA + SC, HA + ST, SC + ST) lasting for 12 weeks. Demographic and clinical characteristics were assessed at baseline (before treatment), interim visit (during treatment after 6 weeks), final visit (after 12 weeks of treatment) and follow-up (after 36 weeks). During the 12-week treatment phase, self-reported tinnitus symptoms were further gathered daily using an electronic EDD via the UNITI smartphone application (126). Additional information about the trial is described elsewhere (173).

Measures

The CGI-I (167) assessed at final visit served as the criterion measure of clinical improvement. In a single query, patients are prompted to assess the overall amelioration of their tinnitus symptoms in relation to the period preceding treatment ("Please rate the total improvement of your tinnitus complaints compared to before beginning of treatment.") using a 7-point Likert scale (1 – very much better, 2 – much

Study II

better, 3 – minimally better, 4 – no change, 5 – minimally worse, 6 – much worse, 7 – very much worse). The validity of the CGI-I as an outcome measure has been examined and it was shown to be sensitive to change (174).

Daily self-reported data on tinnitus symptoms was assessed via an EDD in the UNITI smartphone application, which prompted the patient every evening at 19:30 to rate several EMA and EDD questions on a continuous visual analogue scale (VAS; questions listed in Table 1). EMA questions aimed to assess the condition of the respective moment, while EDD questions were intended to reflect on the whole day. Patients were asked to answer the questions (EMA and EDD) once every day over the 12-week treatment phase, so ideally 84 answered app entries were expected per patient. The assessment could only be submitted once all questions had been answered.

Statistical analysis

Before extracting time-series features from app data, the range of diary answers per patient was calculated. A requirement for meaningful feature extraction and analysis was minimal variability in all app questions; thus, N = 263 patients' time series were eligible for feature extraction (RCT sample size N = 461). We selected 10 features out of a pool of six open-source time-series feature sets (included in the R package *theft*) (175). On one hand, feature selection was based on capturing the dynamics of change as it was expected that symptoms would change as a result of the clinical intervention. On the other hand, the selection was based on citations in the clinical feature-based time-series literature to identify the features that have already been shown to be relevant (see Table 6 for an overview and description of the selected features) (170,171). Each patient generated 10 time-series (10 app questions), thus, extraction of 10 time-series features led to 100 features describing each patient (10 app questions [Table 1] x 10 features [Table 6]). There was no imputation of missing data before feature extraction.

In the second step, we performed a sensitivity analysis to identify the optimal amount of completed diary entries to maximize the model fit of predicting treatment response in our dataset. Due to real-world data usage, the higher the compliance rate, the smaller the sample size available for analysis. We fitted ordinal logistic regression models with a nested cross-validated elastic net penalty for minimal compliance rates between 50% and 90% in 5% steps (i.e., 9 regression models; results are reported in the Figures S1-8). Due to the high number of potentially correlated predictors, elastic

Study II

net was chosen to perform variable selection and shrinkage. CGI-I at final visit was used as outcome variable. Predictors included the 100 scaled times-series features which have been described above. To minimize overfitting, 5-fold nested cross-validation was performed to tune the elastic net across 20 different lambda values (λ) based on the best Akaike Information Criterion (AIC) fit (R package ordinalNet, function ordinalNetCV, method is based on example 5 in work (176)). The nested cross-validation procedure enables a performance evaluation of the tuned model on data that has not been used to train the model. First, the data is split into five folds, each time leaving out the test dataset. Ordinal elastic net regression ($\alpha = 0.5$) is performed on each of the five training datasets on a sequence of 20 λ values (default settings; $\lambda_{\max}$ equals the smallest value that sets every coefficient to zero). The best AIC fit is selected on each fold and the out-of-sample prediction is then assessed on the hold-out test set. The λ value with the best out-of-sample log-likelihood of the five folds is used in the final model to obtain the coefficient estimates. The performance measures on the test set are taken to obtain model fit. This was determined by McFadden pseudo R^2 that compares the log-likelihood of the full model with the log-likelihood of the null model (intercept-only model) (177).

In the third step, the sample with the compliance rate demonstrating the best prediction was subject to further investigation. Spearman correlation coefficients were computed between CGI-I and time-series app features. This served as a first description of the association without reporting any p-values, therefore correction for multiple testing was not applied. The cumulative model was fitted reversed to ensure consistent sign between correlation and regression results (cumulative probabilities $P(Y \geq 2), \dots, P(Y \geq K + 1)$; with $K + 1$ response categories). The coefficients of the ordinal logistic regression are displayed as log odds, which is the logarithm of the odds ratio (probability of success/probability of failure). A positive coefficient means that as the predictor variable increases by one unit, the log odds of being in a higher category (less clinical improvement) increase. In order to deconstruct the results obtained from the regression analysis and identify the relevance of both time-series features and app questions in measuring clinical improvement, absolute values of the log odds were summed for the respective app questions across the time series features and vice versa. All analyses were performed in R (version 4.2.2).

Results

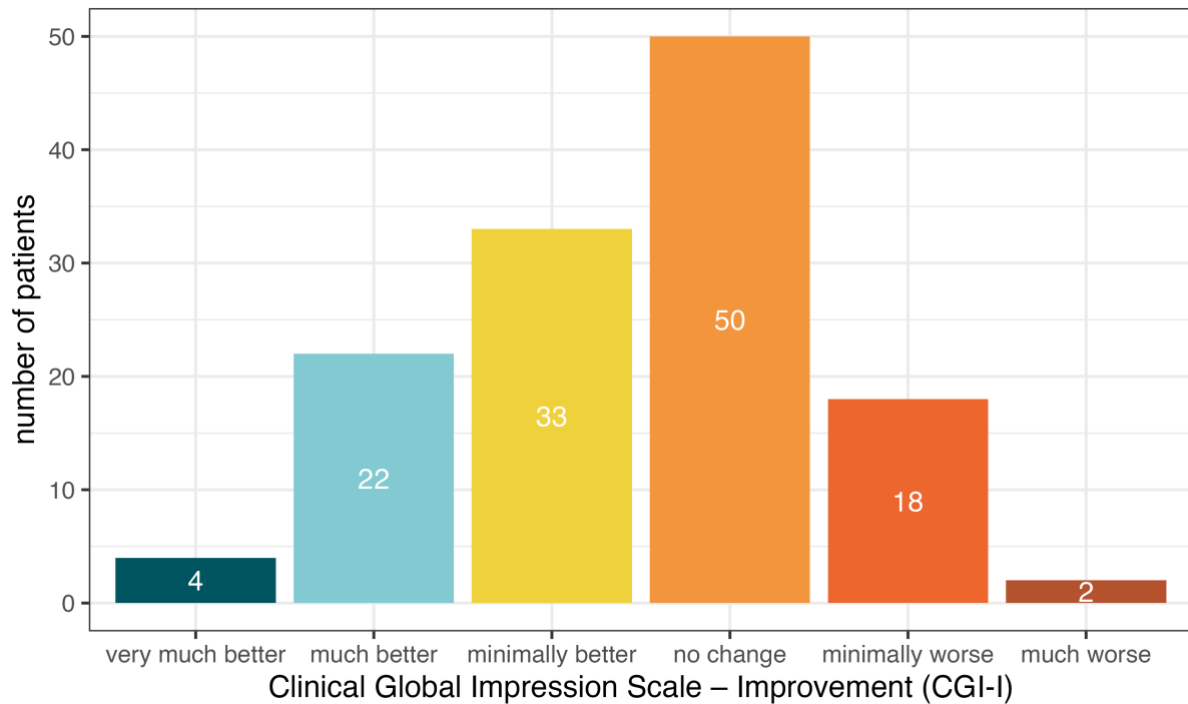
Sample description

The sensitivity analysis revealed that the 50% compliance subsample (corresponding to a minimum of 42 diary entries per patient) yielded the best model fit together with the high compliance subsamples (McFadden pseudo $R^2 = .82$). In this sample, $N = 129$ patients generated 8480 diary entries. Patients were 47% female, on average 55 years old with a mean tinnitus duration of 132 months and a moderate tinnitus handicap (see Table 7). At final visit, 39% ($N = 50$) indicated no change in their tinnitus complaints compared to before treatment, 26% ($N = 33$) improved minimally, 17% ($N = 22$) indicated good and 3% ($N = 4$) very good treatment response. The remaining patients indicated minimal (14%, $N = 18$) and more severe (2%, $N = 2$) worsening in their tinnitus complaints, while no patient indicated greater deterioration (see Figure 7; because of rounding the sum does not precisely equal 100%). A comparison of the 50% compliance subsample with the entire RCT sample ($N = 461$) is reported in the Supplementary Material (Table S3); patients in the subsample were older (RCT sample: Age [years] = 51.1 ± 12.4 [mean $\pm$ SD]).

Table 7. Patient Characteristics at Baseline ($N = 129$)

Participant characteristics	
Sex – no. (%)	
Female	61 (47.3)
Male	68 (52.7)
Age (years)	54.8 ± 12.1
Tinnitus duration (months)	132 ± 117
THI score	48.6 ± 19.6
PHQ-9 score	7.7 ± 5.0
Filled-in diary entries (in %)	78.3 ± 14.0

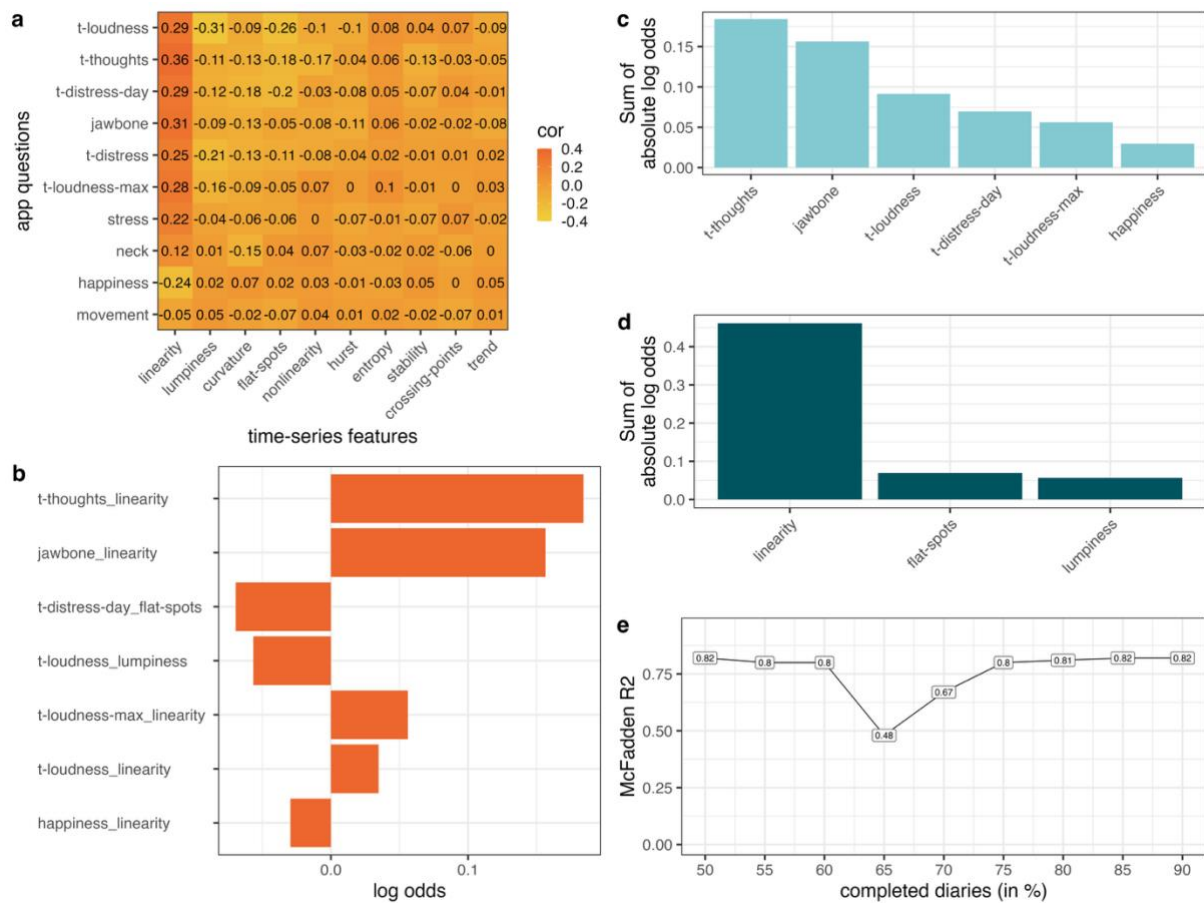
Note. Plus-minus values are means $\pm$ SD. THI = Tinnitus Handicap Inventory; PHQ-9 = Patient Health Questionnaire for Depression.

Figure 7. Distribution of CGI-I

Note. Distribution of clinical improvement operationalized by the CGI-I. N = 129. CGI-I is answered on a 7-point Likert scale (1 – very much better, 2 – much better, 3 – minimally better, 4 – no change, 5 – minimally worse, 6 – much worse, 7 – very much worse).

Correlation of clinical improvement with app features

After the sample had been identified based on the sensitivity analysis, the association of the app features (time-series features of the app questions) with clinical improvement was explored. Figure 8a illustrates the Spearman correlation between the CGI-I and the 100 app features (rows and columns are ordered by mean correlation coefficients of time-series features and app questions respectively). The top three highest correlation coefficients pertain to linearity of tinnitus-related thoughts ($\rho = 0.36$), linearity of jaw tension ($\rho = 0.31$), and lumpiness of momentary tinnitus loudness ($\rho = -0.31$). Thus, patients indicating clinical improvement had linearly decreasing thoughts about their tinnitus and jaw tension as well as changing fluctuations in tinnitus loudness. The highest mean correlation with the CGI-I was found for momentary tinnitus loudness, tinnitus-related thoughts and daily tinnitus distress among the app questions and for linearity, lumpiness, and curvature among the time-series features.

Figure 8. Clinical Improvement measured by Time-Series Features of the App

Note. Clinical improvement measured by time-series features of the app questions. a) Spearman correlation coefficients between CGI-I and app features (rows and columns are ordered by mean correlation coefficients). b) Log odds of the ordinal logistic regression with elastic net regularization and 5-fold nested cross-validation. Outcome: CGI-I, Predictors: App features. McFadden $R^2 = 0.82$. c/d) Log odds were derived from an ordinal logistic regression predicting CGI-I by app time-series features. e) Model fit of predicting CGI-I by app features according to the proportion of completed diary entries. The higher the minimal proportion of completed questionnaires, the smaller the N. Minimum 50% diary entries (N = 129), 55% (N = 125), 60% (N = 113), 65% (N = 102), 70% (N = 91), 75% (N = 75), 80% (N = 64), 85% (N = 53), 90% (N = 37).

Prediction of clinical improvement by app time-series features

Ordinal logistic regression with elastic net penalty was fitted to identify which app questions and time-series features predicted clinical improvement (outcome: CGI-I, predictors: app features). Nested cross validation resulted in an optimal λ of 0.18. This was used to fit the final model which selected 7 out of 100 possible app features (see Figure 8b, Table S4 for the exact coefficients; the remaining features were shrunk to zero). Positive log odds indicate that clinical improvement (i.e. CGI-I values 1-3) is associated with negative values in app time-series features. Negative log odds indicate that clinical improvement is associated with positive values in the respective app time-series feature. Thus, clinical improvement is associated with linearly decreasing tinnitus-related thoughts, jaw tension, maximum tinnitus loudness, and momentary

Study II

tinnitus loudness. Similarly, clinical improvement is associated with linear increases in happiness. Further, clinical improvement is associated with more flat spots in daily tinnitus distress (i.e. higher constancy) and greater lumpiness in momentary tinnitus loudness (i.e. changing fluctuations) according to the model results (see Table 6 for a description of the features). The model achieved a good fit on the test dataset from nested cross-validation (McFadden $R^2 = 0.82$).

Relevance of app questions and time-series features

To enhance comprehension of the significance of both the app questions and the time-series features in predicting the CGI-I, the sum of the absolute log odds was computed for each app question across the time-series features (see Figure 8c) and for each time-series feature across the app questions (see Figure 8d). Among the app questions, the highest predictive relevance for clinical improvement was observed in tinnitus-related thoughts followed by jaw tension, momentary tinnitus loudness, daily tinnitus distress, maximum tinnitus loudness, and happiness. Among the time-series features, linearity was most important for predicting clinical improvement followed by flat-spots and lumpiness. *Linearity* is the fluctuation-corrected linearity of the trend, with negative values indicating negative linear trends and positive values indicating positive linear trends. *Flat spots* is an indicator of constancy, higher values indicate more constancy. *Lumpiness* is an indicator of the change in variance, high values indicate more changes in variance (could be both more or less variance).

Sensitivity analysis

Figure 8e illustrates the model fit, as assessed by McFadden R^2 , for models with varying compliance rates. McFadden R^2 remained approximately constant around 0.8 for compliance rates ranging from 50% to 90% of app entries with a drop at 65% and 70%. The coefficients of the remaining models are reported in the Figures S1-8. Higher lambda values shrink the coefficients stronger which leads to smaller odds ratios and smaller selection of features. Those models are less prone to overfitting and demonstrate better fit to the data, as indicated by higher McFadden R^2 on the test sets. Certain features remain stable across the models with different compliance rates (e.g., t-thoughts_linearity and jawbone_linearity). Notably, due to the small sample size, subsamples with higher compliance rates could not be tested.

Discussion

Our findings underline the feasibility of using daily self-reported symptoms to measure clinical improvement in tinnitus patients undergoing treatment which is in line with findings from other health research fields (49,152,162).

We were able to show how treatment response is reflected longitudinally in tinnitus symptoms. Based on our regression model, patients were more likely to improve if they had a linear decline in tinnitus-related thoughts, jaw tension, maximum tinnitus loudness, and momentary tinnitus loudness, a linear increase in happiness as well as constancy in daily tinnitus distress and a change in fluctuation in momentary tinnitus loudness. The model showed a good fit to the test data, suggesting its generalisability. The results are mostly, but not completely consistent with the correlation analysis which is probably caused by collinearity issues that have been addressed by the elastic net algorithm. Notably, the correlation coefficient of flat spots in daily tinnitus distress with the CGI-I is the smallest among the selected features ($\rho = -0.2$).

The most predictive time-series feature for clinical improvement was linearity which is the linearity of the smoothed trend component corrected from high-frequent fluctuations such as noise and seasonality. Thus, not the magnitude of symptom reduction, but the linearity of the trend was relevant. This suggests that patients may prefer steady progress, with high fluctuations disrupting their sense of improvement. Additionally, though less influential, features such as flat spots and lumpiness were also linked to improvement. Flat spots in tinnitus distress, which indicates constancy, was positively associated with improvement, suggesting that constancy in tinnitus distress is experienced as more beneficial than fluctuations in tinnitus distress. In contrast, lumpiness in tinnitus loudness, which reflects variability in fluctuations, was positively associated with improvement, suggesting that more variability in loudness fluctuation predicts a beneficial outcome. The joint consideration of the other features suggests that a strong fluctuation of symptoms during treatment is unfavourable for treatment response. Thus, we hypothesize that variability in loudness fluctuation unfolds as a reduction in loudness fluctuation. However, it is not specified by the feature and requires further investigation. These findings align with an earlier study investigating the feasibility of using EMA to detect treatment effects in fibromyalgia patients. They found that next to a decline in average pain, a reduction in pain variability contributed incrementally to clinical improvement (162).

Study II

Looking at which symptoms change, most predictive for clinical improvement were changes in tinnitus-related thoughts, jaw tension, momentary tinnitus loudness, daily tinnitus distress, maximum tinnitus loudness, and happiness (relevance in descending order). Those symptoms that are usually applied as outcome measures (31), i.e. tinnitus distress and loudness, are accompanied by psychological and somatic symptoms. This finding adds an interesting dimension to the understanding of clinical improvement in tinnitus, as it underscores the importance of considering general health symptoms alongside tinnitus-specific indicators. While the main mechanism of established tinnitus therapies typically involves reducing tinnitus-related distress, it is noteworthy that patients may prioritize loudness reduction or complete relief from tinnitus (178). Thus, the implementation of patient-centered, personalized EMA, wherein clinically relevant symptoms are identified by the patient and assessments are tailored accordingly, holds promise for enhancing clinical diagnostics and care (166).

The sensitivity analysis demonstrated the stability of key predictors (e.g. `t-thoughts_linearity` and `jawbone_linearity`) as well as consistent model fit across varying compliance rates. McFadden R^2 remained stable around 0.8 for compliance rates of 50%-60% and 75%-90%, suggesting robust modeling of clinical improvement using EMA data even in less compliant samples, which often reflect real-world scenarios. However, models with compliance rates of 65% and 70% appeared overfitted, as evidenced by smaller penalization factors (λ), larger coefficients, and lower fit on the test set. It is important to note that the overall compliance in our study was lower than the typical 79% seen in pure EMA studies (179), likely because the daily diary was a secondary aspect of the clinical trial. A comparison of the 50%-compliance subsample with the entire RCT sample revealed variation only in age, while other baseline characteristics and CGI-I distributions were comparable. Although our findings highlight the feasibility of obtaining robust models from less compliant samples, efforts to increase compliance rates in EMA remain essential. Higher compliance not only improves data quality and statistical power but also enhances representativeness of the sample. EMA experts recommend strategies like improving instructions, encouraging questions, practice, and providing feedback to enhance both compliance and data quality (63).

The following aspects should be taken into account when evaluating the results. The CGI-I was used to assess clinical improvement which asks patients to evaluate

the overall impact of the intervention on their tinnitus. This method has both benefits and limitations. It values patients' self-assessment, treating them as experts on their own condition. However, its retrospective nature relies on patients recalling their pre-therapy state, which can introduce biases. Incorporating clinician-rated assessments alongside patient self-reports could provide a more balanced evaluation, capturing both subjective experiences and clinical evaluations (180). Nevertheless, the "true" measure of treatment efficacy remains elusive in subjective assessments, highlighting the need for a biomarker to gauge the severity of the condition. Further, the features were pre-selected based on an emphasis on capturing the dynamics of change and based on the relevant literature (170,171); a different feature selection could have led to different results. We acknowledge that the use of a linear predictive model, such as logistic regression, may influence feature selection, favoring features with clear linear relationships to the outcome. However, the selected features align with clinically meaningful patterns of improvement. Also, EMA questionnaires need to be carefully evaluated in terms of their clinimetric properties (psychometric criteria transferred to clinical outcome measures). This includes sensitivity (e.g., ability to detect treatment effects, ability to differentiate active treatment from placebo), clinical validity (ability to discriminate between subjects with and without a condition e.g. using cutoff scores), predictive validity (ability to predict treatment response and clinical outcomes), and incremental validity (each item provides distinctive clinical information) (181). Additionally, there is a need for discussion whether clinimetric criteria should be updated to accommodate EMA measures, given that these criteria were originally introduced for traditional questionnaires.

In summary, we provided evidence that EMA is sensitive for change across a range of tinnitus treatments. By longitudinally drawing on the symptom level, EMA was able to show which symptoms change and how they are changing if patients respond to treatment. The latter was realised by the application of time-series features, a computational method that characterizes dynamics of time-series which are not visible to the eye. This allows for an easily applicable, understandable and effective utilization of longitudinal data to understand time dynamics without compromising valuable information or imputing missing data (169). In comparison to questionnaires that often rely on total scores (151), the use of EMA data not only allows treatment effects to be described at the symptom level, but could facilitate an exploratory investigation into which specific symptoms are impacted by a particular treatment and are thus paving

the path towards personalized treatments. Finally, questionnaires depict only few snapshots in time which limits their ability to measure the longitudinal progress of symptoms (146). This study demonstrated the complexity of symptom courses during treatment and the importance of longitudinal data, as linearity and fluctuations were identified to be predictive for treatment success.

Our findings might stimulate both future research and clinical practice in tinnitus. We encourage researchers to replicate our findings and expand them, e.g. by investigating individual differences, early treatment response or different feature sets. Importantly, we need to explore mechanisms to substantially improve compliance towards EMA protocols within clinical studies. Further, while passive sensing has been explored concerning the prediction of clinical status and change (170,171), efforts could be made to investigate whether combining EMA with passive sensing enhances the prediction and understanding of treatment response (152,182). In clinics, EMA could offer a dynamic approach to monitoring patient progress in real-time and reduce costs by the reduction of in-person visits. Clinicians can incorporate EMA techniques into routine clinical assessments, allowing for ongoing tracking of key variables like general stress levels, tinnitus distress, and emotional states. This enables timely adjustments to treatment plans based on individual patient needs and fluctuations in symptoms.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme, Grant Agreement Number 848261. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

RP is affiliated with npj Digital Medicine as an editor. All other authors declare no financial or non-financial competing interests.

Study III: Uncovering Latent Trajectories of Daily Tinnitus Symptoms through App-based Monitoring During Treatment

Milena Engelke¹, Berthold Langguth¹, Thomas Probst², Rüdiger Pryss³, Stefan Schoisswohl¹, Jorge Piano Simões⁴, Peter Ten Klooster⁴, Beat Toedtli⁵, Carsten Vogel³, & Winfried Schlee^{1,5}

¹ Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany

² Division of Psychotherapy, Department of Psychology, Paris Lodron University Salzburg, Salzburg, Austria

³ Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

⁴ Department of Psychology, Health and Technology, University of Twente, Enschede, The Netherlands

⁵ Institute for Information and Process Management, Eastern Switzerland University of Applied Sciences, St. Gallen, Switzerland

The following text, tables and figures were submitted to the journal *Internet Interventions* on the 26th of September 2025 and are currently under review.

© Authors retain copyright.

Abstract

Tinnitus heterogeneity is well-documented across phenotypes, etiologies, risk factors, comorbidities and associated burden. However, variability in treatment response remains insufficiently explored and often masked by the group-level comparisons of clinical studies. Moreover, little is known about the temporal trajectories of symptoms during treatment. Longitudinal monitoring via smartphones using Ecological Momentary Assessment provides rich inter- and intraindividual data on fluctuations and trajectories of symptoms. In this study, we investigated whether individual 12-week trajectories of daily self-reported tinnitus symptoms during treatment could be meaningfully sub-grouped. 147 patients provided 9634 observations while undergoing single or combined applications of hearing aids, cognitive-behavioural therapy, structured counseling, and sound therapy. A four-class growth mixture model best fit the data. One class was characterized by an increase in tinnitus symptoms over time (18%), another showed stable symptom trajectories (40%), while the remaining two classes described symptom reductions with different onsets during treatment (early improvement: 20%; late improvement: 21%). The identified classes did not differ in baseline characteristics, indicating that this information could not be used to predict symptom trajectories. Additionally, all four classes were represented in nearly each treatment arm. Notably, retrospective patient-reported outcome measures (PROMs) did not consistently align with latent symptom trajectories. These results underscore the heterogeneity and non-linearity of symptomatic change both within and across treatment modalities. We propose that app-based trajectories reveal details about symptom improvement that cannot be seen in standard PROMs.

Introduction

Tinnitus is an auditory phantom condition which is perceived as a tonal or composite noise in the absence of an external sound source (90). With a prevalence of 14% among adults and increasing rates with age, it is a common condition especially among the elderly (12). Otological and non-otological factors could have been established as risk factors (14). Tinnitus is characterized by its high heterogeneity in the percept itself, associated etiologies and comorbidities, in tinnitus-triggered distress as well as in response to treatment (8).

While tinnitus heterogeneity has been well documented across perceptual characteristics, risk factors, comorbidities and severity (8), variability in treatment

response remains less understood. There is consensus that no single treatment is universally effective in reducing distress or loudness of the tinnitus percept (23,173). Despite that, clinical studies typically rely on group-level comparisons (183). Although these studies are indispensable for proof of efficacy, this approach can obscure differential effects on individuals or subgroups and may contribute to inconsistent evidence observed across studies (184,185). At the same time, there is a subgroup of patients that benefit from treatment in most clinical tinnitus studies. Recent research aims to predict treatment response from extensive patient-level data assuming that outcomes vary with individual tinnitus characteristics and treatment fit (186). In this context, prescriptive predictors inform clinical practice about optimal treatment allocation for specific patients, while prognostic predictors indicate the likelihood of benefit regardless of treatment type. Yet there are no robust, generalizable results yet in tinnitus research due to diverse interventions and methodologies (143,187).

Simoes et al. crowdsourced data from 5017 tinnitus patients and found that individual characteristics had limited predictive value, but combined explained 16% of treatment outcome variance (188). Niemann et al. identified and replicated four baseline phenotypes, ranging from low to high tinnitus distress with specific psychological or somatic burden (N = 1228 (189), N = 989 (190)). They further described five treatment response subgroups, from deterioration to high improvement, whose distribution varied by baseline phenotype. Yet, predictive utility remained limited, highlighting the complexity of treatment trajectories (190).

Most clinical studies rely on pre- and post-treatment assessments by PROMs, however, tinnitus fluctuates over time depending on internal regulatory processes such as emotion, stress and circadian rhythm (33,70). Further, continuously measuring treatment effects in daily life is rare, yet preliminary evidence describes diverse, non-linear tinnitus trajectories (191–193). Longitudinal research in neighbouring mental-health fields has shown that patients can follow diverse therapeutic paths (194–196). Capturing tinnitus symptom evolution throughout treatment at finer time scales might reveal diverse symptom trajectories as yet another layer of tinnitus heterogeneity.

In this study, we use intensive longitudinal data gathered in daily life by EMA and Growth Mixture Modeling (GMM) to determine whether individual trajectories cluster into unobserved classes. This statistical technique identifies latent classes based on individual developmental trajectories and has been used to uncover

heterogeneity in treatment courses in other mental health fields (194–196). We then describe these classes by their course, size, and patient characteristics to advance understanding of heterogeneity in treatment response.

Methods

Study Design

This exploratory analysis used data from a multicentre RCT on chronic tinnitus conducted between April 2021 and December 2022 in Germany, Greece, Spain, and Belgium (173). The study was approved by the local ethics committee at each site. Patients were randomised to one out of ten treatment arms which were group-based CBT focusing on tinnitus exposure, HA (for patients with a HA indication only), App-based SC, App-based ST and the following combinations CBT + HA, CBT + SC, CBT + ST, HA + SC, HA + ST, and SC + ST. Detailed information on the study design can be obtained from the study protocol (173), the clinical effect of each treatment arm has been published (39).

Subjects

Patients primary complaint was chronic tinnitus (≥ 6 months), they were aged between 18 and 80 years, minimally distressed by their tinnitus (THI ≥ 18) (28), not cognitively impaired (Montreal Cognitive Assessment [MoCa] > 22) (197), and reported stable psychoactive medication for at least 30 days. We excluded patients with objective tinnitus, otosclerosis, acute ear infections, Meniere's disease, serious internal, neurological or psychiatric conditions, drug, medication or alcohol abuse within the last 12 weeks, severe hearing loss or one deaf ear, and if they had started another tinnitus treatment within the previous 12 weeks.

Study Procedure

Patients underwent a baseline visit before they were randomised to a treatment arm. During the 12-week treatment phase, patients were prompted every evening at 7:30 pm by a dedicated smartphone app (126) to fill out a daily EMA questionnaire. The remaining study visits took place after 6, 12, 36 and 48 weeks (in this study, we only use data from baseline and 12-week visit). During those visits, detailed clinical information regarding tinnitus, general health, comorbid diseases as well as audiological data were gathered on-site and online.

Data

We focused on the daily 10 EMA items assessing health-related and tinnitus symptoms (Table 1). Some questions referred to the momentary situation, others to the average over the day. A prior investigation identified changes in *tinnitus-related thoughts* to be best predictive for clinical improvement which was thus chosen as the main outcome (Study II). The distribution of the outcome over all patients and measurement points is shown in Figure S9. The remaining variables predicting clinical improvement were selected to be compared with the predicted trajectories of *tinnitus-related thoughts*. These are *jaw tension*, *momentary tinnitus loudness*, *daily tinnitus distress*, *maximum tinnitus loudness*, and *emotion*. Coding of *emotion* was reversed to *negative emotion* to standardize interpretation that higher scores indicate higher burden.

To describe baseline differences between patients of latent classes we analysed data of tinnitus case history (ESIT-SQ (129), personality (Big Five Inventory-2 [BFI-2]) (198), THI (THI), depression (PHQ-9) (131), and hearing loss (average hearing threshold at the frequencies 500 Hz, 1 kHz, 2 kHz and 4 kHz [PTA-4]).

Data Analysis

Growth Mixture Models

In terms of data preparation, moving averages were calculated with a seven-day time window to smooth out fluctuations and weekly patterns thereby focusing on long-term trends. After visual inspection of the pre-processed outcome, one outlier was removed from the data.

To identify latent trajectories in *tinnitus-related thoughts*, we estimated GMMs (199,200) using the *hlme* function from the *lcmm* package in R. Fixed effects (overall and class-specific) included an intercept and a natural cubic spline function of time in days with three degrees of freedom to allow for non-linear trends over time. A random intercept accounted for subject-level variability within each class with variance constrained to be equal across classes. Models with progressively increasing number of latent classes (1 – 10) were estimated. For models with number of classes ≥ 2 , an automatic grid search with 30 repetitions and 30 iterations was performed to prevent convergence at local maxima. Initialization values were provided from the 1-class model to improve stability and convergence. If the model did not converge, we increased the repetitions to 50. To test whether a linear trend was more appropriate, we repeated the process with a linear function of time in days. We adhered to the Guidelines for Reporting on Latent Trajectory Studies Checklist (Table S5) (201).

Study III

The selection of the final model was based on a triangulation of the relative fit by AIC and Bayesian Information Criteria (BIC), the classification accuracy by the Posterior Probability of Membership (PPM) and Entropy (indicator of the conditional probabilities of individuals' group membership) as well as by the size of the estimated classes and the substantive interpretability and clinical importance of the identified classes. There is no standardized way to identify the optimal result, but model selection should be informed by a variety of several criteria including theory, statistics and past findings (200). Better model fit is indicated by lower AIC and BIC as well as higher PPM and entropy.

Missing Data Handling

The analysis was performed on subjects with at least 50% compliance (42 observations). In the raw format, 22.5% of data was missing. The moving-average procedure reduced missingness to 3%. Missing data was not imputed. Whether missing data is missing at random (MAR) or MNAR cannot be tested, however, we could not detect a systematic pattern of missingness (Table S6, only age was weakly correlated with missingness).

As conservative compliance thresholds may result in selection, we accounted for that with a sensitivity analysis shifting the criterion to a minimum of 10% compliance (8 observations). This reduces potential selection bias by including more participants but allows more missing data per subject at the same time. After pre-processing, this data set contained 20% missing data. Visual inspection revealed two outliers that were removed before analysis. We further compared the full RCT sample with both subsamples (50% and 10% compliance) to examine potential selection bias based on baseline characteristics.

Latent Class Contrasts

Chi-square tests and analysis of variance examined differences in baseline distributions across latent classes (*stats* and *rstatix* packages). Clinical effects according to changes in THI, TFI and PHQ-9 from baseline to 12-week final visit were calculated for each class and compared to MCID estimates (THI = 11, TFI = 9, PHQ-9 = 6) (1,202). The distribution of classes within treatment arms is plotted only, as cell sizes do not allow for statistical testing. *P* values were *Holm-Bonferroni* adjusted for multiple comparisons, with the critical significance level set to $p \leq .05$. All analyses were performed in R (Version 4.4.1).

Trajectories of Remaining Symptoms

The trajectories of the remaining symptoms predicting clinical improvement were plotted as mean trajectories of pre-processed individual time-series (seven-day moving average) and 95% confidence intervals (CI). Based on that, we exploratively assessed the development of the association between daily tinnitus distress and maximum tinnitus loudness for each class with two additional analyses. The underlying assumption is that a reduction of the association can be considered as a desired treatment effect as tinnitus distress evolves to be evaluated more independent of tinnitus loudness (Study II). We computed the weekly correlations between both symptoms for each class and the standardized difference between both symptoms ($[d - \text{mean}\{d\}]/\text{sd}\{d\}$; $d = \text{maximum tinnitus loudness} - \text{daily tinnitus distress}$) for each class. This quotient was standardized intra-individually. A value of 0 in the standardized quotient indicates the individual average difference between daily tinnitus distress and maximum tinnitus loudness. An increase in this quotient over time indicates that the difference between both becomes greater (desired treatment effect), while a decrease over time indicates that the difference becomes smaller (undesired treatment effect).

Results

Sample Description

The sample consisted of 147 chronic tinnitus patients with 9634 observations (see Table 8). 45% were female, patients were on average 55 years old and 44% had graduated from university. The participants provided 65 observations on average (range: 42 – 84).

Table 8. Demographic and Clinical Characteristics of Participants at Baseline (N = 147)

Baseline Characteristics	Mean (SD) N (%)
Gender	
female	66 (44.9%)
male	81 (55.1%)
Age	54.72 (11.86)
Education	
elementary school	16 (10.9%)
high school	23 (15.6%)
middle school	44 (29.9%)
university or higher	64 (43.5%)
THI score	49.17 (20.29)
PHQ9 score	7.58 (4.93)
Hearing loss (PTA4)	20.61 (16.6)
Tinnitus duration (in months)	136.18 (120.98)
Number of comorbidities	2.28 (2.14)
Number of tinnitus treatments	0.32 (0.64)
Extraversion (BFI-2)	39.29 (6.84)
Agreeableness (BFI-2)	47.18 (6.17)
Conscientiousness (BFI-2)	46.94 (8.08)
Neuroticism (BFI-2)	34.19 (8.27)
Openness (BFI-2)	43.72 (7.48)

Notes. THI (Tinnitus Handicap Inventory) score range: 0-100. PHQ-9 (Patient Health Questionnaire-9) score range: 0-27. PTA4: Average hearing threshold at the frequencies 500 Hz, 1 kHz, 2 kHz and 4 kHz. Top five comorbidities: sleep difficulties (N = 37), high blood pressure (N = 26), high cholesterol (N = 26), acid reflux (N = 24), thyroid disorder (N = 21). Number of tinnitus treatments received prior to study participation. BFI-2 dimensions score range: 12-60.

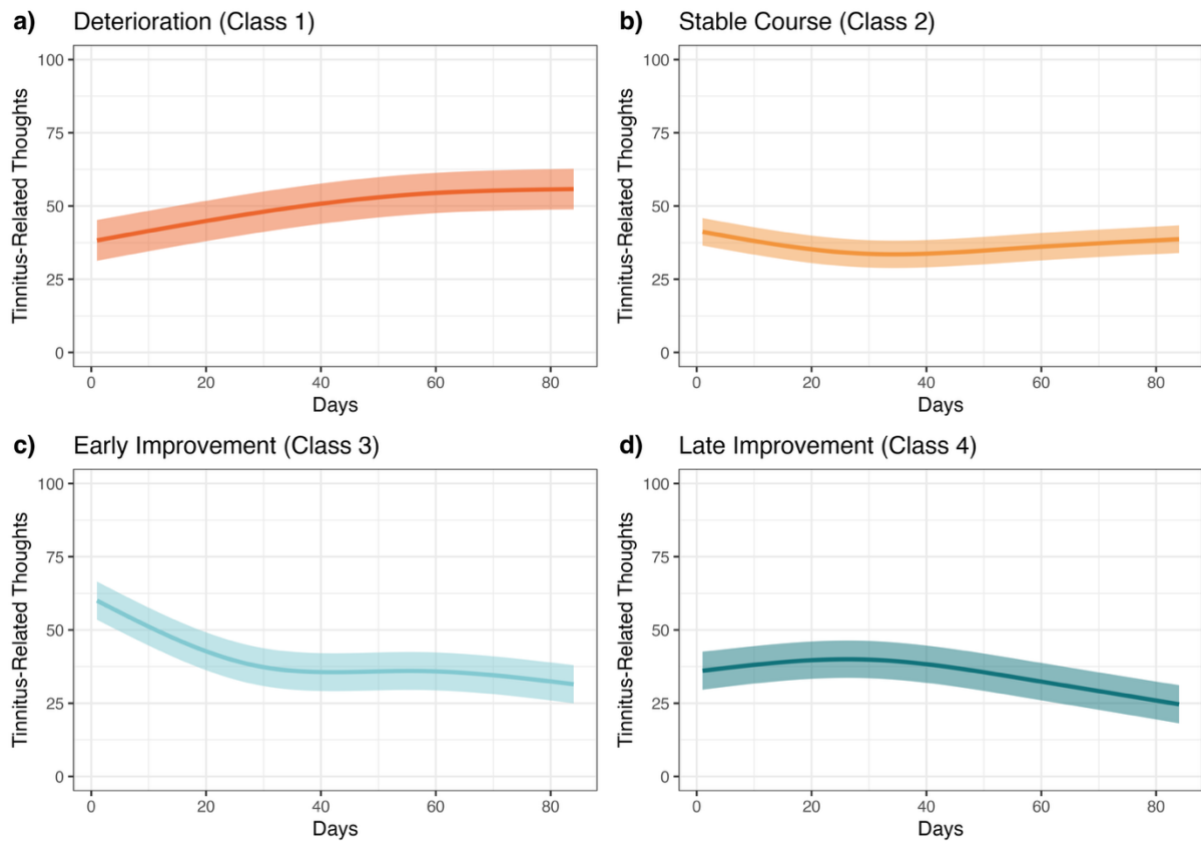
Model Selection

All spline models had a better relative fit than their linear equivalents (Figure S10). The relative fit of the spline models corresponded to an asymptotic reduction of AIC and BIC with increasing number of classes without a reasonable minimum indicating a clearly favourable class solution (Figure S10b). Thus, more weight was put on the other metrics. As comparable literature identified three-, four- and five-class solutions (194–196), these solutions were more closely inspected (Figure S11). PPM and entropy values were high for all three models with a slight tendency towards the four-class solution (Table S7). Inspecting the trajectories, class 4 in the four-class model adds an interesting and potentially relevant group describing a later-onset response. In the five-class model, the new trajectory is captured by class 2, however, the relevance of its distinction from class 3 is ambiguous as neither one describes an improvement (Figure

S11c). Further, class 2 is small (10.9%, N = 16) and CIs overlap strongly in the five-class solution. Thus, to remain conservative and avoid overfitting, we decided that the four-class latent trajectories describe the data best.

Description of Latent Trajectories

The final result consisted of four classes with distinct trajectories (Figure 9). Class 1 depicted increasing tinnitus-related thoughts and was labelled *Deterioration*. (18.4%, N = 27). The estimated trajectory started at 38 (95% CI: 31-45) and showed an increase to 56 (95% CI: 49-63). Class 2 patients showed no change in the outcome and was labelled *Stable Course* (40.1%, N = 59). The trajectory decreased slightly from 41 (95% CI: 37-46) on the first day until 34 (95% CI: 29-38) on day 34 before it increased back to 39 (95% CI: 34-43). Class 3 showed an early response in tinnitus-related thoughts and was labelled *Early Improvement* (20.4%, N = 30). The trajectory depicted the highest baseline of 60 (95% CI: 54-67) and a strong decrease during the first half of the treatment period until 36 (95% CI: 29-42) on day 42 followed by a short counter climb before it decreased again to 32 (95% CI: 25-38). Finally, group 4 (21.1%, N = 31) showed a late-onset response and was labelled *Late Improvement*. The estimation started at 36 (95% CI: 30-43), showed a small increase until 40 (95% CI: 34-46) on day 26 before it decreased to 25 (95% CI: 18-31). The individual observed trajectories are plotted in Figure S12.

Figure 9. Latent Trajectories of the Final Model

Notes. The four-class model was chosen due to its highest entropy and PPM values as well as because class 2 in the five-class model was considered irrelevant from a size and clinical point of view (Figure S11). Trajectories depict estimated values, transparent bars are 95% CIs.

Sensitivity Analysis

The latent trajectories of the four-class result based on the 10% compliance data set are depicted in Figure S13b. Deterioration featured 22%, Stable Course 36%, Early Improvement 15%, and Late Improvement 27% of participants. PPM was high (0.94 – 0.99) and entropy was acceptable (0.88). Trajectories differed minimally from the four-class result based on the 50% compliance sample (see Figure S13a). Both subsets (10% and 50% compliance) were compared with the full RCT sample based on baseline characteristics, only age showed significant differences between samples (see Table S8).

Latent Class Contrasts

Demographic or clinical variables at baseline were not predictive of latent trajectory membership (see Table 9).

Study III

Table 9. Latent Class Contrasts on Baseline Characteristics

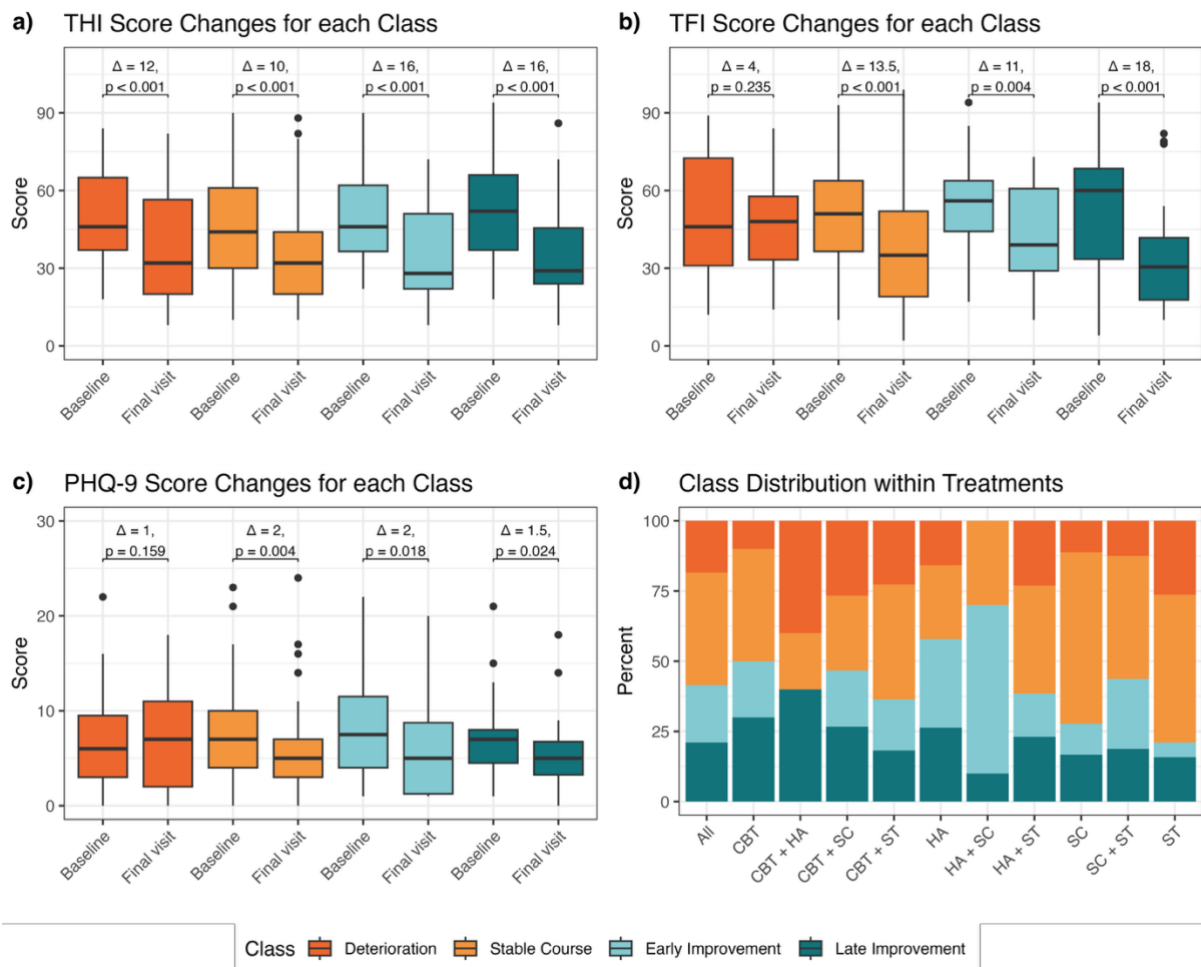
Baseline Characteristics	Class 1: Deterioration	Class 2: Stable Course	Class 3: Early Improvement	Class 4: Late Improvement	<i>p</i> value	<i>adj. p</i> <i>value</i>
Gender						
female	12 (44.4%)	26 (44.1%)	16 (53.3%)	12 (38.7%)	.740	1.00
male	15 (55.6%)	33 (55.9%)	14 (46.7%)	19 (61.3%)	.740	1.00
Education						
Elementary school	4 (14.8%)	4 (6.8%)	3 (10%)	5 (16.1%)	.520	1.00
high school	3 (11.1%)	9 (15.3%)	3 (10%)	8 (25.8%)	.520	1.00
middle school	7 (25.9%)	18 (30.5%)	9 (30%)	10 (32.3%)	.520	1.00
university or higher	13 (48.1%)	28 (47.5%)	15 (50%)	8 (25.8%)	.520	1.00
Age	58.78 (9.82)	54.36 (12.31)	55.63 (10.5)	51 (13.07)	.090	1.00
THI score	49.56 (21.25)	46.71 (20.51)	51.4 (19.19)	51.35 (20.54)	.660	1.00
PHQ9 score	7.26 (5.35)	7.69 (4.99)	8.13 (5.24)	7.13 (4.28)	.860	1.00
Hearing loss (PTA4)	19.07 (14.57)	20.53 (15.5)	21.72 (16.07)	21.02 (20.92)	.940	1.00
Tinnitus duration (in months)	141.4 (123.25)	121.61 (123)	162.86 (136)	134.07 (99.39)	.530	1.00
Extraversion	40.08 (7.94)	37.95 (6.55)	40.4 (7.08)	40.06 (6.01)	.290	1.00
Agreeableness	48.04 (5.82)	46.52 (6.85)	48.13 (6.02)	46.77 (5.28)	.570	1.00
Conscientiousness	45.38 (7.54)	47.21 (8.3)	48.47 (8.31)	46.29 (7.98)	.520	1.00
Neuroticism	34.58 (8.24)	35.26 (7.94)	33.33 (9.45)	32.71 (7.76)	.510	1.00
Openness	42.65 (7.95)	42.91 (8.58)	45.07 (6.34)	44.84 (5.62)	.420	1.00
Number of comorbidities	2.89 (2.82)	2.15 (2.12)	1.97 (1.71)	2.29 (1.88)	.390	1.00
Number of tinnitus treatments	0.41 (0.8)	0.37 (0.67)	0.13 (0.51)	0.32 (0.54)	.330	1.00

Notes. Values depict Mean (SD) or N (%). *P* values were adjusted according to Holm-Bonferroni. THI score range: 0-100. PHQ-9 score range: 0-27. PTA4: Average hearing threshold at the frequencies 500 Hz, 1 kHz, 2 kHz and 4 kHz. BFI-2 dimensions score range: 12-60.

Study III

Figure 10a-c depicts relevant PROM changes across classes. THI score improved significantly for each class from baseline to final visit with clinically relevant improvements for Deterioration, Early and Late Improvement. TFI score improved significantly and clinically relevant for Stable Course, Early and Late Improvement. The PHQ-9 improved significantly for Stable Course, Early and Late Improvement, none reached clinical significance. The distribution of class size for each treatment arm is depicted in Figure 10d.

Figure 10. PROM Changes and Treatments across Classes



Notes. Figures a-c depict PROM changes across classes. THI (Tinnitus Handicap Inventory) score range: 0-100. TFI (Tinnitus Functional Index) score range: 0-100. PHQ-9 (Patient Health Questionnaire-9) score range: 0-27. Figure d depicts the distribution of class size for each treatment arm. Treatments were provided single or in combination: Cognitive Behavioural Therapy, Hearing Aids, Structured Counseling, and Sound Therapy.

Trajectories of Remaining Symptoms

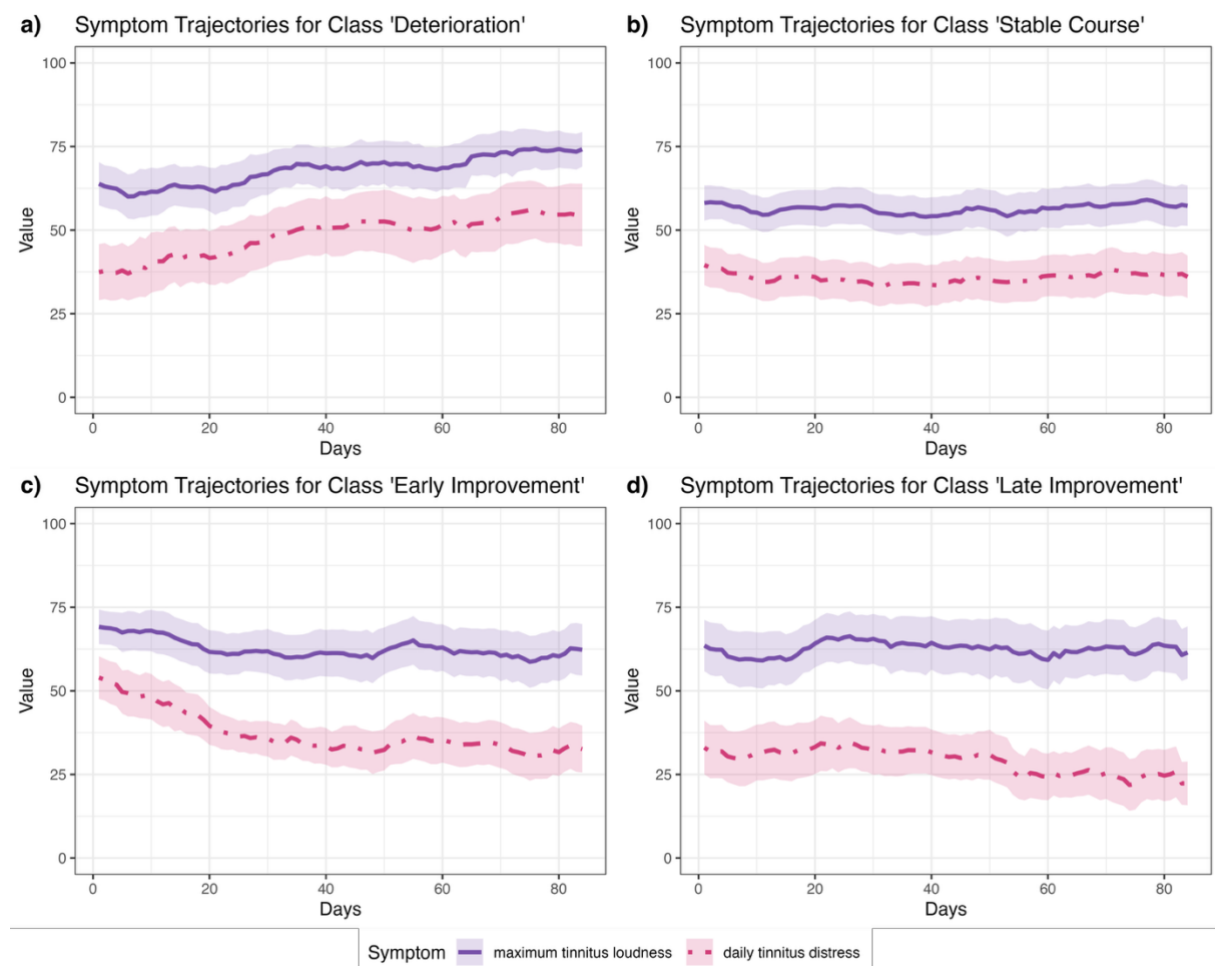
The trajectories of the remaining symptoms relevant for clinical improvement (jaw tension, momentary tinnitus loudness, daily tinnitus distress, maximum tinnitus loudness, and negative emotion) are depicted for each class in Figure S14. Daily

Study III

tinnitus distress varied very closely with tinnitus-related thoughts in each class. Maximum and daily tinnitus loudness was rated the highest across time and classes, jaw tension is rated the weakest. Trends in the tinnitus-related thoughts, if available, are reflected less pronounced in some symptoms and classes.

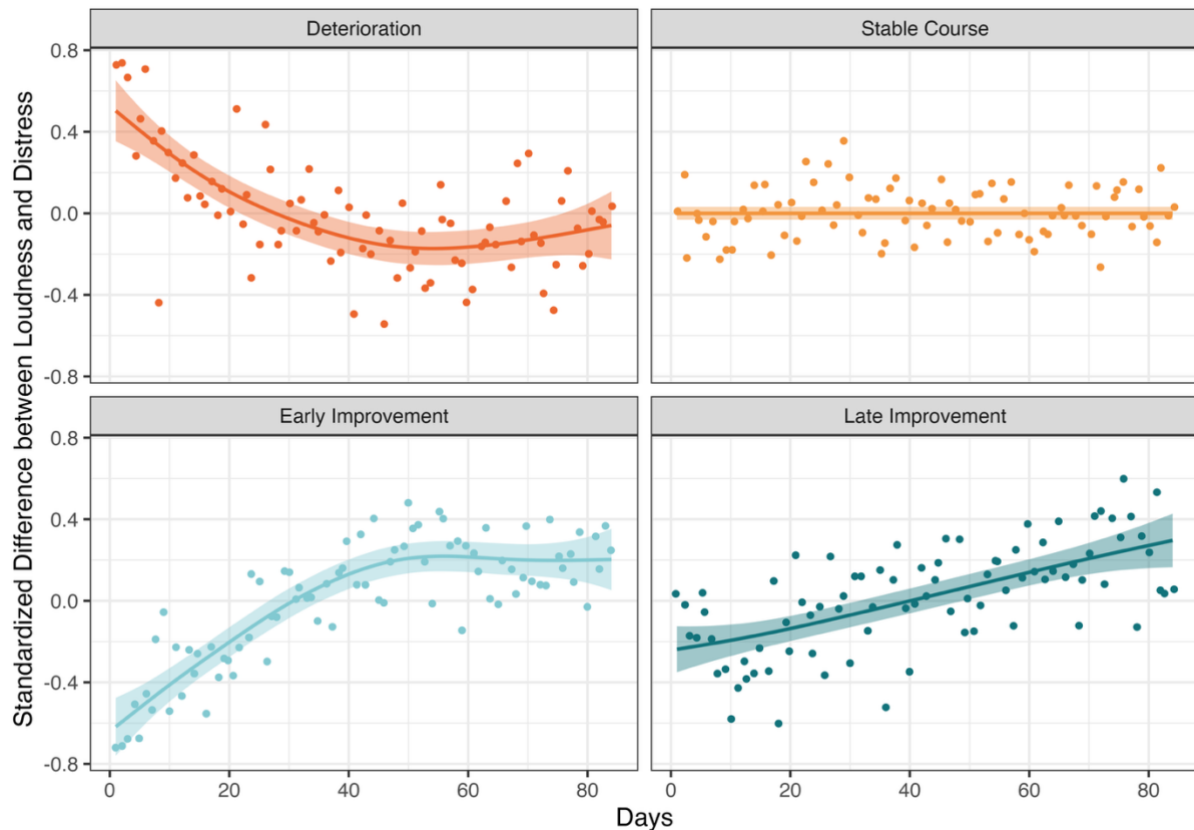
Figure 11 shows the mean trajectories of daily tinnitus distress and maximum tinnitus loudness. The weekly average correlation between both remained descriptively stable for the classes Deterioration and Stable Course and decreased for the classes Early and Late Improvement (Figure S15). When looking at the standardized difference between daily tinnitus distress and maximum tinnitus loudness, the difference decreased in the Deterioration Class, remained stable in the Stable Course Class, and increased in both the Early and Late Improvement Classes (Figure 12).

Figure 11. Trajectories of Tinnitus Loudness and Distress across Classes



Notes. Trajectories are mean values of pre-processed individual time-series (seven-day moving average), transparent bars are 95% CIs. All symptoms were rated on a visual analogue scale from 0 — 100.

Figure 12. Distribution of the Intra-Individual Standardized Difference (maximum tinnitus loudness – daily tinnitus distress) by Class



Note. Distribution of the intra-individual standardized difference between maximum tinnitus loudness and daily tinnitus distress ($[d - \text{mean}\{d\}]/\text{sd}\{d\}$; $d = \text{maximum tinnitus loudness} - \text{daily tinnitus distress}$). Dots represent mean values by class and day. Trend lines depict the smoothed trend and 95% CI (geom_smooth in ggplot). Intra-individual standardization means that positive values indicate a greater difference between maximum tinnitus loudness and daily tinnitus distress compared to the individual's average difference, while negative values indicate a smaller difference compared to their individual average. An increase in this quotient over time indicates that the difference between both becomes greater (desired treatment effect), while a decrease over time indicates that the difference becomes smaller (undesired treatment effect).

Discussion

The present study demonstrated that EMA-measured trajectories in tinnitus-related thoughts while patients received established interventions can be well described by four distinct trajectories. Roughly one-fifth of participants showed a progressive increase (Deterioration, 18%), two-fifths experienced no discernible change (Stable Course, 40%), while the remaining patients split almost evenly between an early response (Early Improvement, 20%) and a late-onset response pattern in tinnitus-related thoughts (Late Improvement, 21%). Despite different metrics and treatment formats used, this is in line with Nieman and colleagues (190). Both studies converge

Study III

on the existence of multiple distinct symptom-change subgroups, prospectively demonstrating heterogeneity in treatment response.

Similar research in mental health has mostly identified three- to four-class solutions, suggesting that the responder/non-responder dichotomy masks clinically relevant heterogeneity (194–196). Our finding of early and late responders aligns with prior evidence that improvement can follow different pacing, sometimes with temporary deterioration before benefits emerge (195,196). Such initial worsening may reflect psychological activation processes (203) and, in tinnitus, increased symptom vigilance during treatment. Communicating this to patients may help manage expectations and reduce dropout, as seen in the “Late Improvement” group (Table S9).

Applying a 50% EMA completion threshold excluded many participants, but replication with a 10% threshold produced nearly identical class trajectories and proportions. Both subsets closely matched the full RCT sample (except for minor age differences), suggesting minimal selection bias. While results appear robust, replication in further samples and treatment settings is needed.

Baseline demographics and clinical data showed no significant class differences, suggesting that generic prediction of treatment response is difficult and highlighting heterogeneity in treatment needs. Random allocation likely resulted in some patients receiving optimal and others sub-optimal interventions, consistent with the presence of all trajectories across treatment arms. Nonetheless, predicting response to specific treatments based on baseline characteristics remains possible. Precision matching approaches, such as the Personalized Advantage Index, may help optimize treatment allocation using prescriptive predictors (204).

GMM revealed subgroups with stable or worsening tinnitus-related thoughts, yet PROMs suggested a more favourable picture. Changes in the TFI mirrored the EMA based latent classes most closely, even though the “Stable Course” class also registered an average improvement in TFI score. In the “Late Improvement” Class, patient reported low TFI scores on average during final visit which could have been influenced by the late-onset decline in EMA measured tinnitus-related thoughts. Further research is needed to clarify the differences between PROMs and EMA data. Average trajectories showed that tinnitus distress closely paralleled tinnitus-related thoughts, suggesting one construct may suffice to reduce burden without loss of information. In contrast, loudness remained high across classes, supporting its partial independence from distress (205,206). Correlation analyses and trajectories of

Study III

standardized difference indicated stable loudness–distress associations in “Deterioration” and “Stable Course,” but divergence in “Early” and “Late Improvement,” potentially reflecting treatment success (Study II). This pattern also suggests current treatments mainly reduce burden rather than loudness, with early improvers plateauing around day 50, while late improvers continued to diverge, possibly with further gains under longer treatment.

The following limitations should be considered. First, we hypothesize that improvements of 28 and 11 points in both response groups are clinically relevant, yet, the available estimation of clinically relevant differences is based on tinnitus distress and loudness (134). Second, daily-measured tinnitus-related thoughts is no established outcome measure even if it revealed some predictive power in an earlier investigation (Study II). Third, our analysis relied on a subsample, posing potential selection bias despite sensitivity analyses and comparisons with the full sample. Smartphone-based data collection and limited sample size add to this risk. Although moving averages reduced missingness, future studies should prioritize maximizing compliance to minimize data loss. Fourth, GMMs require a lot of choices that could influence the number and shape of trajectories. By adhering to the reporting guidelines, we followed the currently existing standard procedure. Fifth, the relative fit indices AIC and BIC failed to pick a clear number of classes, thus, we relied on the other priorly defined criteria to identify the optimal solution. Sixth, due to the large number of treatment arms, our sample lacked power to search for class-by-treatment interactions.

Despite limitations, our findings reveal substantial heterogeneity in symptom trajectories, opening avenues for research and practice. Intensive monitoring can yield valuable insights, ideally complemented by passive data to reduce burden. Precision medicine, the data-driven attempt to individualise treatment, strongly benefits from continuous monitoring of symptom trajectories to regularly check if patients are “on track” and adapt treatment plans to individual needs based on real-time data (207). Together with parallel efforts in prescriptive research initial allocation of patients to treatments could be improved, while continued monitoring beyond treatment may clarify whether improvements consolidate, fade, or require maintenance strategies.

In summary, high-intensive longitudinal monitoring reveals rich, non-linear patterns of change that remain invisible to traditional assessment schedules. Embracing this complexity will be essential to tailor interventions to inter- and

intraindividual needs to ultimately alleviate the burden of tinnitus for the diverse population of sufferers.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme, Grant Agreement Number 848261. The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

We would like to sincerely thank all patients who took part in the UNITI trial.

Conflict of Interest

W.S. and R.P. are stakeholders of the Lenox UG but declare no non-financial competing interests. The Lenox UG aims to translate scientific knowledge into e-health applications and also holds shares of the HealthStudyClub GmbH. W.S. received consulting fees as well as payments for lectures in connection with chronic tinnitus topics but declared no non-financial competing interests. R.P. received consulting fees, reimbursements for congress attendance, and travel expenses as well as payments for lectures in connection with mobile health and e-mental health topics but declares no non-financial competing interests. R.P. serves as Associate Editor of this journal and had no role in the peer review or decision to publish this manuscript. All other authors declare no financial or non-financial competing interests

Discussion

This thesis aimed to explore the clinical utility of intensive longitudinal assessments conducted during tinnitus treatment, focusing on the understanding and characterisation of symptom dynamics and changes, as well as on the development of analytical approaches for longitudinal data. The first study evaluated the feasibility of EMA to capture symptomatic change with high ecological validity using different individual- and group-based analytical approaches. The second study systematically examined whether and how EMA data can depict treatment response by identifying which EMA variables and patterns of change best characterise clinical improvement. The third study investigated whether individual symptom trajectories could be clustered into latent groups, thereby identifying subgroups of patients with similar treatment responses. In the final chapter, the results are discussed across shared conceptual and methodological themes.

Decoupling between Tinnitus Distress and Loudness: Indicator of Treatment Response?

One of the central questions of this thesis was whether and how EMA can evaluate therapy outcome. In the first study, group-level correlations between weekly tinnitus distress and loudness yielded an intriguing finding: Descriptively, their association weakened over time. This decoupling was hypothesised to reflect treatment response to psychological interventions which is supported by neurophysiological evidence of two separate networks underlying tinnitus distress and loudness, which are interconnected only in distressed tinnitus patients (208).

The second study further advanced this question by extracting time-series features from EMA data and selecting those most predictive of clinical improvement, as assessed by the CGI-I. Responders were characterised by fluctuation-corrected linear decreases in tinnitus-related thoughts, jaw tension, maximum and momentary tinnitus loudness, linear increases in happiness, greater stability in daily tinnitus distress, as well as change in variability in momentary tinnitus loudness. Feature importance analysis identified *linearity* as the most influential feature, and *tinnitus-related thoughts* as the most sensitive EMA variable. It should be noted that the preference for linearity may have been influenced by the linear nature of the model.

Yet, these findings emphasised two points: First, clinical improvement is primarily perceived as a reduction in tinnitus symptoms, accompanied by decreased symptom variability. This is consistent with the hypothesis that high fluctuations may contribute to patient distress (72). Second, even if tinnitus-related thoughts appear most sensitive in depicting treatment response, change processes in a variety of symptoms contribute to the clinical improvement of patients. The importance of quality-of-life outcome measures, in addition to mere symptom reduction, in contributing to suffering has been repeatedly emphasised in psychiatric research (209,210).

In the third study, GMM identified four latent trajectory classes based on tinnitus-related thoughts: *Early Improvement* (20.4%), *Late Improvement* (21.1%), *Deterioration* (18.4%), and *Stable Course* (40.1%). Consistent with Study I, both improvement groups displayed a decoupling of daily tinnitus distress and maximum tinnitus loudness over time. While causal factors remain uncertain, tinnitus-related thoughts may mediate the relationship between tinnitus distress and loudness. However, it is possible that this decoupling effect is mainly driven by reductions in distress without substantial changes in loudness (see Figure 11). Future research should disentangle these relationships and examine whether decoupling of distress and loudness may serve as an early indicator of treatment response, informing decisions on treatment continuation or adaptation.

Notably, the existence of the *Late Improvement* group carries clinical significance: patients who show no early improvement may still respond at later stages. This underscores the importance of persistence with treatment, while also complicating early prediction of treatment response.

Tinnitus Symptom Network

Even though tinnitus-related thoughts and their association with distress and loudness emerged as key indicators, the role of other symptoms should not be underestimated. Building on the observations in the third study that certain symptoms fluctuate together and motivated by the intention of reducing participant burden through shorter EMA questionnaire (211), a network analysis was conducted to visualise their interrelations. It was estimated using partial correlations via the *EBICglasso* algorithm from the R package *qgraph* computing a Gaussian graphical model regularised via Lasso to shrink weak associations to zero, thereby producing a sparse network. 44296 EMA data from 1216 patients using the UNITI app were person-mean centred to remove between-

Discussion

subject differences before network estimation. CIs were obtained using non-parametric cluster bootstrapping, resampling participants as clusters to account for the nested EMA data structure. It should be noted that changes over time are not depicted here.

Visually, three sub-networks emerged (see Figure 13); a tinnitus-related network, a somatic network, and a general health network. Within the tinnitus-related network, strong associations were found between momentary distress and loudness (0.53, 95% CI: 0.50-0.56), and between tinnitus-related thoughts and daily distress (0.52, 95% CI: 0.51-0.54). Weaker correlations appeared between daily and momentary distress (0.25, 95% CI: 0.23-0.27), as well as between momentary and maximum loudness (0.4, 95% CI: 0.37-0.42), suggesting that momentary and daily items capture distinct aspects of the tinnitus experience. Intriguingly, distress and loudness were more strongly correlated at the momentary level (0.53, 95% CI: 0.50-0.56) than at the daily level (0.19, 95% CI: 0.17-0.21). Daily distress appeared to be more strongly linked to tinnitus-related thoughts (0.52, 95% CI: 0.51-0.54), than to maximum loudness (0.19, 95% CI: 0.17-0.21). Although weak, daily distress is further correlated with stress (0.08, 95% CI: 0.07-0.09) and positive emotions (-0.06, 95% CI: -0.08 – -0.05). Thus, distress in the moment may be more directly linked to the perceptual intensity, while daily distress may be more influenced by cognitive factors such as tinnitus-related thoughts.

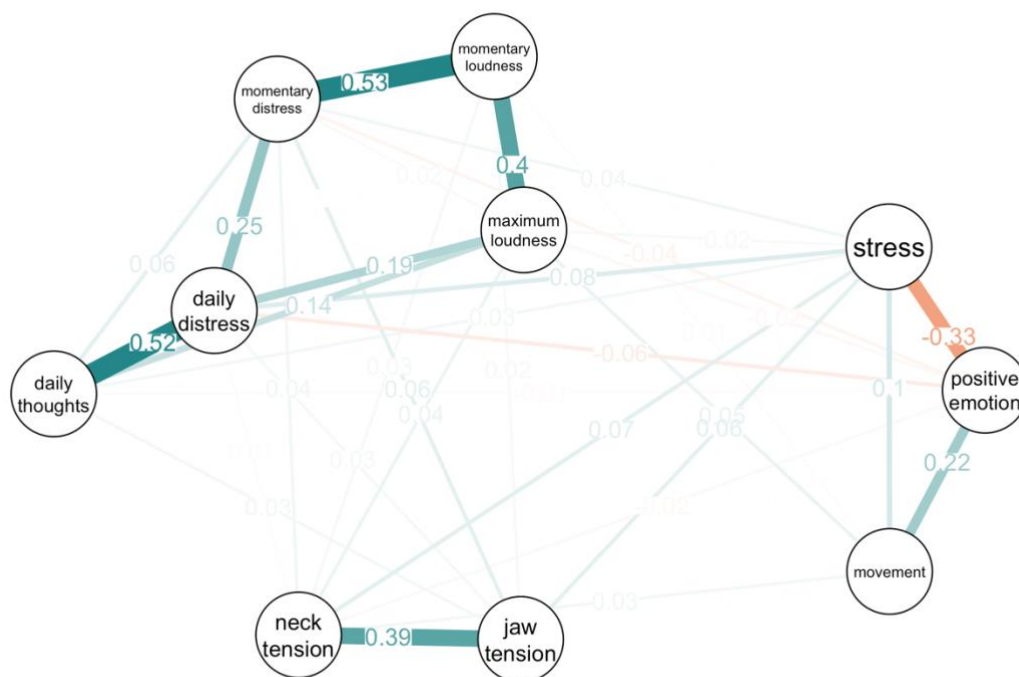
Within the somatic sub-network, jaw and neck tension correlated positively (0.39, 95% CI: 0.36-0.43). Stress, positive emotions, and movement formed a third sub-network. Positive emotions were negatively associated with stress (-0.33, 95% CI: -0.34 – -0.31) and positively with movement (0.22, 95% CI: 0.21-0.24).

Between-sub-network associations were weak (≤ 0.10), suggesting that these sub-networks capture largely distinct but clinically relevant aspects of tinnitus, consistent with findings from Study II. Christensen (2020) applied a heuristic threshold of 0.20 in partial correlations as an indicator of questionnaire item redundancy, albeit in cross-sectional psychometrics (212). Applying this rationale, the present findings suggest that selecting at least one variable from each sub-network could substantially reduce the EMA questionnaire while retaining essential information. Moreover, the weaker day-level coupling between distress and loudness, compared to the momentary level, supports the inclusion of both temporal perspectives in future EMA protocols.

Discussion

Furthermore, the network analysis provided valuable insights into the interrelations among tinnitus-related variables, warranting further investigation. Tinnitus-related thoughts, which are assumed to play a central role in treatment effects, appeared only weakly associated with tinnitus loudness. Although the weak association between loudness and distress has been well documented in the tinnitus literature (84), tinnitus-related thoughts, despite its correlation with daily distress, might represent a distinct construct. An important question arising from these results is what factors influence the occurrence and variability of tinnitus-related thoughts in daily life. EMA data offer a promising avenue to address this question by enabling the analysis of time-contingent associations between cognitive, emotional, and perceptual processes.

Figure 13. Network Analysis of EMA Items



Notes. Gaussian graphical model regularised via Lasso. Nodes represent variables and edges represent partial correlations between variables.

EMA vs. PROMs: Interchangeable or Complementary?

Closely linked to the question of whether EMA can capture treatment outcomes is its relationship with established PROMs. In Study I, pooled baseline and end-of-intervention EMA data did not show group-level treatment effects contrasting with statistically and clinically significant THI improvements. This discrepancy likely arises

from differences in assessment timing: the THI was completed before the baseline EMA phase, with the latter possibly reflecting anticipation effects. Despite this difference, correlations between individual THI improvements and EMA-derived change scores revealed strong associations.

In the second study, correlation coefficients between CGI-I scores and EMA time-series features ranged from small to moderate ($r = 0.01\text{--}0.36$), with a combined model providing a good fit (McFadden $R^2 = 0.82$). In the third study, the TFI corresponded best with EMA-based trajectory groups, although PROMs generally indicated more favourable outcomes than EMA, similar to findings in Study I.

The observed discrepancies are not surprising, given differences in assessment schedules, contexts, and psychometric properties. It is thought that EMA and retrospective measures capture different aspects of the same construct (63). On a small scale, the different associations of momentary and daily distress in the network analysis might demonstrate this. Thus, the method should be carefully selected depending on the theoretical framework of the research question and its respective limitations in mind.

Does compliance limit feasibility?

Compliance remains a key methodological concern, as EMA participation can be demanding for patients and may reduce adherence. In this thesis, response rates varied considerably. While participants in the first study responded on 79% of days on average, only about 30% of participants reached the 50% compliance threshold in the RCT dataset used in the second and third study. This difference may be attributable to study design; The first study involved a small, closely supported pilot sample, whereas the RCT included nearly 500 participants across multiple centres. A meta-analysis identified compliance rates as low as 10% in EMA studies, although 57% of studies reported a compliance of 80% or higher which is likely influenced by publication bias (179).

Sensitivity analyses in the second and third study indicated that results remained stable even when participants with lower compliance were included, with predictive model fits (Study II) and latent class structures (Study III) remaining robust across compliance levels of 10–50%.

Nevertheless, adequate adherence remains essential for generalisability, as high-compliance subsamples may result from selection bias. Missing data are probably

rarely random, as participants in EMA settings may not respond due to deterioration in their symptoms or general health. Future EMA research should implement strategies known to enhance compliance, such as reducing questionnaire length, and integrating financial or psychoeducational incentives (81,179,211).

Addressing within- and between-tinnitus heterogeneity through methodological approaches

A central aim of this thesis was to evaluate methods for analysing EMA data during treatment. Because EMA is still relatively new in health research and its analytical toolbox is evolving, implementation and analysis remain more demanding than with conventional PROMs (213). As shown in Study I, tinnitus symptoms exhibit pronounced within-person fluctuations and substantial between-person heterogeneity. This pattern is often obscured by retrospective assessments but potentially clinically relevant, given that high tinnitus variability has been hypothesised to contribute to greater tinnitus severity (72). Raw time-series plots therefore reveal important features such as non-linearity of symptomatic change or the degree of variability, but are ill-suited, in their unprocessed form, to quantify treatment success.

The second study developed a methodological pipeline to predict treatment response from EMA data. To address the high within-person variability, individual trajectories were summarised into *time-series features* that capture dynamic properties. Thus, feature-based time-series analysis, originating in signal processing and later adopted in medical research (169,170), was applied. Features were pre-selected based on their potential to capture change which was expected given the assigned treatments. Next, an ordinal elastic-net regression was implemented to select and weight time-series features. This approach demonstrates how complex, high-frequency EMA data can be reduced to interpretable features and linked to clinically meaningful outcomes. Future studies could build on this by using non-linear and more advanced machine learning algorithms on larger data sets.

In the third study, the pronounced between-subject heterogeneity observed in the first study was addressed using a modelling approach that assumes the existence of latent classes rather than applying a one-size-fits-all algorithm to the entire sample. Moving-averaged trajectories of tinnitus-related thoughts were used to smooth daily fluctuations prior to class estimation. This approach identified distinct response and

non-response patterns, demonstrating how heterogeneity in idiographic dynamics can be meaningfully aggregated into clinically interpretable subgroups.

Finding the Sweet Spot between Idiographic and Nomothetic Approaches

Altogether, this thesis positions the integration of idiographic and nomothetic perspectives as a central methodological challenge and opportunity for future EMA research. The nomothetic perspective seeks to infer generalisable rules from a population, typically operationalised through group-level analyses. In contrast, the idiographic perspective regards each individual as unique, employing individualised methods to study within-subject variability (3). EMA enables to occupy a conceptual middle ground between these traditions: it provides intensive within-person data while also allowing for aggregation across individuals or subgroups. The challenge lies in identifying the methodological “sweet spot” that preserves individual-specific information while enabling generalisable clinical insights. Where this balance lies is likely condition-specific and depends on the degree of heterogeneity within the disorder.

The empirical studies of this thesis illustrate different ways of navigating this tension. The first study explored several approaches ranging from visual inspection of raw individual trajectories to aggregated group-level analyses thereby identifying weaknesses and advantages of both. The applied study design, a SCED, combines the idiographic principle of within-person experimental manipulation with the possibility of generalisation through between-subject replication. The second study applied feature-based time-series analysis, summarising individual dynamics into mathematically defined indices that could be analysed at the group level. This method retains individual information while yielding interpretable, generalisable predictors of treatment response. The third study clustered individual trajectories into latent subgroups, thereby identifying groups of patients with similar response patterns. For the tinnitus network analysis, raw data were person-mean centred to remove between-subject differences before network estimation. Together, these approaches demonstrate how idiographic information can be aggregated in a way that still respects the individual nature of tinnitus experience.

From a broader perspective, EMA-based tinnitus research should strive for such methodological integration by linking individual symptom trajectories, subgroup

identification, and group-level inferences. This multilevel approach aligns with contemporary trends towards personalised and precision medicine while maintaining the empirical rigour necessary for clinical application (214). Ultimately, identifying this “sweet spot” offers a pathway to models of tinnitus that are both personally meaningful and clinically generalisable.

Limitations

This thesis should be interpreted in light of several limitations. It represents one of the first attempts in tinnitus research to capture treatment effects longitudinally using EMA, and the first to methodologically examine EMA as a tool for evaluating treatment outcomes. Consequently, all findings are preliminary and require external validation.

A further limitation concerns the psychometric properties of the UNITI app items, which have not been systematically validated. Psychometric validation ensures that items accurately capture the intended psychological or behavioural constructs, thereby reducing measurement error, improving data quality, and minimising bias (215). That said, the lack of standardised validation procedures is a broader challenge in EMA research. EMA measures differ from conventional questionnaires in several respects, including their repeated administration in real-world contexts, the use of single-item assessments, and their focus on temporal stability and sensitivity to change. These are characteristics that extend beyond traditional psychometric frameworks (215). In tinnitus research, validated single-item VAS may offer a pragmatic alternative for future studies (216).

In the context of treatment evaluation, EMA should also be scrutinised for reactivity, meaning that frequent self-monitoring might alter the behaviour or experience being measured. Although tinnitus distress generally appears stable, some participants reported increased awareness of their tinnitus during frequent monitoring, suggesting potential reactivity effects (65,66).

Finally, the general methodological and analytical limitations of EMA discussed in the introduction also apply to this work.

Future Directions

Recent advances in psychotherapy research apply EMA within an idiographic framework. One prominent approach is the estimation of idiographic networks from high-frequency multivariate longitudinal data, typically using vector autoregressive modelling (217). These person-specific networks capture dynamic interactions

between symptoms and are thought to reveal the individual pathology underlying a disorder. The central idea is that identifying highly connected symptoms provide primary treatment targets, as effectively addressing these symptoms may initiate a chain reaction that alleviates other symptoms in the network. Network-informed personalised treatments are currently being piloted, particularly in the context of eating disorders (218–220), despite already being criticised for their instability (217). At first glance, this concept appears promising for tinnitus research, given the established effectiveness of CBT for a subset of patients (37). However, access to therapy remains rare, highlighting the need for more scalable eHealth solutions (102). For example, sets of symptom interactions that recur across individuals could inform the design of modular CBT interventions. These modules could then be delivered in a semi-personalised manner, guided by EMA-informed idiographic networks.

While idiographic networks are typically estimated from longitudinal data collected prior to treatment, EMA can also capture symptom dynamics during treatment, thereby enabling flexible adaptation of treatment plans. This principle is operationalised in Ecological Momentary Interventions or Just-in-Time Adaptive Interventions (JITAI), where ambulatory data informs the timely delivery of short interventions. For example, Garland and colleagues used heart rate variability (HRV) to trigger mindfulness exercises in chronic pain patients when HRV exceeded one standard deviation. Although symptoms improved after prompts, adherence was low, indicating that the nomothetic trigger rule was suboptimal. The authors proposed more individualised, closed-loop systems that adapt interventions based on biomarker feedback. Given the pronounced within-subject variability in tinnitus, JITAI represent a promising approach to alleviating periods of heightened tinnitus distress. Realising this potential, however, requires progress in identifying individualised treatment strategies and tailoring intervention triggers to the dynamics of each patient.

Recent shifts towards precision medicine have also reached tinnitus research (221). Precision medicine, broadly defined, represents a move towards data-informed tailoring of interventions to the individual patient. Within the UNITI project, an RCT compared different treatment strategies with the overarching aim of developing a clinical decision-support system (CDSS) that can empirically guide clinicians and patients in selecting the most appropriate treatment based on an individual's clinical profile (186,222). However, preliminary results remain limited and especially lack external validation. EMA data offer a promising avenue not only to support such

Discussion

validation efforts but also to fine-tune CDSS by continuously monitoring whether patients remain on track during treatment. A key example comes from psychotherapy research where Lutz and colleagues developed the Trier Treatment Navigator (TTN), a decision-support tool that informs clinicians both prior to and during treatment regarding treatment choices in psychotherapy (223). The TTN employs a nearest-neighbour algorithm to generate predictions based on baseline information, and dynamically updates these predictions through regular patient progress assessments. When a patient deviates from the expected trajectory, the system alerts the therapist and recommends alternative treatment strategies. Translating this vision to tinnitus care could involve a CDSS that recommends, for example, HA based on a patient's individual profile, while regular EMA or passive sensing data provide continuous feedback on treatment response. If the patient fails to respond as predicted, the attending physician could be informed and treatment adaptations initiated. The identification of latent subgroups in the third study can be regarded as a step towards precision medicine, as it provides a framework for tailoring treatment strategies to different patient profiles. Rather than treating tinnitus patients as a homogeneous group, subgroup modelling allows for more nuanced inferences that can inform personalised treatment pathways. Importantly, these longitudinal data could also be fed back into the prediction model, iteratively improving its accuracy and clinical utility over time.

Conclusion

In summary, this thesis demonstrates that daily symptom monitoring through EMA can substantially enrich tinnitus research by capturing the temporal complexity of symptoms and revealing how patients change over the course of treatment. By uniting idiographic precision with nomothetic generalisability through novel analytical approaches, EMA opens the door to a more dynamic, personalised, and clinically relevant understanding of tinnitus.

References

1. Engelke M, Basso L, Langguth B, Zeman F, Schlee W, Schoisswohl S, et al. Estimation of Minimal Clinically Important Difference for Tinnitus Handicap Inventory and Tinnitus Functional Index. *Otolaryngol Head Neck Surg.* 2025;173(1):69–79.
2. Allport G. The general and the unique in psychological science. *Personality.* 1962;30:405–22.
3. Barlow DH, Nock MK. Why Can't We Be More Idiographic in Our Research? *Perspect Psychol Sci.* 2009;4(1):19–21.
4. Böschén I. Changes in methodological study characteristics in psychology between 2010-2021. *PLoS ONE.* 2023;18(5):e0283353.
5. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. *Proc Natl Acad Sci.* 2018;115(27):e6106–15.
6. Langguth B, De Ridder D. Tinnitus, Tinnitus Disorder, and Other Phantom Perceptions. In: Schlee W, Langguth B, De Ridder D, Vanneste S, Kleinjung T, Möller AR, editors. *Textbook of Tinnitus.* Cham: Springer International Publishing; 2024. p. 17–24.
7. De Ridder D, Schlee W, Vanneste S, Londero A, Weisz N, Kleinjung T, et al. Tinnitus and tinnitus disorder: Theoretical and operational definitions (an international multidisciplinary proposal). In: Schlee W, Langguth B, Kleinjung T, Vanneste S, De Ridder D, editors. *Progress in Brain Research.* Amsterdam: Elsevier; 2021. p. 1–25.
8. Langguth B. Tinnitus Heterogeneity, Different Types of Tinnitus, and Gender Aspects. In: Schlee W, Langguth B, De Ridder D, Vanneste S, Kleinjung T, Möller AR, editors. *Textbook of Tinnitus.* Cham: Springer International Publishing; 2024. p. 71–9.
9. Watts EJ, Fackrell K, Smith S, Sheldrake J, Haider H, Hoare DJ. Why Is Tinnitus a Problem? A Qualitative Analysis of Problems Reported by Tinnitus Patients. *Trends in Hearing.* 2018;22:1–10.
10. Langguth B, Landgrebe M. Tinnitus and Psychiatric Comorbidities. In: Schlee W, Langguth B, De Ridder D, Vanneste S, Kleinjung T, Möller AR, editors. *Textbook of Tinnitus.* Cham: Springer International Publishing; 2024. p. 515–25.
11. Park SN. Objective Tinnitus. *Hanyang Med Rev.* 2016;36(2):99–108.

12. Jarach CM, Lugo A, Scala M, Van Den Brandt PA, Cederroth CR, Odone A, et al. Global Prevalence and Incidence of Tinnitus: A Systematic Review and Meta-analysis. *JAMA Neurol.* 2022;79(9):888–900.
13. Langguth B, De Ridder D, Schlee W, Kleinjung T. Tinnitus: Clinical Insights in Its Pathophysiology-A Perspective. *J Assoc Res Otolaryngol.* 2024;25(3):249–58.
14. Biswas R, Genitsaridi E, Trpchevska N, Lugo A, Schlee W, Cederroth CR, et al. Low Evidence for Tinnitus Risk Factors: A Systematic Review and Meta-analysis. *J Assoc Res Otolaryngol.* 2022;24(1):81–94.
15. Eggermont JJ, Roberts LE. The neuroscience of tinnitus. *Trends in Neurosciences.* 2004;27(11):676–82.
16. Shore SE, Roberts LE, Langguth B. Maladaptive plasticity in tinnitus-triggers, mechanisms and treatment. *Nat Rev Neurol.* 2016;12(3):150–60.
17. Elgoyhen AB, Langguth B, De Ridder D, Vanneste S. Tinnitus: perspectives from human neuroimaging. *Nat Rev Neurosci.* 2015;16(10):632–42.
18. Langguth B, Kreuzer PM, Kleinjung T, De Ridder D. Tinnitus: causes and clinical management. *The Lancet Neurology.* 2013;12(9):920–30.
19. De Ridder D, Vanneste S. The Bayesian brain in imbalance: Medial, lateral and descending pathways in tinnitus and pain: A perspective. In: *Progress in Brain Research.* Elsevier; 2021. p. 309–34.
20. Schaette R, McAlpine D. Tinnitus with a Normal Audiogram: Physiological Evidence for Hidden Hearing Loss and Computational Model. *J Neurosci.* 2011;31(38):13452–7.
21. Mohan A, Luckey A, Weisz N, Vanneste S. Predisposition to domain-wide maladaptive changes in predictive coding in auditory phantom perception. *NeuroImage.* 2022;248:118813.
22. Berger JI, Owen W, Wilson CA, Hockley A, Coomber B, Palmer AR, et al. Gap-induced reductions of evoked potentials in the auditory cortex: A possible objective marker for the presence of tinnitus in animals. *Brain Research.* 2018;1679:101–8.
23. Cima RFF, Mazurek B, Haider H, Kikidis D, Lapira A, Noreña A, et al. A multidisciplinary European guideline for tinnitus: diagnostics, assessment, and treatment. *HNO.* 2019;67(1):10–42.
24. Langguth B. Diagnosis and Flowchart. In: Schlee W, Langguth B, De Ridder D, Vanneste S, Kleinjung T, Møller AR, editors. *Textbook of Tinnitus.* Springer; 2024. p. 315–20.

25. Tunkel DE, Bauer CA, Sun GH, Rosenfeld RM, Chandrasekhar SS, Cunningham ER, et al. Clinical Practice Guideline: Tinnitus. *Otolaryngol Head Neck Surg.* 2014;151(2S):S1–40.
26. Langguth B, Gilles A. Tinnitus Questionnaires. In: Schlee W, Langguth B, De Ridder D, Vanneste S, Kleinjung T, Möller AR, editors. *Textbook of Tinnitus.* Springer; 2024. p. 329–43.
27. Viergever K, Kraak JT, Bruinewoud Elsm, Ket JCF, Kramer SE, Merkus P. Questionnaires in otology: a systematic mapping review. *Syst Rev.* 2021;10(1):119.
28. Newman C, Jacobson G, Spitzer J. Development of the Tinnitus Handicap Inventory. *Arch Otolaryngol Head Neck Surg.* 1996;122(2):143–8.
29. Meikle MB, Henry JA, Griest SE, Stewart BJ, Abrams HB, McArdle R, et al. The Tinnitus Functional Index: Development of a New Clinical Measure for Chronic, Intrusive Tinnitus. *Ear & Hearing.* 2012;33(2):153–76.
30. Hallam R, Jakes S, Hinchcliffe R. Cognitive variables in tinnitus annoyance. *Br J Clin Psychol.* 1988;27:213–22.
31. Hall DA, Haider H, Szczepek AJ, Lau P, Rabau S, Jones-Diette J, et al. Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults. *Trials.* 2016;17(1):270.
32. Langguth B, De Ridder D. Minimal Clinically Important Difference of Tinnitus Outcome Measurement Instruments—A Scoping Review. *J Clin Med.* 2023;12(22):7117.
33. Schlee W, Pryss R, Probst T, Schobel J, Bachmeier A, Reichert M, et al. Measuring the Moment-to-Moment Variability of Tinnitus: The TrackYourTinnitus Smart Phone App. *Front Aging Neurosci.* 2016;8.
34. Shiffman S, Stone AA, Hufford M. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4:1–32.
35. Husain FT, Gander PE, Jansen JN, Shen S. Expectations for Tinnitus Treatment and Outcomes: A Survey Study of Audiologists and Patients. *J Am Acad Audiol.* 2018;29(4):313–36.
36. Mazurek B, Hesse G, Dobel C, Kratzsch V, Lahmann C, Sattel H. Clinical practice guideline: Chronic tinnitus—diagnosis and treatment. *Dtsch Arztebl Int.* 2022;119:219–25.
37. Fuller T, Cima R, Langguth B, Mazurek B, Vlaeyen J, Hoare D. Cognitive behavioural therapy for tinnitus. *Cochrane Database Syst Rev.* 2020;1(CD012614).

38. Kleinjung T, Peter N, Schecklmann M, Langguth B. The Current State of Tinnitus Diagnosis and Treatment: a Multidisciplinary Expert Perspective. *J Assoc Res Otolaryngol*. 2024;25(5):413–25.
39. Schoisswohl S, Basso L, Simoes J, Engelke M, Langguth B, Mazurek B, et al. Single versus Combination Treatment in Tinnitus: A Randomized, Multicenter Trial. *Nat Commun*. in press;
40. Schlee W, Langguth B, Pryss R, Allgaier J, Mulansky L, Vogel C, et al. Using Big Data to Develop a Clinical Decision Support System for Tinnitus Treatment. In: Searchfield GD, Zhang J, editors. *The Behavioral Neuroscience of Tinnitus*. Cham: Springer International Publishing; 2021. p. 175–89.
41. De Ridder D, De Ridder-Symoens H. History of Tinnitus. In: Schlee W, Langguth B, De Ridder D, Vanneste S, Kleinjung T, Møller AR, editors. *Textbook of Tinnitus*. Springer; 2024. p. 3–16.
42. Flügel JC. A quantitative study of feeling and emotion in everyday life. *Br J Psychol*. 1925;15(4):318–55.
43. Csikszentmihalyi M, Larson R, Prescott S. The ecology of adolescent activity and experience. *J Youth Adolesc*. 1977;6(3):281–94.
44. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. *J Nerv Ment Dis*. 1987;175(9):526–36.
45. Csikszentmihalyi M. *Beyond Boredom and Anxiety: The Experience of Play in Work and Games*. San Francisco: Jossey-Bass; 1975.
46. Stone AA, Shiffman S. Ecological Momentary Assessment (Ema) in Behavioral Medicine. *Ann Behav Med*. 1994;16(3):199–202.
47. Shiffman S, Stone AA. Introduction to the special section: Ecological momentary assessment in health psychology. *Health Psychol*. 1998;17(1):3–5.
48. Doherty K, Balaskas A, Doherty G. The Design of Ecological Momentary Assessment Technologies. *Interact Comput*. 2020;32(3):257–78.
49. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, et al. Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry*. 2018;17(2):123–32.
50. Stieger S, Reips UD. Well-being, Smartphone Sensors, and Data from Open-access Databases: A Mobile Experience Sampling Study. *Field Methods*. 2019;31(3):277–91.

51. Houben M, Van Den Noortgate W, Kuppens P. The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychol Bull.* 2015;141(4):901–30.
52. MacKerron G, Mourato S. Happiness is greater in natural environments. *Glob Environ Change.* 2013;23(5):992–1000.
53. Bakolis I, Hammoud R, Smythe M, Gibbons J, Davidson N, Tognin S, et al. Urban Mind: Using Smartphone Technologies to Investigate the Impact of Nature on Mental Well-Being in Real Time. *BioScience.* 2018;68(2):134–45.
54. Myin-Germeys I, Peeters F, Havermans R, Nicolson NA, DeVries MW, Delespaul P, et al. Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. *Acta Psychiatr Scand.* 2003;107(2):124–31.
55. Oorschot M, Lataster T, Thewissen V, Lardinois M, Wichers M, Van Os J, et al. Emotional Experience in Negative Symptoms of Schizophrenia—No Evidence for a Generalized Hedonic Deficit. *Schizophr Bull.* 2013;39(1):217–25.
56. Triantafyllou S, Saeb S, Lattie EG, Mohr DC, Kording KP. Relationship Between Sleep Quality and Mood: Ecological Momentary Assessment Study. *JMIR Ment Health.* 2019;6(3):e12613.
57. Thewissen V, Bentall RP, Oorschot M, à Campo J, van Lierop T, van Os J, et al. Emotions, self-esteem, and paranoid episodes: An experience sampling study. *Br J Clin Psychol.* 2011;50(2):178–95.
58. Barge-Schaapveld D. Changes in daily life experience associated with clinical improvement in depression. *J Affect Disord.* 1995;34(2):139–54.
59. Geschwind N, Nicolson NA, Peeters F, van Os J, Barge-Schaapveld D, Wichers M. Early improvement in positive rather than negative emotion predicts remission from depression after pharmacotherapy. *Eur Neuropsychopharmacol.* 2011;21(3):241–7.
60. Barge-Schaapveld DQCM, Nicolson NA. Effects of Antidepressant Treatment on the Quality of Daily Life: An Experience Sampling Study. *J Clin Psychiatry.* 2002;
61. Ram N, Brinberg M, Pincus AL, Conroy DE. The Questionable Ecological Validity of Ecological Momentary Assessment: Considerations for Design and Analysis. *Res Hum Dev.* 2017;14(3):253–70.
62. Ben-Zeev D, Young MA. Accuracy of Hospitalized Depressed Patients' and Healthy Controls' Retrospective Symptom Reports: An Experience Sampling Study. *J Nerv Ment Dis.* 2010;198(4):280–5.

63. Stone AA, Schneider S, Smyth JM. Evaluation of Pressing Issues in Ecological Momentary Assessment. *Annu Rev Clin Psychol*. 2023;19:1.1-1.25.
64. Scott B, Lindberg P, Lyttkens L, Melin L. Psychological Treatment of Tinnitus. An Experimental Group Study. *Scand Audiol*. 1985;14(4):223–30.
65. Henry J, Galvez G, Turbin M, Thielman E, McMillan G, Istvan J. Pilot Study to Evaluate Ecological Momentary Assessment of Tinnitus. *Ear Hear*. 2012;33(2):179–290.
66. Goldberg R, Piccirillo M, Nicklaus J, Skillington A, Lenze E, Rodebaugh T, et al. Evaluation of Ecological Momentary Assessment for Tinnitus Severity. *JAMA Otolaryngol Head Neck Surg*. 2017;143(7):700.
67. Gerull K, Kallogjeri D, Piccirillo M, Rodebaugh T, Lenze E, Piccirillo J. Feasibility of Intensive Ecological Sampling of Tinnitus in Intervention Research. *Otolaryngol Head Neck Surg*. 2019;161(3):485–92.
68. Lourenco MPCG, Cima RFF, Vlaeyen JWS. Chapter 8 - Effects of ecological momentary assessment (EMA) induced monitoring on tinnitus experience: A multiple-baseline single-case experiment. In: Langguth B, Kleinjung T, Ridder DD, Schlee W, Vanneste S, editors. *Progress in Brain Research*. Elsevier; 2021. p. 153–70.
69. Lourenco MPCG, Simoes J, Vlaeyen JWS, Cima RFF. The Daily Experience of Subjective Tinnitus: Ecological Momentary Assessment Versus End-of-Day Diary. *Ear Hear*. 2022;43(1):45–52.
70. Pryss R, Probst T, Schlee W, Schobel J, Langguth B, Neff P, et al. Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus–stress associations based on the TrackYourTinnitus mobile platform. *Int J Data Sci Anal*. 2019;8(4):327–38.
71. Dode A, Mehdi M, Rüdiger Pryss, Schlee W, Probst T, Reichert M, et al. Using a visual analog scale (VAS) to measure tinnitus-related distress and loudness: Investigating correlations using the Mini-TQ results of participants from the TrackYourTinnitus platform. *Prog Brain Res*. 2021;263:171–90.
72. Wilson M, Kallogjeri D, Joplin C, Gorman M, Krings JG, Lenze E, et al. Ecological Momentary Assessment of Tinnitus Using Smartphone Technology: A Pilot Study. *Otolaryngol Head Neck Surg*. 2015;152(5):897–903.
73. Probst T, Pryss R, Langguth B, Schlee W. Emotional states as mediators between tinnitus loudness and tinnitus distress in daily life: Results from the “TrackYourTinnitus” application. *Sci Rep*. 2016;6(1):20382.

74. Probst T, Pryss R, Langguth B, Schlee W. Emotion dynamics and tinnitus: Daily life data from the “TrackYourTinnitus” application. *Sci Rep*. 2016;6(1):31166.
75. Probst T, Pryss R, Langguth B, Rauschecker J, Schobel J, Reichert M, et al. Does Tinnitus Depend on Time-of-Day? An Ecological Momentary Assessment Study with the “TrackYourTinnitus” Application. *Front Aging Neurosci*. 2017;9:253.
76. Simoes J, Bulla J, Neff P, Pryss R, Marcum S, Langguth B, et al. Daily Contributors of Tinnitus Loudness and Distress: An Ecological Momentary Assessment Study. *Front Neurosci*. 2022;16:883665.
77. Kraft R, Langguth B, Simoes J, Reichert M, Schlee W, Pryss R. Global 10 year ecological momentary assessment and mobile sensing study on tinnitus and environmental sounds. *npj Digit Med*. 2025;8(1):162.
78. Muniandi LP, Schlee W, Pryss R, Reichert M, Schobel J, Kraft R, et al. Finding Tinnitus Patients with Similar Evolution of Their Ecological Momentary Assessments. In: 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS). Karlstad: IEEE; 2018. p. 112–7.
79. Pryss R, Schlee W, Hoppenstedt B, Reichert M, Spiliopoulou M, Langguth B, et al. Applying Machine Learning to Daily-Life Data From the TrackYourTinnitus Mobile Health Crowdsensing Platform to Predict the Mobile Operating System Used With High Accuracy: Longitudinal Observational Study. *J Med Internet Res*. 2020;22(6):e15547.
80. Schleicher M, Unnikrishnan V, Neff P, Simoes J, Probst T, Pryss R, et al. Understanding adherence to the recording of ecological momentary assessments in the example of tinnitus monitoring. *Sci Rep*. 2020;10(1):22459.
81. Unnikrishnan V, Schleicher M, Shah Y, Jamaludeen N, Pryss R, Schobel J, et al. The Effect of Non-Personalised Tips on the Continued Use of Self-Monitoring mHealth Applications. *Brain Sci*. 2020;10(12):924.
82. Prakash S, Unnikrishnan V, Pryss R, Kraft R, Schobel J, Hannemann R, et al. Interactive System for Similarity-Based Inspection and Assessment of the Well-Being of mHealth Users. *Entropy (Switzerland)*. 2021;23(12):1695.
83. Shahania S, Unnikrishnan V, Pryss R, Kraft R, Schobel J, Hannemann R, et al. Predicting Ecological Momentary Assessments in an App for Tinnitus by Learning From Each User’s Stream With a Contextual Multi-Armed Bandit. *Front Neurosci*. 2022;16.
84. Kleinau A, Flügel S, Pryss R, Vogel C, Engelke M, Schlee W, et al. Predicting Patient-Based Time-Dependent Mobile Health Data. In: 2023 IEEE 36th International

Symposium on Computer-Based Medical Systems (CBMS). L'Aquila, Italy: IEEE; 2023. p. 79–84.

85. Allgaier J, Schlee W, Probst T, Pryss R. Prediction of Tinnitus Perception Based on Daily Life MHealth Data Using Country Origin and Season. *J Clin Med*. 2022;11(15):4270.

86. Breitmayer M, Stach M, Kraft R, Allgaier J, Reichert M, Schlee W, et al. Predicting the presence of tinnitus using ecological momentary assessments. *Sci Rep*. 2023;13(1):8989.

87. Kraft R, Hofmann F, Reichert M, Pryss R. Dealing With Inaccurate Sensor Data in the Context of Mobile Crowdsensing and mHealth. *IEEE J Biomed Health Inform*. 2022;26(11):5439–49.

88. Winter M, Langguth B, Schlee W, Pryss R. Process mining in mHealth data analysis. *npj Digit Med*. 2024;7(1):299.

89. Schlee W, Simoes J, Pryss R. Auricular Acupressure Combined with Self-Help Intervention for Treating Chronic Tinnitus: A Longitudinal Observational Study. *J Clin Med*. 2021;10(18):4201.

90. Baguley D, McFerran D, Hall D. Tinnitus. *Lancet*. 2013;382(9904):1600–7.

91. Biswas R, Lugo A, Akeroyd M, Schlee W, Gallus S, Hall D. Tinnitus prevalence in Europe: a multi-country cross-sectional population study. *Lancet Reg Health Eur*. 2022;12:100250.

92. Jarach C, Lugo A, Scala M, Brandt P, Cederroth C, Odone A, et al. Global Prevalence and Incidence of Tinnitus: A Systematic Review and Meta-analysis. *JAMA Neurology*. 2022 Aug 8;79.

93. Cederroth C, Gallus S, Hall D, Kleinjung T, Langguth B, Maruotti A, et al. Editorial: Towards an Understanding of Tinnitus Heterogeneity. *Front Aging Neurosci*. 2019;11:53.

94. Møller A. Different forms of tinnitus. In: Møller A, Langguth B, De Ridder D, Kleinjung T, editors. *Textbook of Tinnitus*. New York, NY: Springer New York; 2011. p. 9–12.

95. Goodey R. Introduction. In: Møller A, Langguth B, De Ridder D, Kleinjung T, editors. *Textbook of Tinnitus*. New York, NY: Springer New York; 2011. p. 523–6.

96. Langguth B, Landgrebe M, Kleinjung T, Sand P, Hajak G. Tinnitus and depression. *World J Biol Psychiatry*. 2011;12(7):489–500.

97. Stobik C, Weber R, Münte T, Walter M, Frommer J. Evidence of psychosomatic influences in compensated and decompensated tinnitus. *Int J Audiol*. 2005;44(6):370–8.
98. Hallam R, Rachman S, Hinchcliffe R. Psychological aspects of tinnitus. In: Rachman S, editor. *Contributions to Medical Psychology*. 1984. p. 31–53.
99. Searchfield G, Magnusson J, Shakes G, Biesinger E, Kong O. Counseling and Psycho-Education for Tinnitus Management. In: Møller A, Langguth B, De Ridder D, Kleinjung T, editors. *Textbook of Tinnitus*. New York: Springer; 2011. p. 535–56.
100. McKenna L, Handscomb L, Hoare D, Hall D. A Scientific Cognitive-Behavioral Model of Tinnitus: Novel Conceptualizations of Tinnitus Distress. *Front Neurol*. 2014;5.
101. Coles R, Hallam R. Tinnitus and its management. *Br Med Bull*. 1987;43(4):983–98.
102. Kleinjung T, Langguth B. Avenue for Future Tinnitus Treatments. *Otolaryngol Clin North Am*. 2020;53(4):667–83.
103. Weise C, Kleinstäuber M, Andersson G. Internet-Delivered Cognitive-Behavior Therapy for Tinnitus: A Randomized Controlled Trial. *Psychosom Med*. 2016;78(4):501–10.
104. Hoare D, Gander P, Collins L, Smith S, Hall D. Management of tinnitus in English NHS audiology departments: an evaluation of current practice: Tinnitus management in England. *J Eval Clin Pract*. 2012;18(2):326–34.
105. Aazh H, Swanepoel D, Moore B. Telehealth tinnitus therapy during the COVID-19 outbreak in the UK: uptake and related factors. *Int J Audiol*. 2021;60(5):322–7.
106. Trochidis I, Lugo A, Borroni E, Cederroth C, Cima R, Kikidis D, et al. Systematic Review on Healthcare and Societal Costs of Tinnitus. *Int J Environ Res Public Health*. 2021;18(13):6881.
107. Maes I, Cima R, Vlaeyen J, Anteunis L, Joore M. Tinnitus: A Cost Study. *Ear Hear*. 2013;34(4):508–14.
108. Messner E, Probst T, O'Rourke T, Stoyanov S, Baumeister H. mHealth Applications: Potentials, Limitations, Current Quality and Future Directions. In: Baumeister H, Montag C, editors. *Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics*. Cham: Springer International Publishing; 2019. p. 235–48. (Studies in Neuroscience, Psychology and Behavioral Economics).

109. Bendig E, Bauereiß N, Ebert D, Snoek F, Andersson G, Baumeister H. Internet-Based Interventions in Chronic Somatic Disease. *Dtsch Arztebl Int.* 2018;(115):659–65.
110. Mehdi M, Dode A, Pryss R, Schlee W, Reichert M, Hauck F. Contemporary Review of Smartphone Apps for Tinnitus Management and Treatment. *Brain Sci.* 2020;10(11):867.
111. Oron Y, Ben David B, Doron G. Brief cognitive-behavioral training for tinnitus relief using a mobile application: A pilot open trial. *Health Inform J.* 2022;28(1):1–12.
112. Kutyla J, Gos E, Jędrzejczak W, Raj-Koziak D, Karpiesz L, Niedzialek I, et al. Effectiveness of tinnitus therapy using a mobile application. *Eur Arch Otorhinolaryngol.* 2022;279(3):1257–67.
113. Henry J, Thielman E, Zaugg T, Kaelin C, Choma C, Chang B, et al. Development and field testing of a smartphone “App” for tinnitus management. *Int J Audiol.* 2017;56(10):784–92.
114. Kim S, Chang M, Hong M, Yoo S, Oh D, Park M. Tinnitus therapy using tailor-made notched music delivered via a smartphone application and Ginko combined treatment: A pilot study. *Auris Nasus Larynx.* 2017;44(5):528–33.
115. Schlee W, Neff P, Simoes J, Langguth B, Schoisswohl S, Steinberger H, et al. Smartphone-Guided Educational Counseling and Self-Help for Chronic Tinnitus. *J Clin Med.* 2022;(11):1825.
116. Xiang T, Zhong J, Lu T, Pu J mei, Liu L, Xiao Y, et al. Effect of educational counseling alone on people with tinnitus: Meta-analysis of randomized controlled trials. *Patient Educ Couns.* 2020;103(1):44–54.
117. Bauer C, Brozoski T. Effect of Tinnitus Retraining Therapy on the Loudness and Annoyance of Tinnitus: A Controlled Trial. *Ear Hear.* 2011;32(2):145–55.
118. Henry J, Schechter M, Zaugg T, Griest S, Jastreboff P, Vernon J, et al. Clinical trial to compare tinnitus masking and tinnitus retraining therapy. *Acta Otolaryngol.* 2006;126(sup556):64–9.
119. Johansen J, Skellgaard P, Caporali S. Effect of Counseling, Amplification and Fractal Tones in Tinnitus Management. *Commun Disord Deaf Stud Hearing Aids.* 2014;02(04).
120. Lobo M, Moeyaert M, Baraldi Cunha A, Babik I. Single-Case Design, Analysis, and Quality Assessment for Intervention Research. *J Neurol Phys Ther.* 2017;41(3):187–97.

121. Krasny-Pacini A, Evans J. Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Ann Phys Rehabil Med*. 2018;61(3):164–79.
122. Tate R, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, et al. Revision of a method quality rating scale for single-case experimental designs and *n*-of-1 trials: The 15-item Risk of Bias in *N*-of-1 Trials (RoBiNT) Scale. *Neuropsychol Rehabil*. 2013;23(5):619–38.
123. Levin J, Ferron J, Gafurov B. Comparison of randomization-test procedures for single-case multiple-baseline designs. *Dev Neurorehabil*. 2018;21(5):290–311.
124. Koehler M, Levin J. Regulated randomization: A potentially sharper analytical tool for the multiple-baseline design. *Psychol Methods*. 1998;3(2):206–17.
125. Schlee W, Schoisswohl S, Staudinger S, Schiller A, Lehner A, Langguth B, et al. Towards a unification of treatments and interventions for tinnitus patients: The EU research and innovation action UNITI. In: Schlee W, Langguth B, Kleinjung T, Vanneste S, De Ridder D, editors. *Tinnitus - An Interdisciplinary Approach Towards Individualized Treatment*. Amsterdam: Elsevier; 2021. p. 441–51. (Progress in Brain Research).
126. Vogel C, Schobel J, Schlee W, Engelke M, Pryss R. UNITI Mobile—EMI-Apps for a Large-Scale European Study on Tinnitus. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Mexico: IEEE; 2021. p. 2358–62.
127. Nelting M, Rienhoff N, Hesse G, Lamparter U. The assessment of subjective distress related to hyperacusis with a self-rating questionnaire on hypersensitivity to sound. *Laryngo Rhino Otol*. 2002;81(5):327–34.
128. Kroenke K, Spitzer R, Williams J. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–13.
129. Genitsaridi E, Partyka M, Gallus S, Lopez-Escamez J, Schecklmann M, Mielczarek M, et al. Standardised profiling for tinnitus research: The European School for Interdisciplinary Tinnitus Research Screening Questionnaire (ESIT-SQ). *Hear Res*. 2019;377:353–9.
130. RDocumentation. *aregImpute: Multiple Imputation using Additive Regression, Bootstrapping, and Predictive Mean Matching* [Internet]. [cited 2022 July 6]. Available from: <https://www.rdocumentation.org/packages/Hmisc/versions/4.7-0/topics/aregImpute>

131. Kroenke K, Spitzer R. The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatr Ann.* 2002;32(9):509–15.
132. McCombe A, Baguley D, Coles R, McKenna L, McKinney C, Windle-Taylor P. Guidelines for the grading of tinnitus severity: the results of a working group commissioned by the British Association of Otolaryngologists, Head and Neck Surgeons, 1999. *Clin Otolaryngol.* 2001;26(5):388–93.
133. Zeman F, Koller M, Figueiredo R, Aazevedo A, Rates M, Coelho C, et al. Tinnitus Handicap Inventory for Evaluating Treatment Effects: Which Changes Are Clinically Relevant? *Otolaryngol Head Neck Surg.* 2011;145(2):282–7.
134. Adamchic I, Langguth B, Hauptmann C, Tass P. Psychometric Evaluation of Visual Analog Scale for the Assessment of Chronic Tinnitus. *Am J Audiol.* 2012;21(2):215–25.
135. Wilbert J. Using the SCDA package (0.8) for analysing single and multiple case AB designs [Internet]. 2014. Available from: https://www.uni-potsdam.de/fileadmin/projects/inklusion/PDFs/SCDA_Documentation.pdf
136. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis. *J Med Internet Res.* 2020;22(9):e20283.
137. Engelke M, Schlee W. Welche Therapie für welchen Tinnitus? *HNO Nachrichten.* 2021;51(3):48–9.
138. Flückiger C, Del Re A, Wampold B, Horvath A. The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy (Chic).* 2018;55(4):316.
139. Ebert D, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. *J Affect Disord.* 2015;176:9–17.
140. Mira A, Soler C, Alda M, Baños R, Castilla D, Castro A, et al. Exploring the Relationship Between the Acceptability of an Internet-Based Intervention for Depression in Primary Care and Clinical Outcomes: Secondary Analysis of a Randomized Controlled Trial. *Front Psychiatry.* 2019;10:325.
141. Schlee W, Kraft R, Schobel J, Langguth B, Probst T, Neff P, et al. Momentary Assessment of Tinnitus—How Smart Mobile Applications Advance Our Understanding of Tinnitus. In: Baumeister H, Montag C, editors. *Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics.* Cham: Springer; 2019. p. 209–20.

142. Weise C, Kleinstäuber M, Hesser H, Westin V, Gerhard Andersson. Acceptance of Tinnitus: Validation of the Tinnitus Acceptance Questionnaire. *Cogn Behav Ther*. 2013;42(2):100–15.
143. Niemann U, Boecking B, Brueggemann P, Mazurek B, Spiliopoulou M. Gender-Specific Differences in Patients With Chronic Tinnitus—Baseline Characteristics and Treatment Effects. *Front Neurosci*. 2020;14:487.
144. Seydel C, Haupt H, Olze H, Szczepek A, Mazurek B. Gender and Chronic Tinnitus: Differences in Tinnitus-Related Distress Depend on Age and Duration of Tinnitus. *Ear Hear*. 2013;34(5):661–72.
145. Ariel B, Sutherland A, Bland M. The trick does not work if you have already seen the gorilla: how anticipatory effects contaminate pre-treatment measures in field experiments. *J Exp Criminol*. 2021;17(1):55–66.
146. Targum S, Sauder C, Evans M, Saber J, Harvey P. Ecological momentary assessment as a measurement tool in depression trials. *J Psychiatr Res*. 2021;136:256–64.
147. Moore RC, Depp CA, Wetherell JL, Lenze E. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. *J Psychiatr Res*. 2016;75:116–23.
148. Jarach CM, Karydou K, Trochidis I, Bernal-Robledano A, Van Den Brandt PA, Cima R, et al. The out-of-pocket expenses of people with tinnitus in Europe. 2024;
149. Churrua K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect*. 2021;24(4):1015–24.
150. Trull TJ, Ebner-Priemer UW. Using Experience Sampling Methods/Ecological Momentary Assessment (ESM/EMA) in Clinical Assessment and Clinical Research: Introduction to the Special Section. *Psychol Assess*. 2009;21(4):457–62.
151. Mofsen AM, Rodebaugh TL, Nicol GE, Depp CA, Miller JP, Lenze EJ. When All Else Fails, Listen to the Patient: A Viewpoint on the Use of Ecological Momentary Assessment in Clinical Trials. *JMIR Ment Health*. 2019;6(5):e11845.
152. Yim SJ, Lui LMW, Lee Y, Rosenblat JD, Ragguett RM, Park C, et al. The utility of smartphone-based, ecological momentary assessment for depressive symptoms. *J Affect Disord*. 2020;274:602–9.

153. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, Os J van. Experience sampling research in psychopathology: opening the black box of daily life. *Psychol Med*. 2009;39(9):1533–47.
154. So SHW, Peters ER, Swendsen J, Garety PA, Kapur S. Changes in delusions in the early phase of antipsychotic treatment – An experience sampling study. *Psychiatry Res*. 2014;215(3):568–73.
155. Lataster J, Van Os J, De Haan L, Thewissen V, Bak M, Lataster T, et al. Emotional Experience and Estimates of D₂ Receptor Occupancy in Psychotic Patients Treated With Haloperidol, Risperidone, or Olanzapine: An Experience Sampling Study. *J Clin Psychiatry*. 2011;72(10):1397–404.
156. Geschwind N, Peeters F, Drukker M, Van Os J, Wichers M. Mindfulness training increases momentary positive emotions and reward experience in adults vulnerable to depression: A randomized controlled trial. *J Consult Clin Psychol*. 2011;79(5):618–28.
157. Forbes EE, Stepp SD, Dahl RE, Ryan ND, Whalen D, Axelson DA, et al. Real-World Affect and Social Context as Predictors of Treatment Response in Child and Adolescent Depression and Anxiety: An Ecological Momentary Assessment Study. *J Child Adolesc Psychopharmacol*. 2012;22(1):37–47.
158. Peeters F, Berkhof J, Rottenberg J, Nicolson NA. Ambulatory emotional reactivity to negative daily life events predicts remission from major depressive disorder. *Behav Res Ther*. 2010;48(8):754–60.
159. Wichers M, Lothmann C, Simons CJP, Nicolson NA, Peeters F. The dynamic interplay between negative and positive emotions in daily life predicts response to treatment in depression: A momentary assessment study. *Br J Clin Psychol*. 2012;51(2):206–22.
160. Oreel TH, Delespaul P, Hartog ID, Henriques JPS, Netjes JE, Vonk ABA, et al. Ecological momentary assessment versus retrospective assessment for measuring change in health-related quality of life following cardiac intervention. *J Patient Rep Outcomes*. 2020;4(1):98.
161. Peterson CB, Engel SG, Crosby RD, Strauman T, Smith TL, Klein M, et al. Comparing integrative cognitive-affective therapy and guided self-help cognitive-behavioral therapy to treat binge-eating disorder using standard and naturalistic momentary outcome measures: A randomized controlled trial. *Int J Eat Disord*. 2020;53(9):1418–27.

162. Schneider S, Junghaenel DU, Ono M, Broderick JE, Stone AA. III. Detecting treatment effects in clinical trials with different indices of pain intensity derived from ecological momentary assessment. *J Pain*. 2021;22(4):386–99.
163. Scollon CN, Kim-Prieto C, Scollon CN. Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. *J Happiness Stud*. 2003;4(1):5–34.
164. Jacobs PG, Kaye JA. Ubiquitous Real-World Sensing and Audiology-Based Health Informatics. *J Am Acad Audiol*. 2015;26(9):777–83.
165. Holube I, Von Gablenz P, Bitzer J. Ecological Momentary Assessment in Hearing Research: Current State, Challenges, and Future Directions. *Ear Hear*. 2020;41:79S-90S.
166. Bos FM, Von Klipstein L, Emerencia AC, Veermans E, Verhage T, Snippe E, et al. A Web-Based Application for Personalized Ecological Momentary Assessment in Psychiatric Care: User-Centered Development of the PETRA Application. *JMIR Ment Health*. 2022;9(8):e36430.
167. Guy W. ECDEU assessment manual for psychopharmacology. U.S. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs; 1976. 612 p.
168. Padhi A, Fineberg N. Clinical Global Impression Scales. In: Stolerian IP, editor. *Encyclopedia of Psychopharmacology*. Berlin: Springer; 2010. p. 303–303.
169. Fulcher BD. Feature-based time-series analysis [Internet]. arXiv; 2017 [cited 2023 Oct 4]. Available from: <http://arxiv.org/abs/1709.08055>
170. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. *npj Digital Med*. 2019;2(1):1–3.
171. Jacobson NC, Feng B. Digital phenotyping of generalized anxiety disorder: using artificial intelligence to accurately predict symptom severity using wearable sensors in daily life. *Transl Psychiatry*. 2022;12(1):1–7.
172. Wang L, Yin Z, Puppala M, Ezeana C, Wong K, He T, et al. A Time-Series Feature-Based Recursive Classification Model to Optimize Treatment Strategies for Improving Outcomes and Resource Allocations of COVID-19 Patients. *IEEE J Biomed Health Inform*. 2022;26(7):3323–9.
173. Schoisswohl S, Langguth B, Schecklmann M, Bernal-Robledano A, Boecking B, Cederroth CR, et al. Unification of Treatments and Interventions for Tinnitus Patients

(UNITI): a study protocol for a multi-center randomized clinical trial. *Trials*. 2021;22(1):875.

174. Berk M, Ng F, Dodd S, Callaly T, Campbell S, Bernardo M, et al. The validity of the CGI severity and improvement scales as measures of clinical effectiveness suitable for routine clinical use. *J Eval Clin Pract*. 2008;14(6):979–83.

175. Henderson T, Fulcher BD. Feature-Based Time-Series Analysis in R using the theft Package [Internet]. arXiv; 2023 [cited 2024 Jan 30]. Available from: <http://arxiv.org/abs/2208.06146>

176. Wurm MJ, Rathouz PJ, Hanlon BM. Regularized Ordinal Regression and the ordinalNet R Package. *J Stat Softw*. 2021;99(6):10.18637/jss.v099.i06.

177. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. *Frontiers in Econometrics*. New York: Academic Press; 1974. p. 104–42.

178. Goedhart H. Tinnitus Hub. 2023 [cited 2024 Mar 27]. Tinnitus Sufferers Take The Lead in Research. Available from: <https://www.tinnitushub.com/tinnitus-sufferers-take-the-lead-in-research/>

179. Wrzus C, Neubauer AB. Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment*. 2022;1–22.

180. Forkmann T, Scherer A, Boecker M, Pawelzik M, Jostes R, Gauggel S. The clinical global impression scale and the influence of patient or staff perspective on outcome. *BMC Psychiatry*. 2011;11(1):83.

181. Carrozzino D, Patierno C, Guidi J, Berrocal Montiel C, Cao J, Charlson ME, et al. Clinimetric Criteria for Patient-Reported Outcome Measures. *Psychother Psychosom*. 2021;90(4):222–32.

182. Kraft R, Reichert M, Pryss R. Mobile Crowdsensing in Ecological Momentary Assessment mHealth Studies: A Systematic Review and Analysis. *Sensors (Basel)*. 2024;24(2):472.

183. Hoare DJ, Kowalkowski VL, Kang S, Hall DA. Systematic review and meta-analyses of randomized controlled trials examining tinnitus management. *The Laryngoscope*. 2011;121(7):1555–64.

184. McKenna L, Irwin R. Sound therapy for tinnitus – sacred cow or idol worship?: An investigation of the evidence. *Audiol Med*. 2008;6(1):16–24.

185. Hobson J, Chisholm E, El Refaie A. Sound therapy (masking) in the management of tinnitus in adults. *Cochrane Database Syst Rev*. 2012;2012(11):CD006371.
186. Bromis K, Sarafidis M, Manta O, Kouris I, Vellidou E, Schlee W, et al. Predicting the optimal therapeutic intervention for tinnitus patients using random forest regression: A preliminary study of UNITI's decision support system model. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2022. p. 2655–8.
187. Ivansic D, Palm J, Pantev C, Brüggemann P, Mazurek B, Guntinas-Lichius O, et al. Prediction of treatment outcome in patients suffering from chronic tinnitus – from individual characteristics to early and long-term change. *Journal of Psychosomatic Research*. 2022;157:110794.
188. Simoes J, Neff P, Schoisswohl S, Bulla J, Schecklmann M, Harrison S, et al. Toward Personalized Tinnitus Treatment: An Exploratory Study Based on Internet Crowdsensing. *Front Public Health*. 2019;7:157.
189. Niemann U, Brüeggemann P, Boecking B, Mebus W, Rose M, Spiliopoulou M, et al. Phenotyping chronic tinnitus patients using self-report questionnaire data: cluster analysis and visual comparison. *Sci Rep*. 2020;10(1):16411.
190. Niemann U, Boecking B, Brüeggemann P, Spiliopoulou M, Mazurek B. Heterogeneity in response to treatment across tinnitus phenotypes. *Sci Rep*. 2024;14(1):2111.
191. Cuesta M, Cobo P. Enriched Acoustic Environment as a customized treatment for tinnitus: A non-controlled longitudinal study. *J Otol*. 2024;19(2):63–71.
192. Seydel C, Haupt H, Szczepek AJ, Klapp BF, Mazurek B. Long-Term Improvement in Tinnitus after Modified Tinnitus Retraining Therapy Enhanced by a Variety of Psychological Approaches. *Audiol Neurotol*. 2010;15(2):69–80.
193. De Ridder D, Langguth B, Schlee W. Mourning for Silence: Bereavement and Tinnitus—A Perspective. *J Clin Med*. 2025;14(7):2218.
194. Ulvenes P, Soma CS, Melsom L, Wampold BE. A latent trajectory analysis of inpatient depression treatment. *Psychotherapy (Chic)*. 2022;59(1):113–24.
195. Saunders R, Buckman JEJ, Cape J, Fearon P, Leibowitz J, Pilling S. Trajectories of depression and anxiety symptom change during psychological therapy. *J Affect Disord*. 2019;249:327–35.

196. Espel-Huynh HM, Zhang F, Boswell JF, Thomas JG, Thompson-Brenner H, Juarascio AS, et al. Latent trajectories of eating disorder treatment response among female patients in residential care. *Intl J Eating Disorders*. 2020;53(10):1647–56.
197. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695–9.
198. Soto CJ, John OP. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *J Pers Soc Psychol*. 2017;113(1):117.
199. Jung T, Wickrama K a. S. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Soc Pers Psychol Compass*. 2008;2(1):302–17.
200. Ram N, Grimm KJ. Methods and Measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *Int J Behav Dev*. 2009;33(6):565–76.
201. Van De Schoot R, Sijbrandij M, Winter SD, Depaoli S, Vermunt JK. The GRoLTS-Checklist: Guidelines for Reporting on Latent Trajectory Studies. *Struct Equ Model*. 2017;24(3):451–67.
202. Hudgens S, Floden L, Blackowicz M, Jamieson C, Popova V, Fedgchin M, et al. Meaningful Change in Depression Symptoms Assessed with the Patient Health Questionnaire (PHQ-9) and Montgomery-Åsberg Depression Rating Scale (MADRS) Among Patients with Treatment Resistant Depression in Two, Randomized, Double-blind, Active-controlled Trials of Esketamine Nasal Spray Combined With a New Oral Antidepressant. *J Affect Disord*. 2021;281:767–75.
203. Lin T, Anderson T, Erekson DM, Ogles BM. Trajectories of change in weekly and biweekly therapy. *J Couns Psychol*. 2024;71(1):77–87.
204. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The Personalized Advantage Index: Translating Research on Prediction into Individualized Treatment Recommendations. A Demonstration. *PLoS One*. 2014;9(1):e83875.
205. Ueyama T, Donishi T, Ukai S, Ikeda Y, Hotomi M, Yamanaka N, et al. Brain Regions Responsible for Tinnitus Distress and Loudness: A Resting-State fMRI Study. *PLoS One*. 2013;8(6):e67778.

206. Wallhäußer-Franke E, Brade J, Balkenhol T, D'Amelio R, Seegmüller A, Delb W. Tinnitus: distinguishing between subjectively perceived loudness and tinnitus-related distress. *PLoS One*. 2012;7(4):e34583.
207. Delgadillo J, Lutz W. Precision mental health care for depression [Internet]. 2026. Available from: https://www.researchgate.net/profile/Jaime-Delgadillo/publication/369842004_Precision_Mental_Health_Care_for_Depression/links/6777fdbbc1b01354650b9a65/Precision-Mental-Health-Care-for-Depression.pdf
208. Vanneste S, Congedo M, De Ridder D. Pinpointing a Highly Specific Pathological Functional Connection That Turns Phantom Sound into Distress. *Cereb Cortex*. 2014;24(9):2268–82.
209. Davidoff J, Christensen S, Khalili DN, Nguyen J, IsHak WW. Quality of life in panic disorder: looking beyond symptom remission. *Qual Life Res*. 2012;21(6):945–59.
210. IsHak WW, Greenberg JM, Balayan K, Kapitanski N, Jeffrey J, Fathy H, et al. Quality of life: the ultimate outcome measure of interventions in major depressive disorder. *Harv Rev Psychiatry*. 2011;19(5):229–39.
211. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, et al. The Effects of Sampling Frequency and Questionnaire Length on Perceived Burden, Compliance, and Careless Responding in Experience Sampling Data in a Student Population. *Assessment*. 2022;29(2):136–51.
212. Christensen AP. Towards a network psychometrics approach to assessment: Simulations for redundancy, dimensionality, and loadings. The University of North Carolina at Greensboro; 2020.
213. Lutz W, Schwartz B, Delgadillo J. Measurement-Based and Data-Informed Psychological Therapy. *Annu Rev Clin Psychol*. 2022;18(1):71–98.
214. Deisenhofer AK, Barkham M, Beierl ET, Schwartz B, Aafjes-van Doorn K, Beevers CG, et al. Implementing precision methods in personalizing psychological therapies: Barriers and possible ways forward. *Behav Res Ther*. 2024;172:104443.
215. Degroote L, DeSmet A, De Bourdeaudhuij I, Van Dyck D, Crombez G. Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: a systematic review. *Int J Behav Nutr Phys Act*. 2020;17(1):35.

216. Raj-Koziak D, Gos E, Swierniak W, Rajchel JJ, Karpiesz L, Niedzialek I, et al. Visual Analogue Scales as a Tool for Initial Assessment of Tinnitus Severity: Psychometric Evaluation in a Clinical Population. *Audiol Neurotol*. 2018;23(4):229–37.
217. Bringmann LF. Person-specific networks in psychopathology: Past, present, and future. *Curr Opin Psychol*. 2021;41:59–64.
218. Levinson CA, Cusack C, Brown ML, Smith AR. A network approach can improve eating disorder conceptualization and treatment. *Nat Rev Psychol*. 2022;1(7):419–30.
219. Butler RM, Ortiz AML, Pennesi JL, Crumby EK, Cusack CE, Levinson CA. A pilot randomized controlled trial of transdiagnostic network-informed personalized treatment for eating disorders versus enhanced cognitive behavioral therapy. *Int J Eat Disord*. 2023;56(8):1674–80.
220. Harris LM, Vanzhula IA, Cash ED, Levinson CA, Smith AR. Using network analysis to personalize treatment for individuals with co-occurring restrictive eating disorders and suicidality: a proof-of-concept study. *J Eat Disord*. 2025;13(1):156.
221. Tzounopoulos T, Balaban C, Zitelli L, Palmer C. Towards a Mechanistic-Driven Precision Medicine Approach for Tinnitus. *J Assoc Res Otolaryngol*. 2019;20(2):115–31.
222. Unnikrishnan V, Puga C, Schleicher M, Niemann U, Langguth B, Schoisswohl S, et al. Training and validating a treatment recommender with partial verification evidence. *Artif Intell Med*. 2025;160:103062.
223. Lutz W, Rubel JA, Schwartz B, Schilling V, Deisenhofer AK. Towards integrating personalized feedback research into clinical practice: Development of the Trier Treatment Navigator (TTN). *Behav Res Ther*. 2019;120:103438.

List of Figures

Figure 1. Scientific Articles on EMA in Tinnitus	15
Figure 2. CONSORT Flow Diagram.....	29
Figure 3. Improvement of THI from Begin of Baseline Phase to Final Visit.....	31
Figure 4. Course of Tinnitus Distress and Loudness for each Case (EMA)	32
Figure 5. Change of Correlation between Distress and Loudness over the Study Period (EMA)	35
Figure 6. Correlation of Δ THI with Improvement in EMA Tinnitus Distress	37
Figure 7. Distribution of CGI-I.....	52
Figure 8. Clinical Improvement measured by Time-Series Features of the App	53
Figure 9. Latent Trajectories of the Final Model	68
Figure 10. PROM Changes and Treatments across Classes	70
Figure 11. Trajectories of Tinnitus Loudness and Distress across Classes	71
Figure 12. Distribution of the Intra-Individual Standardized Difference (maximum tinnitus loudness – daily tinnitus distress) by Class.....	72
Figure 13. Network Analysis of EMA Items.....	79

List of Tables

Table 1. Participant Demographics, Tinnitus Characteristics and App Use	30
Table 2. Improvement of THI from Begin of Baseline Phase to Final Visit	31
Table 3. Improvement in Tinnitus Distress from Baseline to End of Intervention Phase	33
Table 4. Improvement in Tinnitus Loudness from Baseline to End of Intervention Phase	34
Table 5. Mixed Effect Model	36
Table 6. Description of Time-Series Features	47
Table 7. Patient Characteristics at Baseline (N = 129).....	51
Table 8. Demographic and Clinical Characteristics of Participants at Baseline (N = 147)	66
Table 9. Latent Class Contrasts on Baseline Characteristics	69

Supplementary Material

Table S1. EMA Items and Naming Conventions of Study I – III

Question (EMA/EDD)	VAS scale [lowest anchor - highest anchor]	Naming Study I	Naming Study II (Abbreviation)	Naming Study III
How burdensome do you find your tinnitus at the moment? (EMA)	not burdensome – very burdensome	Tinnitus distress	Momentary tinnitus distress (t-distress)	Momentary tinnitus distress
To what extent did you feel affected by the tinnitus today? (EDD)	not at all – the whole day		Daily tinnitus distress (t-distress-day)	Daily tinnitus distress
How loud is your tinnitus at the moment? (EMA)	inaudible – very loud	Tinnitus loudness	Momentary tinnitus loudness (t-loudness)	Momentary tinnitus loudness
What was the maximum tinnitus volume today? (EDD)	inaudible – very loud		Maximum tinnitus loudness (t-loudness-max)	Maximum tinnitus loudness
How often have you thought about tinnitus today? (EDD)	not at all – the whole day		Tinnitus-related thoughts (t-thoughts)	Tinnitus-related thoughts
What emotion would you use to describe today? (EDD)	[sad emoji] – [happy emoji]		Happiness	Negative Emotion*
How tense does your jaw feel right now? (EMA)	not at all tense – very tense		Jaw tension (jawbone)	Jaw tension
How much did you move today? (EDD)	not at all – very much		Movement	Movement
How tense does your neck feel right now? (EMA)	not at all tense – very tense		Neck tension (neck)	Neck tension
How stressed did you feel today? (EDD)	not at all stressed – very stressed		Stress	Stress

Note. EMA: Ecological Momentary Assessment question; EDD: End-of-day diary question. Questions were answered on a continuous visual analogue scale ranging from 0-100 (numbers were invisible to the patient). *Coding was reversed to standardise interpretation that higher scores indicate higher burden

Table S2. EMA subset Demographics (Study I)

Participant demographics	
Participants	<i>N</i> = 14
Female, <i>n</i> (%)	6 (42.9%)
Age in years, mean (range)	52.7 (33 – 68)
PHQ-9, mean (SD)	5.6 (3.7)
Tinnitus characteristics	
THI, mean (SD)	38.1 (15.8)
Tinnitus duration in months, mean (range)	184 (20 – 350)
GUEF Hyperakusis, mean (SD)	10.8 (6.7)

Table S3. Comparison of 50% App Compliance Subsample with entire RCT Sample (Study II)

	50% compliance (N = 129)	RCT (N = 461)	p value
Sex			
female	61 (47.3%)	190 (41.2%)	0.258
male	68 (52.7%)	271 (58.8%)	
Age (years)			
Mean (SD)	54.8 (12.1)	51.1 (12.4)	0.002
Median [Min, Max]	56.0 [23.0, 77.0]	53.0 [19.0, 77.0]	
Tinnitus duration (months)			
Mean (SD)	132 (117)	119 (113)	0.31
Median [Min, Max]	105 [6.00, 480]	84.0 [6.00, 480]	
THI score baseline			
Mean (SD)	48.6 (19.6)	48.0 (19.7)	0.738
Median [Min, Max]	46.0 [18.0, 94.0]	46.0 [8.00, 96.0]	
TFI score baseline			
Mean (SD)	51.5 (21.2)	48.6 (20.3)	0.159
Median [Min, Max]	54.0 [4.00, 94.0]	49.0 [2.00, 95.0]	
PHQ-9 score baseline			
Mean (SD)	7.73 (5.02)	7.32 (4.87)	0.42
Median [Min, Max]	7.00 [0, 23.0]	7.00 [0, 27.0]	
CGI-I score final visit			
very much better	4 (3.1%)	16 (3.5%)	0.854
much better	22 (17.1%)	56 (12.1%)	
minimally better	33 (25.6%)	112 (24.3%)	
no change	50 (38.8%)	132 (28.6%)	
minimally worse	18 (14.0%)	43 (9.3%)	
much worse	2 (1.6%)	5 (1.1%)	

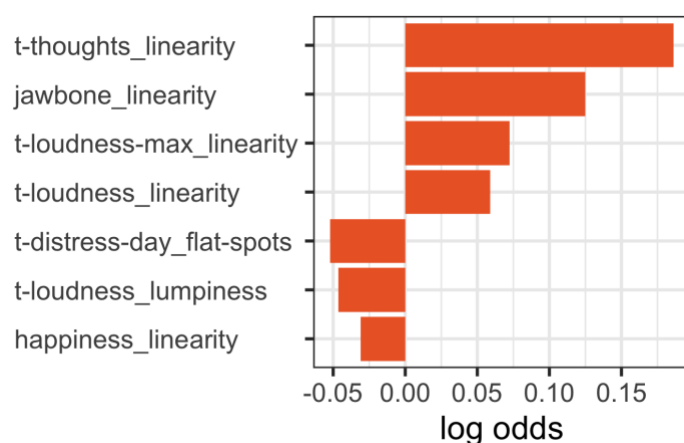
Note. Plus-minus values are means $\pm$ SD. THI = Tinnitus Handicap Inventory; TFI = Tinnitus Functional Index; PHQ-9 = Patient Health Questionnaire for Depression; CGI-I = Clinical Global Impression Scale – Improvement. Final visit: 12 weeks after baseline (treatment end).

Table S4. Coefficients of the Main Model (Study II)

Feature	Log odds	Odds ratio
t-thoughts_linearity	0.18	1.20
jawbone_linearity	0.16	1.17
t-distress-day_flatspots	-0.07	0.93
t-loudness_lumpiness	-0.06	0.95
t-loudness-max_linearity	0.06	1.06
t-loudness_linearity	0.03	1.04
happiness_linearity	-0.03	0.97

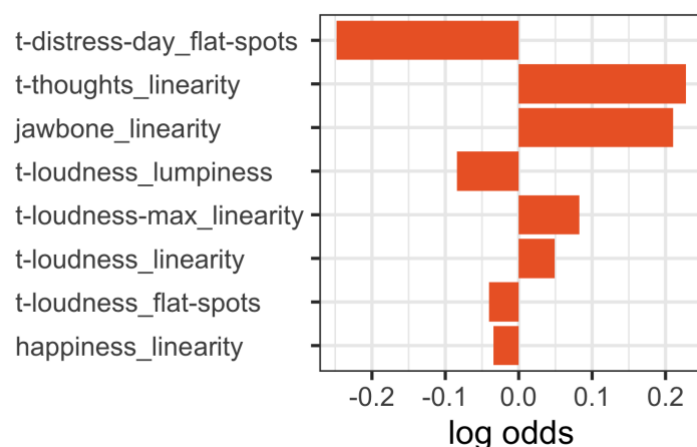
Note. Log odds of the ordinal logistic regression with elastic net regularization and 5-fold cross-validation. Outcome: CGI-I, Predictors: EMA times-series features.

Figure S1. Model results with minimal 55% completion rate (N = 125) (Study II)



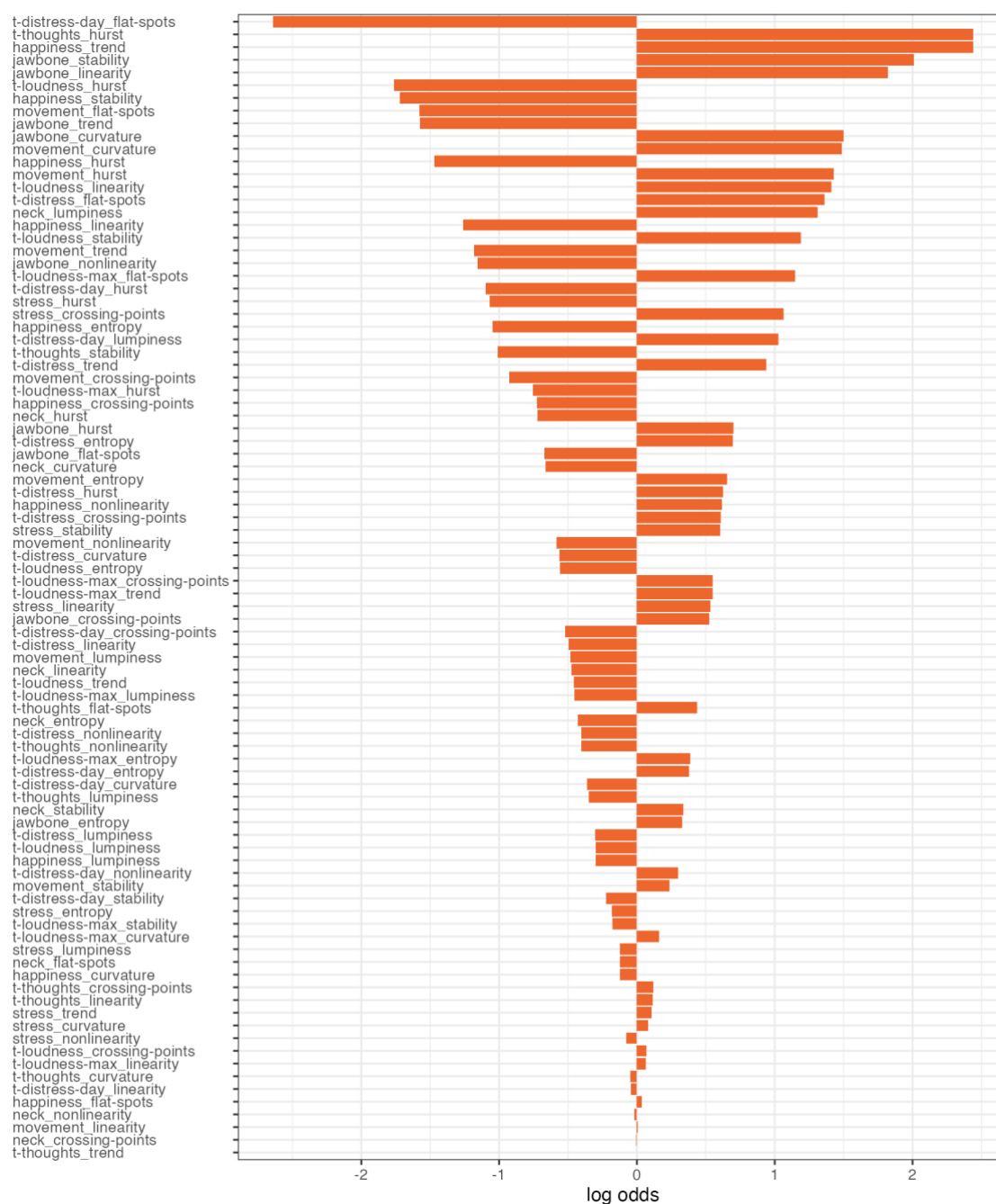
Note. N = 125, Lambda = 0.19, Loglik = -36.4. Log odds of the ordinal logistic regression with elastic net regularization and 5-fold cross-validation. Outcome: CGI-I, Predictors: EMA times-series features.

Figure S2. Model results with minimal 60% completion rate (N = 113) (Study II)



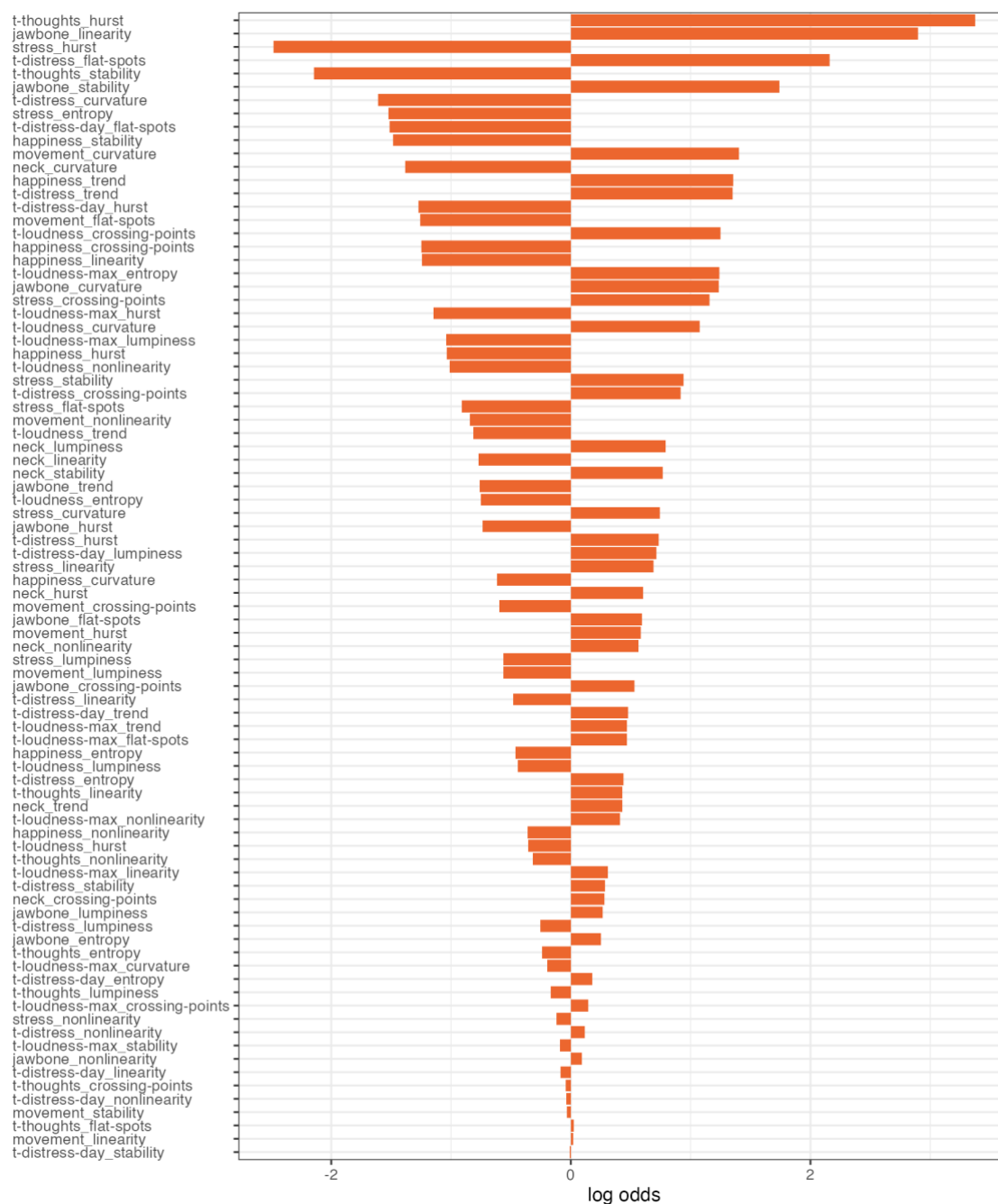
Note. N = 113, Lambda = 0.15, Loglik = -30.8. Log odds of the ordinal logistic regression with elastic net regularization and 5-fold cross-validation. Outcome: CGI-I, Predictors: EMA times-series features.

Figure S3. Model results with minimal 65% completion rate (N = 102) (Study II)



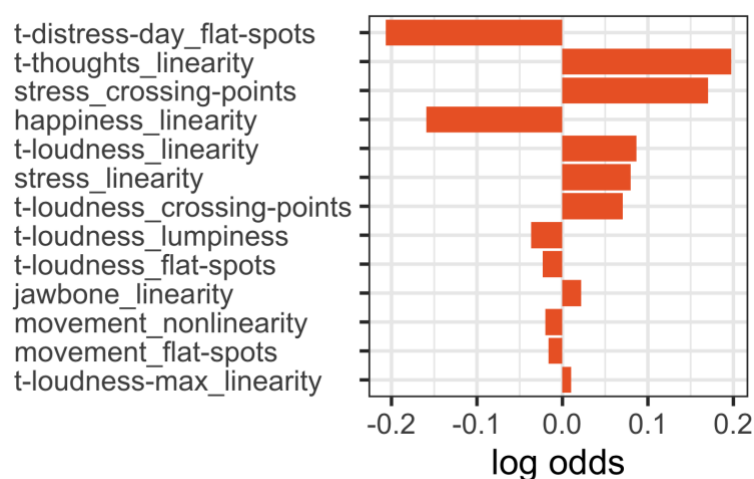
Note. N = 102, Lambda = 0.004, Loglik = -74.8. Log odds of the ordinal logistic regression with elastic net regularization and 5-fold cross-validation. Outcome: CGI-I, Predictors: EMA times-series features.

Figure S4. Model results with minimal 70% completion rate (N = 91) (Study II)



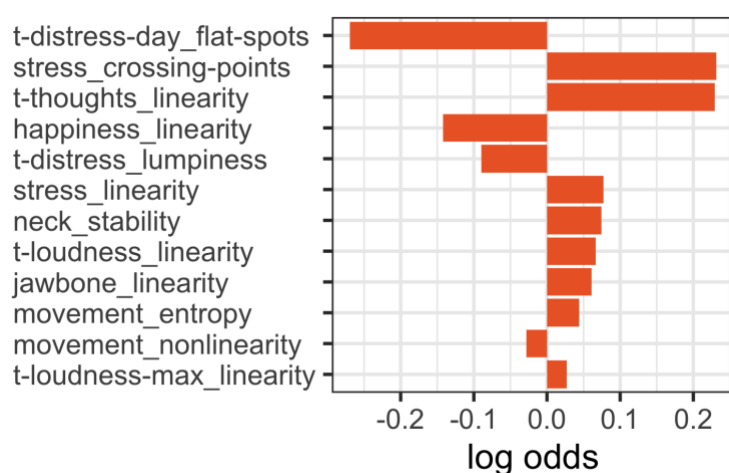
Note. N = 91, Lambda = 0.04, Loglik = -41.3. Log odds of the ordinal logistic regression with elastic net regularization and 5-fold cross-validation. Outcome: CGI-I, Predictors: EMA times-series features.

Figure S5. Model results with minimal 75% completion rate (N = 75) (Study II)



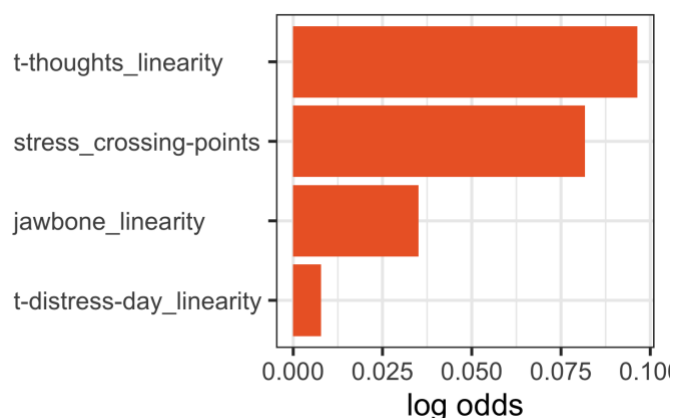
Note. N = 75, Lambda = 0.19, Loglik = -21.2. Log odds of the ordinal logistic regression with elastic net regularization and 5-fold cross-validation. Outcome: CGI-I, Predictors: EMA times-series features.

Figure S6. Model results with minimal 80% completion rate (N = 64) (Study II)



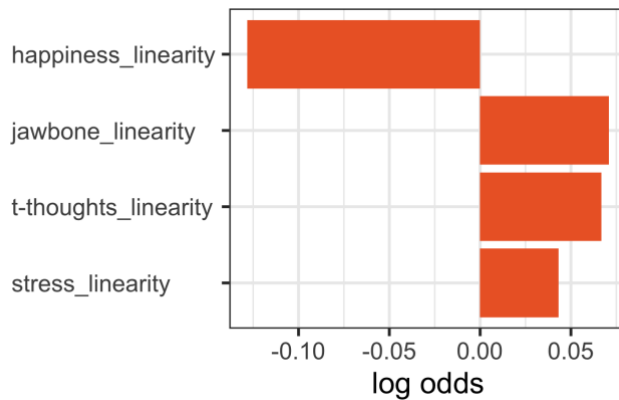
Note. N = 64, Lambda = 0.19, Loglik = -16.0. Log odds of the ordinal logistic regression with elastic net regularization and 5-fold cross-validation. Outcome: CGI-I, Predictors: EMA times-series features.

Figure S7. Model results with minimal 85% completion rate (N = 53) (Study II)



Note. N = 53, Lambda = 0.37, Loglik = -12.9. Log odds of the ordinal logistic regression with elastic net regularization and 5-fold cross-validation. Outcome: CGI-I, Predictors: EMA times-series features.

Figure S8. Model results with minimal 90% completion rate (N = 37) (Study II)



Note. N = 37, Lambda = 0.51, Loglik = -9.0. Log odds of the ordinal logistic regression with elastic net regularization and 5-fold cross-validation. Outcome: CGI-I, Predictors: EMA times-series features.

Figure S9. Distribution of tinnitus-related thoughts (Study III)

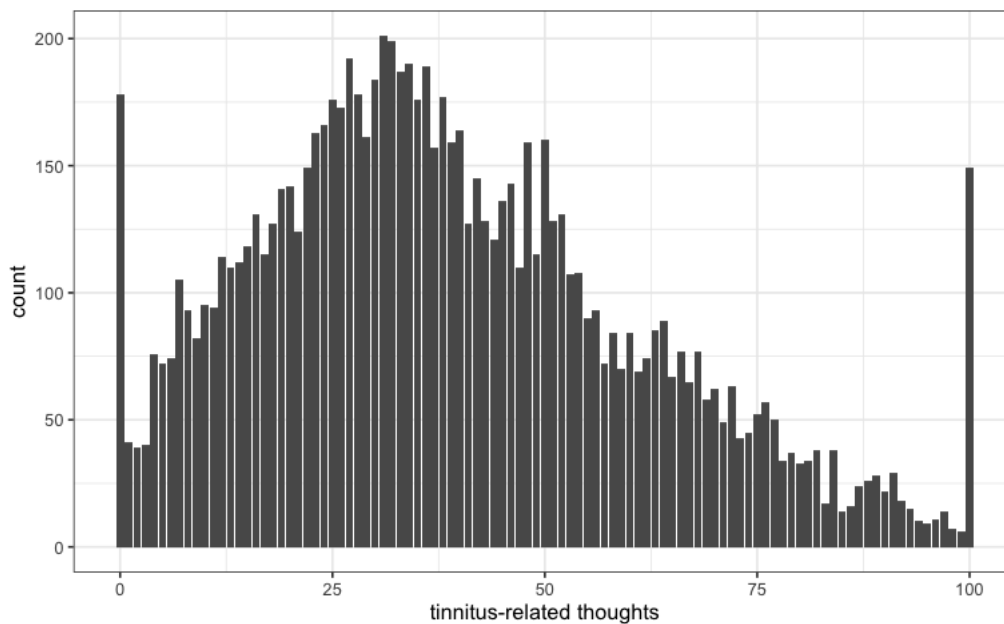


Table S5. Guidelines for Reporting on Latent Trajectory Studies (GRoLTS) Checklist (Study III)

Checklist item	reported
1. Is the metric of time used in the statistical model reported?	yes
2. Is information presented about the mean and variance of time within a wave?	yes
3a. Is the missing data mechanism reported?	yes
3b. Is a description provided of what variables are related to attrition/missing data?	yes
3c. Is a Description Provided for How Missing Data Were Handled in the Analyses?	yes
4. Is information about the distribution of the observed variables included?	yes
5. Is the software mentioned?	yes

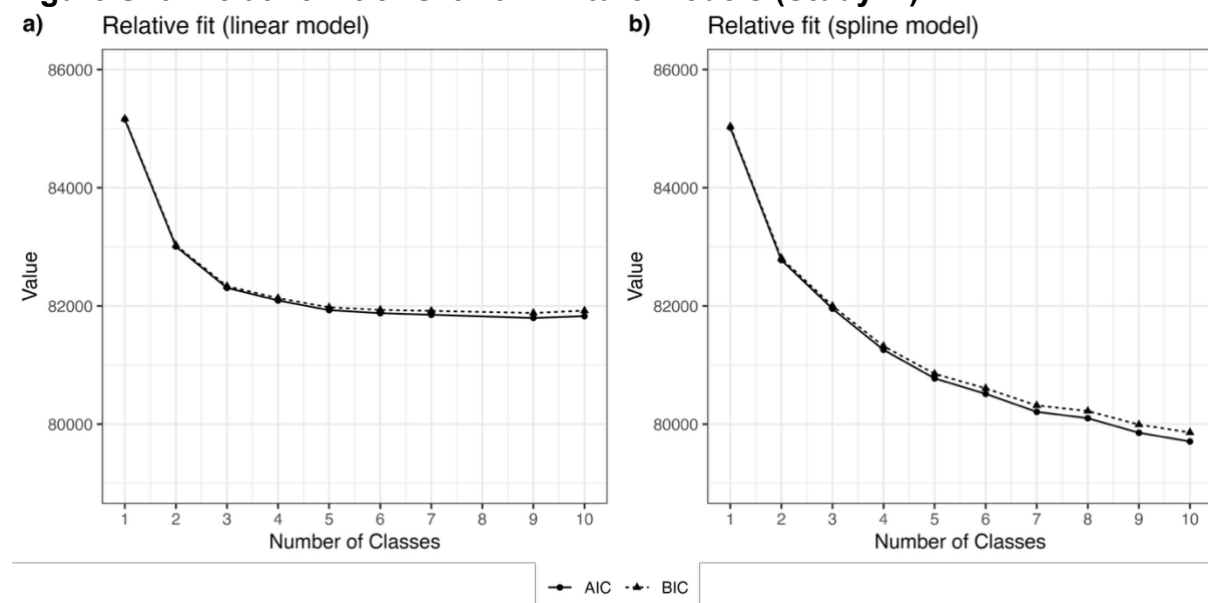
6a. Are alternative specifications of within-class heterogeneity considered (e.g., LGCA vs. LGMM) and clearly documented? If not, was sufficient justification provided as to eliminate certain specifications from consideration?	yes
6b. Are Alternative Specifications of the BetweenClass Differences in Variance–Covariance Matrix Structure Considered and Clearly Documented?	yes
7. Are Alternative Shape and Functional Forms of the Trajectories Described?	yes
8. If Covariates Have Been Used, Can Analyses Still Be Replicated?	yes
9. Is Information Reported About the Number of Random Start Values and Final Iterations Included?	yes
10. Are the model comparison (and selection) tools described from a statistical perspective?	yes
11. Are the total number of fitted models reported, including a one-class solution?	yes
12. Are the number of cases per class reported for each model (absolute sample size, or proportion)?	yes
13. If classification of cases in a trajectory is the goal, is entropy reported?	yes
14a. Is a plot included with the estimated mean trajectories of the final solution?	yes
14b. Are plots included with the estimated mean trajectories for each model?	yes
14c. Is a plot included of the combination of estimated means of the final model and the observed individual trajectories split out for each latent class?	yes
15. Are characteristics of the final class solution numerically described (i.e., means, SD/SE, n, CI, etc.)?	yes
16. Are the syntax files available (either in the appendix, supplementary materials, or from the authors)?	yes

Table S6. Correlation with missingness (Study III)

variable	correlation	p value	adjusted p value
CGI-I	0.06	0.46	1.00
THI score baseline	0.10	0.25	1.00
THI score final visit	0.02	0.85	1.00
TFI score baseline	0.03	0.74	1.00
TFI score final visit	-0.06	0.50	1.00
PHQ9 score baseline	0.05	0.52	1.00
PHQ9 score final visit	0.09	0.30	1.00
Age	-0.26	0.00	0.02
Extraversion	0.02	0.78	1.00
Agreeableness	-0.06	0.47	1.00
Conscientiousness	-0.02	0.85	1.00
Neuroticism	0.06	0.46	1.00
Openness	-0.01	0.86	1.00
Hearing loss (PTA4)	-0.03	0.70	1.00

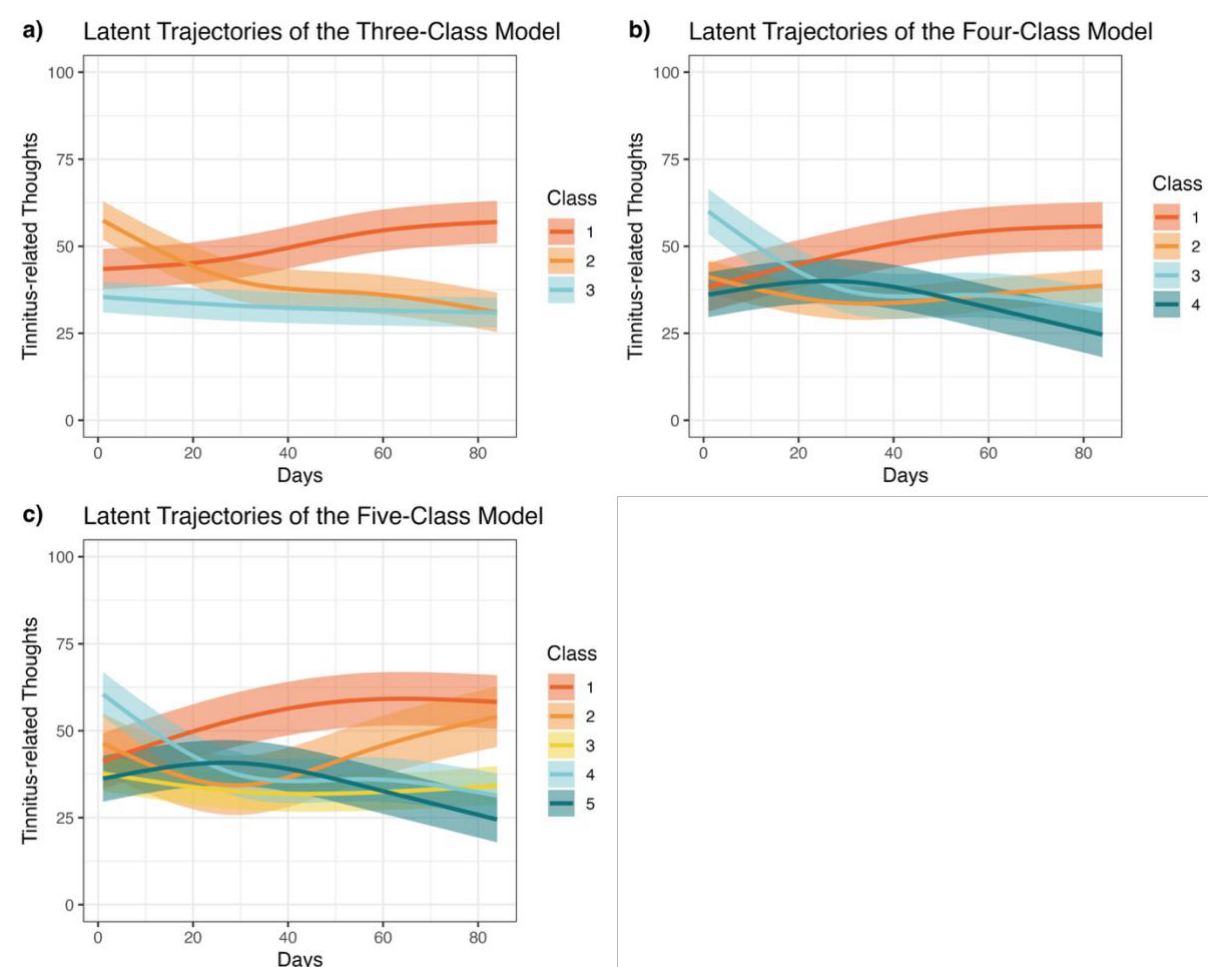
Note. Correlation of patient characteristics with amount of missingness in EMA data to check for a systematic pattern of missingness. P values were corrected according to Holm.

Figure S10. Relative Fit of Growth Mixture Models (Study III)



Notes. Better relative fit of spline models than linear models. Due to asymptotic reduction of AIC and BIC among spline models, more weight was put on clinical interpretability in the selection of the final model. The linear eight-class model did not converge.

Figure S11. Model Selection Based on Clinical Interpretability (Study III)



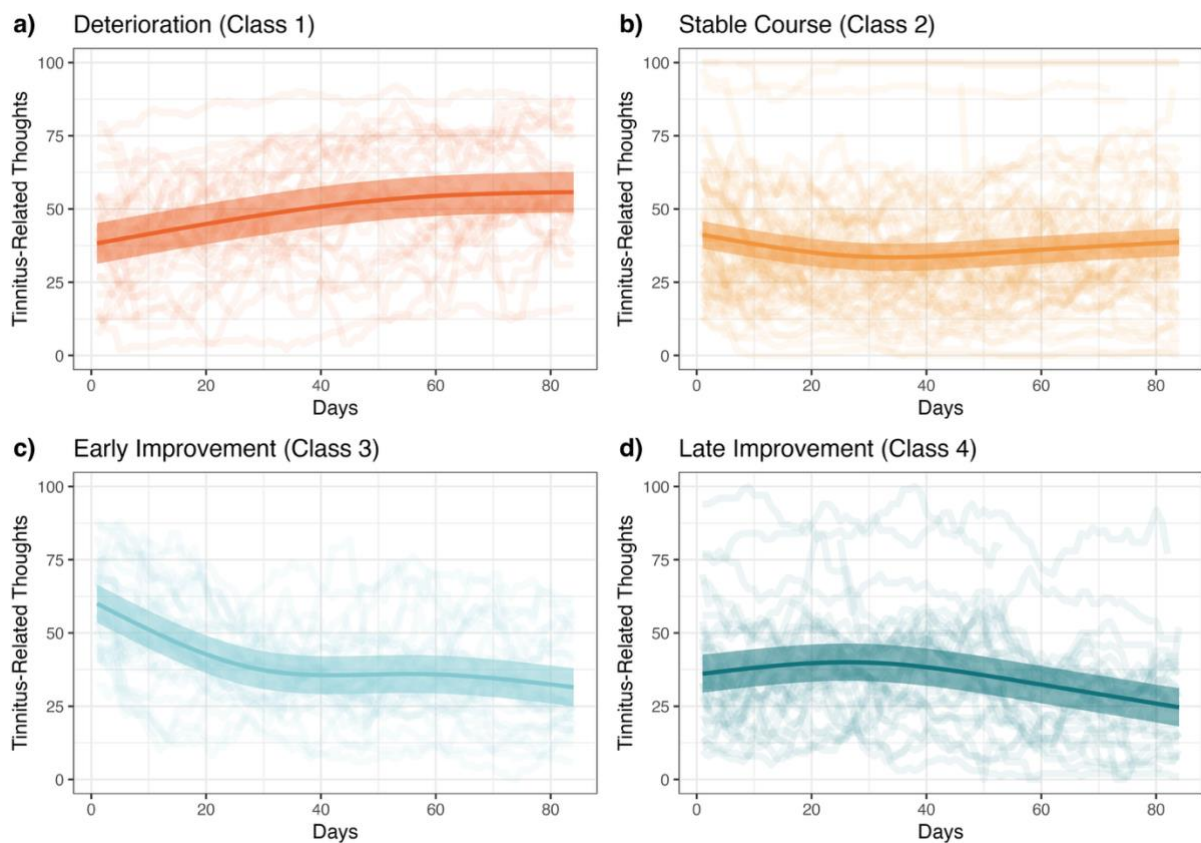
Notes. Trajectories depict estimated values, transparent bars are 95% CIs. Results were derived from growth mixture analysis estimating trajectories of tinnitus-related thoughts during treatment with varying number of latent classes.

Table S7. Model Selection Based on Entropy, Class Size, and PPM (Study III)

Model	Entropy	Class Size		Posterior Probability of Membership	
Three-class	0.91	Class 1	25.9%	Class 1	0.98
		Class 2	27.2%	Class 2	0.98
		Class 3	46.9%	Class 3	0.98
Four-class	0.97	Class 1	18.4%	Class 1	>0.99
		Class 2	40.1%	Class 2	0.99
		Class 3	20.4%	Class 3	>0.99
Five-class	0.96	Class 4	21.1%	Class 4	0.99
		Class 1	15.0%	Class 1	>0.99
		Class 2	10.9%	Class 2	0.99
		Class 3	35.4%	Class 3	0.98
		Class 4	19.7%	Class 4	0.99
		Class 5	19.1%	Class 5	0.99

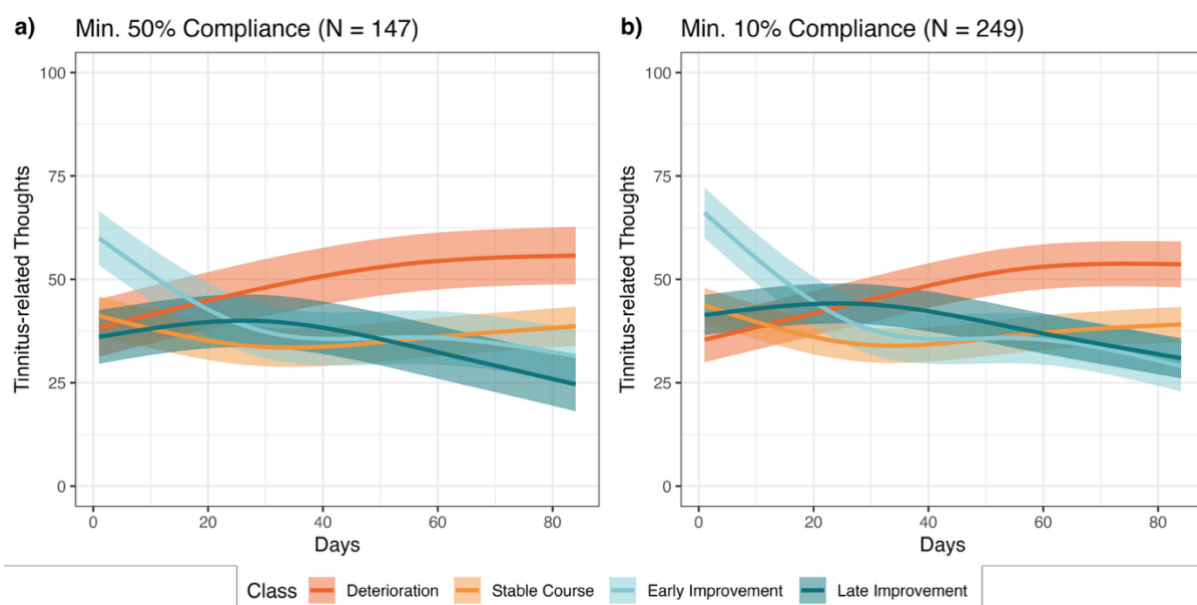
Notes. Models were derived from growth mixture analysis estimating trajectories of tinnitus-related thoughts during treatment with varying number of latent classes.

Figure S12. Observed Trajectories for Each Class (Study III)



Note. The bold trajectories depict estimated values, transparent bars are 95% CIs. The transparent lines depict the observed trajectories.

Figure S13. Latent Trajectories of the Sensitivity Analysis (Study III)



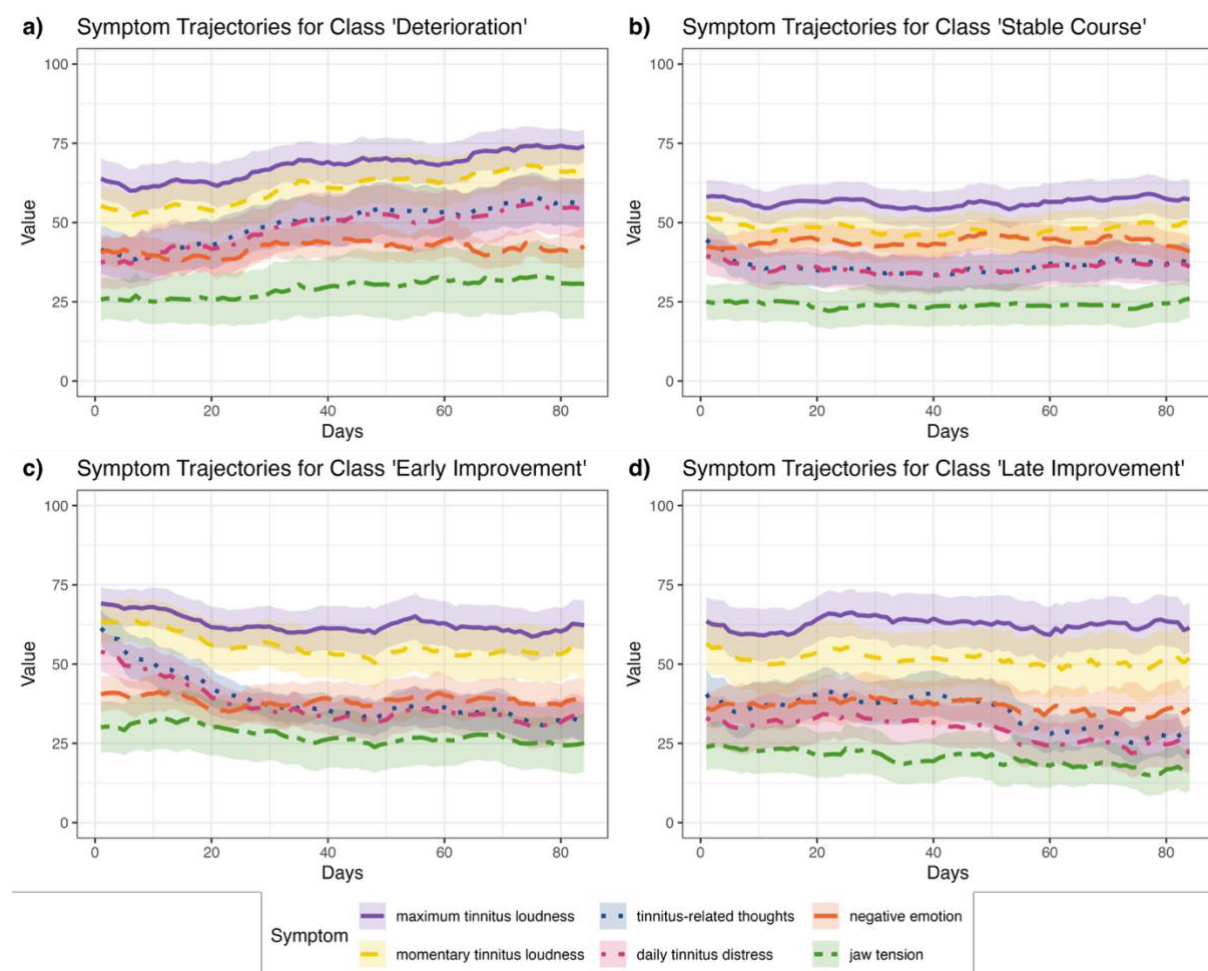
Notes. Comparison of results of the growth mixture analysis with two different cut-off values regarding compliance. The analysis in a) included participants with at least 50% compliance to the EMA protocol, the analysis in b) included participants with at least 10% compliance to the EMA protocol. Trajectories depict estimated values, transparent bars are 95% CIs.

Table S8. Comparison of Subsamples to the RCT Sample on Baseline Characteristics (Study III)

Baseline Characteristics	10% compliance	50% compliance	RCT sample	<i>p</i> value	<i>adjusted p</i> value
Gender					
female	110 (44.2%)	66 (44.9%)	195 (40.8%)	.540	1.00
male	139 (55.8%)	81 (55.1%)	283 (59.2%)	.540	1.00
Age	51.92 (12.66)	54.72 (11.86)	50.95 (12.42)	.010*	0.07
THI score	49.05 (20.06)	49.17 (20.29)	48.02 (19.73)	.730	1.00
PHQ9 score	7.62 (5.19)	7.58 (4.93)	7.35 (4.87)	.750	1.00
Hearingloss (PTA4)	19.28 (15.98)	20.61 (16.6)	19.21 (15.38)	.630	1.00
Tinnitus duration (in months)	124.33 (113.61)	136.18 (120.98)	119.48 (113.37)	.330	1.00
Extraversion	40.37 (7.25)	39.29 (6.84)	40.6 (7.53)	.170	1.00
Agreeableness	47.51 (6.18)	47.18 (6.17)	47.1 (5.97)	.700	1.00
Conscientiousness	47.06 (8.16)	46.94 (8.08)	47.19 (7.94)	.940	1.00
Neuroticism	34.4 (8.16)	34.19 (8.27)	34.85 (8.07)	.630	1.00
Openness	44.35 (7.6)	43.72 (7.48)	44.49 (7.63)	.580	1.00

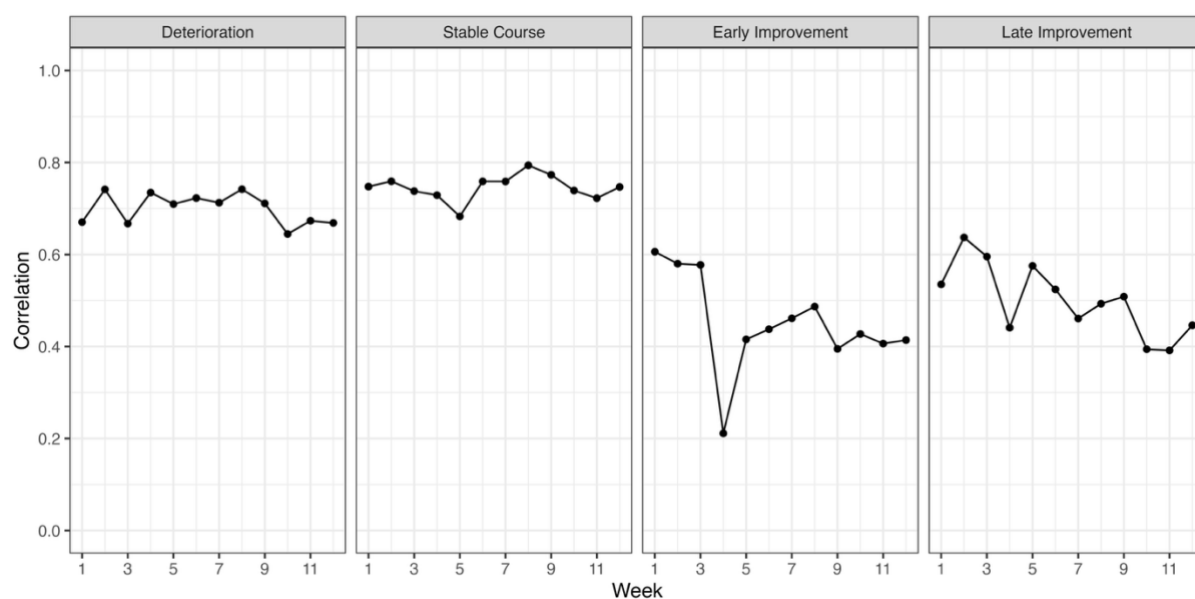
Note. THI score range: 0-100. PHQ-9 score range: 0-27. PTA4: Average hearing threshold at the frequencies 500 Hz, 1 kHz, 2 kHz and 4 kHz. BFI-2 dimensions score range: 12-60.

Figure S14. Trajectories of Remaining Symptoms Across Classes (Study III)



Notes. Trajectories are mean values of pre-processed individual time-series (seven-day moving average), transparent bars are 95% CIs. All symptoms were rated on a visual analogue scale from 0 — 100.

Figure S15. Trajectories of Remaining Symptoms Across Classes (Study III)



Note. Dots depict weekly average correlations between maximum tinnitus loudness and daily tinnitus distress (raw values) for each class.

Table S9. Dropouts and discontinuation during treatment by class (Study III)

Class	N	%
Deterioration	4	14.8
Stable Course	5	8.5
Early Improvement	1	3.3
Late Improvement	5	16.1

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used generative AI for copy-assisted editing. After using this tool, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgements

Mein besonderer Dank gilt meinem Doktorvater Winfried Schlee, der meine wissenschaftliche Arbeit mit großem Engagement und anhaltender Begeisterung begleitet hat. Seine fachliche Kompetenz, die anregenden Diskussionen auf Augenhöhe und sein ermutigendes Zusprechen in Phasen des Zweifelns haben diese Arbeit in besonderem Maße geprägt.

Ebenso danke ich Berthold Langguth, der mich sowohl in meiner wissenschaftlichen als auch in meiner klinischen Laufbahn bestärkt und unterstützt.

Mein herzlicher Dank gilt zudem Thomas Probst für die wertvolle Unterstützung aus der Ferne und den wissenschaftlich inspirierenden Monat, den ich in Salzburg verbringen durfte.

Für das kritische Mitdenken, die kollegiale Unterstützung und die angenehme Zusammenarbeit danke ich meinen (ehemaligen) Kolleginnen und Kollegen Stefan Schoisswohl, Jorge Simões und Nina Wunder. Ein besonderer Dank gilt Susanne Staudinger, die stets ein offenes Ohr hatte und dem ganzen wissenschaftlichen Team den Rücken freihält.

Von Herzen danke ich meinen Eltern, die mich zur Selbstständigkeit erzogen und niemals an mir gezweifelt haben, sowie meinen Schwestern, die stets mehr Vertrauen in mich hatten, als ich selbst. Mein tiefster Dank gilt meinem Partner, der die Jahre meiner Promotion an meiner Seite war und jede meiner Schwankungen mit Geduld und Zuversicht ausgeglichen hat.

Affidavit

Ich, Engelke, Milena geboren am 07.02.1997 in Augsburg, erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Insbesondere habe ich nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater*in oder andere Personen) in Anspruch genommen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Regensburg, den 16.10.2026

Ort, Datum



eigenhändige Unterschrift

des/der Promovenden*in