Systematic Review

The Natural History of Subjective Tinnitus in Adults: A Systematic Review and Meta-Analysis of No-Intervention Periods in Controlled Trials

John S. Phillips, FRCS (ORL-HNS) ©; Don J. McFerran, MA, FRCS; Deborah A. Hall, BSc, PhD; Derek J. Hoare, BSc, MRes, PhD

Objectives: Tinnitus is a prevalent condition, but little has been published regarding the natural history of the condition. One technique for evaluating the long-term progression of the disease is to examine what happens to participants in the no-intervention control arm of a clinical trial. The aim of this study was to examine no-intervention or waiting-list data reported in trials, in which participants on the active arm received any form of tinnitus intervention.

Data Sources: CINAHL, PsychINFO, EMBASE, ASSIA, PubMed, Web of Science, Science Direct, EBSCO Host, and Cochrane.

Methods: Inclusion criteria followed the PICOS principles: Participants, adults with tinnitus; Intervention, none; Control, any intervention for alleviating tinnitus; Outcomes, a measure assessing tinnitus symptoms using a multi-item patient-reported tinnitus questionnaire. Secondary outcome measures included multi-item patient-reported questionnaires of mood and health-related quality of life and measures that quantified change in tinnitus loudness; Study design, randomized controlled trials or observational studies utilizing a no-intervention or waiting-list control group. Data were extracted and standardized mean difference was calculated for each study to enable meta-analysis.

Results: The evidence strongly favored a statistically significant decrease in the impact of tinnitus over time, though there was significant heterogeneity and clinical significance cannot be interpreted. Outcome data regarding secondary measures did not demonstrate any clinically significant change.

Conclusions: Participants allocated to the no-intervention or waiting-list control arm of clinical trials for a tinnitus intervention show a small but significant improvement in self-reported measures of tinnitus with time; the clinical significance of this finding is unknown. There is, however, considerable variation across individuals. These findings support previous work and can cautiously be used when counseling patients.

Key Words: Tinnitus, natural history, outcomes, control, waiting list.

INTRODUCTION

Part of the counseling provided to tinnitus patients by practitioners involves reassurance that both the perceived loudness of the tinnitus sounds and the emotional symptoms of tinnitus generally improve with time. Although this may be true, data to support the validity of this statement and to quantify any improvement in symptoms have been poorly presented in the literature. There are a small number of longitudinal studies of tinnitus, which give some support to the suggestion that tinnitus impact lessens with time.1–3 However, participants in these studies could access healthcare services for their symptom, and it is therefore difficult to ascertain whether any change is natural improvement with
time or treatment effect. One technique used to study what happens to symptoms over time among people receiving no treatment is to examine the outcome of participants on a no-intervention or waiting-list control arm of clinical trials, and this methodology has a long pedigree of usage in the field of mental health. By amalgamating the control groups of multiple trials, meta-analysis of the outcome is viable. A limited study of what happens to patients with tinnitus while on a no-intervention and waiting-list control group has previously been undertaken, but this was restricted to studies that had incorporated cognitive behavior therapy as the active arm of the trial. Restricting participants to those willing to embrace psychological therapies for their tinnitus potentially produces a study population that is not representative of the wider tinnitus population. The aim of the current study was to expand that original work by looking at people with tinnitus who had been allocated to a no-intervention or waiting-list control group in the context of a trial evaluating any form of tinnitus therapy.

The following research questions were posited: 1) During a period of no intervention or waiting list, what changes occur in self-reported measure of tinnitus? 2) During a period of no intervention or waiting list, what changes occur in self-reported measures of tinnitus-related problems of mood and quality of life? 3) During a period of no-intervention waiting, what changes occur in perceived tinnitus loudness? The first research question was our primary outcome measure and our second and third research questions were our secondary outcome measures.

METHODS

Study Registration

Details of the proposed study eligibility criteria, information sources, search strategy, selection and data collection processes, as well as data synthesis methods were registered at PROSPERO, the international database of prospectively registered systematic reviews (PROSPERO 2013:CRD42013003334). Reporting of the review has been conducted using the criteria recommended by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Presentation of the meta-analysis complies with MOOSE Guidelines for Meta-Analyses and Systematic Reviews of Observational Studies.

Study Selection

In the protocol registered in PROSPERO, the condition of interest was referred to as “watchful waiting.” Because this term implies some degree of symptom monitoring, which was not necessarily evident in the records found, and because our study selection strategy did not necessarily seek to exclude study designs in which a group was not anticipating receiving an intervention, we refer instead to this group throughout as “no-intervention” or “waiting-list” control. Inclusion criteria were formed using the Participants, Intervention, Control, Outcomes, and Study designs (PICOS) strategy. These are: Participants, adults with tinnitus; Intervention, no-intervention or waiting-list control; Comparator: any intervention for tinnitus; Outcomes, primary, one or more tinnitus-specific measure using a multi-item patient-reported questionnaire; Study design, randomized controlled trial or observational study with a control group involving no intervention.

Studies that were not available in English were also excluded as we did not have the resources to translate them. Records that had not been through a peer-review process (grey literature) were excluded as a quality-control measure.

The search was not explicitly time limited, but the first multi-item patient-reported tinnitus-related questionnaire was published in 1988. Hence, no clinical trials meeting our inclusion criteria would have been published prior to this date. For the purposes of the review, adult was defined as aged 16 years or older.

Appropriate Outcome Measures

Eligible studies were those reporting at least one patient-reported outcome relating to tinnitus, measured using a multi-item patient-reported tinnitus-specific questionnaire with scores that were reported both before and after the time period corresponding to the intervention for the active comparator group. Examples of acceptable measurement instruments are shown in Supporting Table SI in the online version of this article. This is not an exhaustive list, and, if encountered, other tools were considered. Outcomes that were considered as a secondary question in this review were those multi-item patient-reported questionnaires of mood and quality of life, and tools for estimating a change in tinnitus percept, namely loudness, with scores that were reported both before and after the time period corresponding to the intervention for the comparator group. Such assessments were not prerequisites for study inclusion, but where such information was available it was extracted and analyzed. In a change to the study design as registered in PROSPERO, we did not investigate the change in audiological or physiological outcome measures as secondary questions.

Appropriate Study Design

Eligible study designs were randomized controlled trials in which adult participants were allocated to a no-intervention control group receiving no support. Observational studies in which there was a no-intervention group were also eligible. Cross-over designs were included if a no-intervention period preceded an active intervention comparator and data from the preintervention period could be separately extracted.

Search Strategy

A systematic search of the literature was conducted by one of the authors (D.J.H.) to identify relevant articles from eight literature search platforms: CINAHL, PsychINFO, Embase, ASSIA, PubMed, Web of Science, Science Direct, and EBSCO Host. For each database, the search was run using the Boolean search term: tinnitus AND waiting OR wait * OR waiting-list OR watchful OR observation. For interest, a sample search strategy (generated by PubMed) in executing the search is given in Supporting Table SII in the online version of this article.

In addition, hand-searching of the reference lists of all articles returned from the search was undertaken, and articles published by shortlisted authors were screened to identify any relevant articles that may not have been returned by the initial database searches. Cochrane and other relevant systematic reviews were searched. In October 2015, hand searches were conducted of articles published in issues since April 2013 of the prespecified journals (Acta Otologynlogica, Ear and Hearing, Hearing Research, Journal of Psychosomatic Medicine, Psychosomatic Research, International Journal of Audiology, Laryngoscope 128: January 2018 Phillips et al.: Natural History of Tinnitus

218
Data Management

All identified records were saved into a Microsoft Excel (Microsoft Corp., Redmond, WA) master file where records were tracked through the screening and data collection process by a unique study identification number. A simple system of record annotation was implemented to capture reasons for exclusion. Two authors (J.S.P., D.A.H.) independently assessed the search results to identify studies for inclusion in the review and extracted the relevant data. Any discrepancies in study selection or data extraction were resolved in discussion with a third author (D.A.H. or D.J.H.). One of the authors (J.S.P.) was the data guarantor.

Data Extraction

Data extracted included study design, participants (demographics, baseline characteristics), context of waiting (waiting list for crossover or no intervention), comparator, outcomes measured, study findings, and conclusions. A data extraction form was developed and piloted for the purpose. Where data were missing or unclearly reported, an attempt was made to contact the relevant corresponding author of the study; the most common problem was that the results had been presented graphically and numerical data for the meta-analysis could not be extracted. Supporting Table SIII in the online version of this article provides a summary of 18 study records for which we sought clarification or additional information. Of those, only three did not reply; six did reply but were unable to provide the data requested.

Risk of Bias (Quality) Assessment

Risk of bias assessment was guided by Higgins et al.13 and was conducted by three authors (J.S.P., D.A.H., D.J.H.) on those study records included in the meta-analysis. The following terminology was specified: 1) Selection bias refers to how participants were allocated to the intervention arms of the trial and was assessed according to two criteria, namely sequence generation for the randomization process and allocation concealment to ensure that the schedule of random assignments prevented advance knowledge about the forthcoming allocations. 2) Attribution bias refers to how participants withdrew from any trial and was assessed by identifying incomplete outcome data. 3) Detection bias refers to how the outcomes were determined and was assessed according to the blinding of participants and outcome assessors assessing patient- or clinician-reported questionnaires, respectively. In the protocol registered in PROSPERO, these three categories of risk of bias were described as 1) study design, 2) compliance and drop out, and 3) blinding. Sample size was not evaluated in this section because this is a marker of quality, not risk of bias.14 Similarly, external validity of the study sample (i.e., specialist subgroups such as occupational setting, tertiary clinic, severe tinnitus only) was not formally evaluated.

Measures of Effect

From each study a standardized mean difference (SMD) was calculated for every included score obtained on all tinnitus questionnaires. SMD was calculated for each post-baseline time point and was defined as the difference between the group mean questionnaire score at baseline and after n weeks of no-intervention waiting, divided by the pooled standard deviation. A positive SMD indicated an improvement over time. This difference was then converted to Hedges’ g15 a commonly used measure of effect that controls for the bias in effect size that might be introduced by studies with small participant sample size. The test-retest correlation between the repeated time points was set to 90% for all questionnaires. Where multiple questionnaires were used at the same time point, a mean effect size was calculated by averaging the individual effect sizes.

Meta-analysis

Mean effect sizes across studies were calculated using Comprehensive Meta-Analysis (version 2.2.048, Biostat, Englewood, NJ). For the primary synthesis, the latest time point in each study was selected, and a random effects model was run. A random effects model assumes that the true effect may vary from study to study; here it was assumed that changes in the impact of tinnitus over time are not likely a constant effect but may be influenced by study factors such as age of participants, duration of tinnitus, education level, or general health. Sensitivity analyses were conducted pooling effect sizes per time from baseline (≤6 weeks, ≤12 weeks, ≤6 months). For all meta-analyses, it was reasonably assumed that the multi-item questionnaires included showed sufficient convergent validity to be pooled; tinnitus questionnaires are generally demonstrated to measure the same underlying construct of the everyday impact of tinnitus.

RESULTS

Literature searches were conducted in December 2013 and updated in October 2015. The initial database searches yielded 902 records. A further 23 eligible records were identified through hand searches (see Fig. 1 for PRISMA flowchart). Following screening, 50 records were retrieved for full-text review. Twenty-five studies were included16–40 (Table I), and 21 studies were suitable for inclusion in the meta-analyses.16–18,21–26,28,29,31–40 Among these, one record reported two independent control groups that were retained as independent groups.18 Ross et al. (2007) also used multiple control groups depending on tinnitus duration, and this enabled two independent control groups to be used for the meta-analysis.35 Thus, the meta-analyses examined up to 23 no-intervention or waiting-list control groups from 21 study records.

Missing Data

Data queries were satisfactorily answered for eight study records and partly answered for one other (see Supporting Table SIII in the online version of this article).

Data Synthesis

The period spent on the no-intervention or waiting-list period varied from 1 to 52 weeks, with an average of 12 weeks. Information about the individual percentage and effect size of change in tinnitus severity, as measured by tinnitus questionnaire score is provided in Table II. Two studies (Fackrell et al., 2016; Krick et al., 2015) were excluded from the meta-analysis because the

Laryngoscope 128: January 2018

Phillips et al.: Natural History of Tinnitus

219
interval between assessments for most or all patients was as little as 7 days. Caffier et al. (2006) was excluded, as numerical data were not sufficiently available. Jakes et al. (1992) was excluded from the meta-analysis, as their tinnitus outcome questionnaire was the Tinnitus Effects Questionnaire, which does not yield a global score.

Across the remaining studies, over the longest period reported, there was a small decrease in global tinnitus of 2.3%, indicating a trend for improvement over time. How clinically meaningful that is cannot be interpreted; although it was assumed that tinnitus questionnaires measure the same construct of the everyday impact of tinnitus, clinically meaningful change scores on those questionnaires differ. Strikingly, no study demonstrated statistically significant worsening of tinnitus over time. There was, however, considerable heterogeneity across studies. Reports of changes in depression, anxiety, quality of life, and tinnitus loudness were few and not significant.

Risk of Bias Assessment

A summary of the risk of bias of the 21 study records that were included in the meta-analysis is shown in Table III. Low risk of bias was achieved on 51% of occasions. Six studies had a high or unclear risk of bias on two of the criteria, whereas one study had a high or unclear risk of bias on all three criteria. Detection bias was the most poorly reported. Support for judgement concerning selection bias, attrition bias, and detection bias is provided in Supporting Tables IV–VI in the online version of this article, respectively.
| Record | No-Intervention Group, Mean Age, yr, Where Reported | No-Intervention Group, Gender | Comparison Intervention(s) | Tinnitus Questionnaire(s) | Anxiety, Depression, Quality of Life Questionnaire(s) | Other Measures | Study Design |
|--------|-----------------------------------------------|-----------------------------|----------------------------|--------------------------|-----------------------------------------------|----------------|-------------|
| Andersson 2002 | 47.2 | 48%F | Internet-based CBT | TRQ | HADS, ASI | VAS loudness | RCT |
| Andersson 2005 | — | — | CBT | TRQ | HADS, ASI | VAS loudness | RCT |
| Biesinger 2010 | 39.3 | 12F, 8M | Qigong | TBF-12 | — | VAS annoyance | RCT |
| Biesinger 2010 | 39.9 | 9F, 10M | Qigong | TBF-12 | — | VAS annoyance | RCT |
| Caffer 2006 | — | — | Tinnitus coping therapy ± sound therapy | TQ (German) | — | VAS loudness, annoyance, awareness | RCT |
| Fackrell 2016 | — | — | No intervention | TR, THI, THQ | BDI, BAI, WHOQOL-BREF | VAS loudness | Questionnaire validation |
| Henry 1996 | — | — | Cognitive education | TEQ, TRQ, THQ, TCQ | BDI | — | RCT |
| Henry 1998 | — | — | Attention control and imagery training, cognitive restructuring | TRQ, THQ, TEQ, TCQ, TCSQ | BDI, ATQ | — | RCT |
| Henry 2002 | 62.0 | 3F, 88M | Educational counseling and support | TSI | — | — | Questionnaire validation |
| Henry 2016 | 59.2 | 13F, 77M | Telephone tinnitus education | TFI, THI | — | — | RCT |
| Herraiz 2005 | — | — | Tinnitus retraining therapy | THI | — | VAS intensity annoyance | Nonrandomized clinical assay |
| Herraiz 2010 | 47 | 14F, 12M | Auditory discrimination training | THI | — | VAS intensity | RCT |
| Jakes 1992 | 54.2 | 6M, 8F | Group cognitive therapy | TEQ | CCEI | IWDA | RCT |
| Kaldo 2007 | 48.5 | 18F, 20M | Self-help book | THI, TRQ | HADS | VAS loudness, distress, stress, ISI | RCT |
| Kreuzer 2012 | 51.7 | 10F, 8M | Mindfulness and body psychotherapy | THI, TQ | BDI | VAS loudness, discomfort, annoyance, distractibility, unpleasantness | RCT |
| Krick 2015 | 42.6 | 9F, 13M | Music therapy | TQ | — | — | Nonrandomized controlled study |
| Kroner-Herwig 2003 | 47.3 | 10F, 10M | Cognitive behavioral group training | TQ, TDI | ADS | COPE | RCT |
| Malouf 2010 | 57.8 | 31F, 47M | Bibliotherapy | TSS, TRQ | GHQ-12 | — | RCT |
| Neri 2009 | 44.2 | 20F, 14M | Melatonin and sulodexide | THI | — | — | RCT |
| Rief 2005 | 48.0 | 12F, 8M | Psychophysiolgic treatment | TQ | GSI | — | RCT |
| Ross 2007 | >16 | — | Ericksonian hypnosis | TQ | SF-36 | — | Nonrandomized longitudinal controlled study |
| Sadlier 2008 | 54.3 | 6F, 5M | Mindfulness meditation CBT | TQ | HADS | VAS problem level | Nonrandomized controlled study |
Effects Over Time on Global Tinnitus

Twenty-three study groups (788 participants) in 21 study records reported changes in tinnitus over time. Effect sizes (Hedges’ $g$) for the maximum interval within studies ranged from $-0.17$ to 0.55. The primary meta-analysis pooled data across studies using the longest timeframe reported in each study record, irrespective of the absolute length of time (23 study groups, $M = 12$ weeks, range = 4–52 weeks). There was significant heterogeneity across studies ($Q(df = 22) = 112.97, P < .001, I^2 = 80.53$). In a random effects model, the mean effect size was statistically significant in favor of tinnitus improving (Hedges’ $g = 0.122$, 95% confidence interval [CI]: 0.055 to 0.188, $P < .001$) (Fig. 2).

For 14 study groups, tinnitus questionnaire data at up to 2 months from baseline were available ($M = 6.6$ weeks, range = 2–8 weeks). There was significant heterogeneity ($Q(df = 13) = 48.7, P < .001, I^2 = 73.29$). In a random effects model, the mean effect size was statistically significant toward improvement in tinnitus (Hedges’ $g = 0.097$, 95% CI: 0.019 to 0.176, $P = .015$). The pattern was the same for 10 study groups reporting questionnaire data up to 4 months ($M = 10.6$ weeks, range = 9–12 weeks, $n = 238$). There was significant heterogeneity ($Q(df = 8) = 51.6, P < .001, I^2 = 82.57$). In a random effects model, the mean effect size was still significant (Hedges’ $g = .154$, 95% CI: 0.027 to 0.281, $P = .018$). Longer-term effects (measured in seven study groups) were not significant, however ($M = 29$ weeks, range = 17–52, $n = 256$) (Hedges’ $g = 0.112$, 95% CI: $-0.013$ to 0.236, $P = .079$).

Effects Over Time on Depression

Eight studies (301 participants) reported changes in depression questionnaire scores, using the Hospital Anxiety Depression Scale-Depression$^{41}$ or Beck Depression Index (BDI)$^{32}$ questionnaires over intervals ranging 6 to 26 weeks ($M = 16.2$). Henry et al. (1998)$^{22}$ measured the BDI score at three time intervals (baseline, 26 weeks, and 52 weeks later). Hedges’ $g$ across the eight studies ranged from 0.469 to 0.182, with one study favoring a worsening and two studies favoring an improvement in scores over time (Fig. 3). The pooled effect size across all studies (using the 26-week measure from Henry et al., 1996$^{21}$) was positive but not significant (Hedges’ $g = 0.006$, 95% CI: $-0.045$ to 0.057, $P = .828$) indicating no significant change in depression over time.

Effects Over Time on Generalized Anxiety

Five studies (161 participants) reported changes in anxiety questionnaire scores, using the Hospital Anxiety Depression Scale-Anxiety (HADS-A) (Zigmond and Snaith 1983) over intervals ranging from 6 to 12 weeks ($M = 8$). Hedges’ $g$ across studies ranged from 0.089 to 0.206, with one study favoring an improvement in scores over time (Fig. 4). The pooled effect size across all studies was positive but not significant (Hedges’ $g = 0.058$, 95% CI: $-0.012$ to 0.127, $P = .104$), indicating no significant change in anxiety over time. Andersson et al.
### TABLE II.
Percentage and Effect Size of Change in Tinnitus Questionnaire Score.

| Study         | Sample Size | Interval (Weeks) | % Change | TRQ | THI | TQ | THQ | TBF-12 | TSI | TDI | TFI | TCQ     | SMD   | SE of SMD |
|---------------|-------------|------------------|----------|-----|-----|----|-----|--------|-----|-----|-----|---------|-------|-----------|
| Andersson 2002 | 48          | 6                | −7.3     | 0.1190 | 0.0648 |
| Andersson 2005 | 11          | 6                | 10.5     | −0.1675 | 0.1358 |
| Biesinger 2010  | 20          | 5                | 2.3      | 0.0689  | 0.1001 |
| Biesinger 2010  | 20          | 9                | −0.8     | 0.0230  | 0.1000 |
| Biesinger 2010  | 20          | 17               | −3.8     | 0.1075  | 0.1003 |
| Biesinger 2010  | 19          | 5                | 0        | 0.0000  | 0.1026 |
| Biesinger 2010  | 19          | 9                | −2.1     | 0.0625  | 0.1027 |
| Fackrell 2016   | 44          | 2                | 0.6      | −0.0152 | 0.0674 |
| Henry 1996      | 20          | 6                | −1.8     | 0.0395  | 0.1000 |
| Henry 1996      | 20          | 6                | 1.6      | −0.0136 | 0.1000 |
| Henry 1998      | 12          | 8                | −13.5    | 0.2690  | 0.1314 |
| Henry 1998      | 12          | 8                | −5.2     | 0.3421  | 0.1328 |
| Henry 1998      | 12          | 8                | −18.4    | 0.4956  | 0.1368 |
| Henry 2007      | 72          | 26               | 0        | 0.0000  | 0.0523 |
| Henry 2007      | 75          | 52               | −2.3     | 0.0575  | 0.0517 |
| Henry 2016      | 82          | 26               | NR       | 0.0146  | 0.0471 |
| Henry 2015      | 82          | 26               | NR       | 0.0553  | 0.0472 |
| Herraiz 2005    | 21          | 6                | 6.5      | −0.1005 | 0.0978 |
| Herraiz 2010    | 26          | 4                | 5        | −0.0730 | 0.0878 |
| Kaldo 2007      | 36          | 6                | −0.3     | 0.0044  | 0.0745 |
| Kaldo 2007      | 36          | 6                | −4.9     | 0.1271  | 0.0748 |
| Kreuzer 2012    | 18          | 7                | −7       | 0.1787  | 0.1034 |
| Kreuzer 2012    | 18          | 9                | −12.5    | 0.3210  | 0.1052 |
| Kreuzer 2012    | 18          | 24               | −14      | 0.3444  | 0.1085 |
| Kreuzer 2012    | 18          | 7                | −4.9     | 0.1227  | 0.1058 |
| Kreuzer 2012    | 18          | 9                | −12.1    | 0.3033  | 0.1078 |
| Kreuzer 2012    | 18          | 24               | −12.3    | 0.2992  | 0.1077 |
| Krick 2015      | 22          | 1                | −5       | 0.1071  | 0.0956 |
| Kroner–Herwig 2003 | 20  | NR               | −7       | −0.0985 | 0.1002 |
| Kroner–Herwig 2003 | 20  | NR               | 6.7      | 0.1225  | 0.1004 |
| Malouff 2010    | 70          | 8                | −7.6     | 0.1605  | 0.0538 |
| Neri 2009      | 34          | 12               | 1.2      | −0.0228 | 0.0767 |
| Neri 2009      | 34          | 18               | 1        | −0.0183 | 0.0767 |
| Rief 2005      | 20          | 2                | −21.5    | 0.5208  | 0.1066 |
| Rief 2005      | 20          | 6                | −22.5    | 0.4944  | 0.1059 |
| Ross 2007      | 100         | 8                | −5.2     | 0.1422  | 0.0449 |
| Ross 2007      | 283         | 8                | −7.1     | 0.2043  | 0.0269 |
| Ross 2007      | 225         | 8                | −7.8     | 0.2269  | 0.0302 |
| Sadlier 2008   | 11          | 12               | 2.8      | −0.0629 | 0.0660 |
| Seydel 2010    | 45          | 12               | −0.5     | 0.1160  | 0.0700 |
| Sönmez 2013    | 15          | 26               | 17.4     | 0.5480  | 0.1230 |
| Weise 2008     | 59          | 12               | −9.1     | 0.4464  | 0.0611 |
| Westin 2011    | 21          | 10               | −2       | 0.0507  | 0.0977 |

NR = baseline not reported; SE = standard error; SMD = standardized mean difference; TBF-12 = Tinnitus Beeinträchtigung Fragebogen (i.e., Tinnitus Impairment Questionnaire); TCQ = Tinnitus Cognitions Questionnaire; THI = Tinnitus Handicap Inventory; THQ = Tinnitus Handicap Questionnaire; TQ = Tinnitus Questionnaire; TRQ = Tinnitus Reaction Questionnaire; TSI = Tinnitus Severity Index.

Negative percentage change indicates decreased questionnaire score (tinnitus improves); positive percentage change indicates increased questionnaire score (tinnitus worse).
(2002) additionally measured “fear of anxiety-related somatic sensation” using the Anxiety Sensitivity Index, noting a slight improvement over time; the mean score was reduced from 19.1 (standard deviation [SD] = 12.7) at baseline to 17.8 (SD = 12.1) at 6 weeks. In contrast, Andersson et al. (2005) report an increase in Anxiety Sensitivity Index score in their waiting-list control group after about 6 weeks; the mean score increased from 18.9 (±10.0) to 26.3 (±10.5).

**Effects Over Time on Quality of Life**

Only two studies reported on changes in quality of life over time, with similar results. Rief et al. (2005) reported on change on the Questions of Life Satisfaction questionnaire; after 8 weeks their participants reported a 15.5-point (on a 100-point scale) increase in questionnaire scores, suggesting improvement in subjective satisfaction with quality of life, but at the 6-month follow-up there was no difference from baseline. Westin et al. (2011) used the Quality of Life Inventory and measured change at 10 weeks, finding a 0.2 increase in score on a three-point scale, which was not significant.

**Effects Over Time on Tinnitus Loudness**

In six studies, tinnitus loudness was measured using a visual analogue scale. However, Jakes et al. (1992) reported abandoning the measure during the study for several reasons including poor compliance, and Andersson et al. (2005) did not report numerical values. Of the four remaining, three used a 0

| Study Name | Low | Low | Unclear | Detection |
|------------|-----|-----|---------|-----------|
| Andersson 2002 | Low | Unclear | High |
| Andersson 2005 | Low | Low | High |
| Biesinger 2010 | Low | Low | Unclear |
| Henry 1996 | Low | Low | High |
| Henry 1999 | Low | High | High |
| Henry 2007 | Low | Unclear | Unclear |
| Henry 2016 | Low | Low | Unclear |
| Herrera 2005 | High | Low | Unclear |
| Herrera 2010 | High | Low | Unclear |
| Kaldo 2007 | Low | Low | Unclear |
| Kreuzer 2012 | Low | Low | Unclear |
| Kroner-Herwig 2003 | Low | Low | Unclear |
| Malouff 2010 | Low | Low | Unclear |
| Neri 2009 | Low | High | High |
| Rief 2005 | Low | High | High |
| Ross 2007 | High | Low | High |
| Sadler 2008 | High | Unclear | Unclear |
| Seydel 2007 | High | Low | Unclear |
| Sonmez 2013 | Unclear | Unclear | Low |
| Weise 2008 | Low | Low | Unclear |
| Westin 2011 | Low | Low | Low |

Selection bias includes sequence generation for the randomization process and allocation concealment. Bias was judged for those study records included in the meta-analysis, and a full description of the review authors’ judgements about risk of bias for each included study is given in the Supporting Information, Tables SIV–SVI.
to 10 scale and reported a 0.8-point decrease, no change, and a 0.1-point increase in scores, respectively, after 6 to 8 weeks watchful waiting. One study used a 0 to 4 scale and reported a reduction of <0.1 after 4 weeks.

Although single-item measures of tinnitus show good correlation with each other, they do not measure meaningful tinnitus-related constructs, so these data were not subjected to meta-analysis.

**DISCUSSION**

This systematic review with a meta-analysis presents the most inclusive evaluation of the natural time course of tinnitus under controlled experimental conditions to date. The random effects meta-analysis gives a reliable overall summary of findings, because the analysis accounts for heterogeneity and is weighted by sample size. This revealed a small but significant improvement in global tinnitus severity up to 4 months, but studies with longer assessment periods did not reveal any change. This finding may reflect a lack of statistical power for this subgroup analysis or an insensitivity of tinnitus questionnaire measures over longer periods. Even for the 2- and 4-month analyses, it must be cautioned that we cannot ascertain with certainty whether the small statistically significant improvement is equivalent to a clinically meaningful improvement that is noticeable to people with tinnitus. Clinical interpretation of the findings by anchoring numerical values against patient-reported experience is under-reported to date.

In contrast to the small improvements in global tinnitus severity, our meta-analyses did not reveal statistically significant improvement in measures of mood. This finding contradicts that of a study by Posternak and Miller (2001), which looked at mental health conditions in isolation and found improvement while on waiting-list control groups. It is possible that the null findings in the current study simply represent the relatively low number of tinnitus studies that had incorporated a measure of depression or generalized anxiety.

Although this systematic review accepted studies testing any form of tinnitus intervention, the meta-analysis was biased toward psychological interventions, with 11 out of the 21 included studies testing a psychological management modality. The current study, although skewed toward psychological treatment trials, adds to previous work by Hesser et al. (2011), because

**Fig. 3.** Meta-analysis of change in self-reported depression shows no significant change over time. Black square = effect size (Hedges’ g) in that study. Black diamond = pooled effect size. The relative sample size and hence relative influence of individual studies on the pooled effect size is indicated by the size of the black square. CI = confidence interval.

**Fig. 4.** Meta-analysis of change in self-reported anxiety shows no significant change over time. Black square = effect size (Hedges’ g) in that study. Black diamond = pooled effect size. The relative sample size and hence relative influence of individual studies on the pooled effect size is indicated by the size of the black square. CI = confidence interval.
it assessed a much broader range of tinnitus experiences than this previous work. Our findings incorporated those participants enrolled into a range of tinnitus intervention studies, namely tinnitus retraining therapy, education, auditory discrimination training, self-help using books, drug treatment, and Qigong (a combination of body posture, breathing control, and meditation developed in China). We believe that this inadvertent bias toward psychological interventions is in large part a reflection of the type of control group favored by trial designs assessing different types of tinnitus study. Pharmacological intervention studies will typically use a placebo medication as control, whereas studies assessing a device treatment, such as repetitive transcranial magnetic stimulation or laser therapy, will generally employ a sham treatment as control. For psychological therapies, such as cognitive behavior therapy, acceptance and commitment therapy, or mindfulness meditation, a placebo or sham psychological therapy control is unethical for the trial design, and so those trials are therefore much more likely to use a no-intervention or waiting-list control. Moreover, psychological therapies present a routine therapeutic option for people with bothersome tinnitus, often with a natural waiting list for an initial appointment, and so a no-intervention control is often a straightforward pragmatic option. One limitation for our interpretation of the findings is that it is not clear whether tinnitus patients consenting to participate in a psychological treatment trial are representative of tinnitus patients in general, and more specifically whether they are equivalent to those consenting to join a drug trial or medical device study. Two studies drew their participants directly from US military veterans, hence participants were more likely to be male, have been exposed to a greater than average risk of noise-induced hearing loss, and to the psychological stress associated with military service. These limitations are mitigated to some degree by the meta-analysis, which pooled findings from a wide range of studies. One further potential limitation of note is the exclusion of studies not available in English (because of limited resources), and studies that appear only in the grey literature. Although excluding grey literature may have introduced a publication bias, including grey literature could in itself introduce bias if the included sample of unpublished studies was not representative of all unpublished studies. It would be interesting to explore this issue in further analyses.

CONCLUSION

Participants enrolled into clinical trials assessing tinnitus interventions generally demonstrate a small but statistically significant improvement in self-reported global tinnitus severity scores over time, despite receiving no intervention. This finding provides statistical evidence that tinnitus generally improves over time, albeit the effect is highly variable across individuals, and how clinically meaningful the effect is cannot be interpreted at a general level. This evidence can therefore cautiously be used when counseling patients.

BIBLIOGRAPHY

1. Andersson G, Vreblad P, Larsen HC, Lyttkens L. Longitudinal follow-up of tinnitus complaints. Arch Otolaryngol Head Neck Surg 2001;127:175–179.
2. Nondahl DM, Cruickshanks KJ, Wiley TL, Klein R, Klein BE, Tweed TS. Prevalence and 5-year incidence of tinnitus among older adults: the epidemiology of hearing loss study. J Am Acad Audiol 2002;13:323–331.
3. Nondahl DM, Cruickshanks KJ, Wiley TL, et al. The 10-year incidence of tinnitus among older adults. J Int Audiol 2010;49:580–585.
4. Posternak MA, Miller J. Underestimation of true prevalence of depression: a meta-analysis of outcomes from studies using wait-list control groups. J Affect Disord 2001;66:159–146.
5. Young C. What happens when people wait for treatment? Assessing the clinical significance of the changes observed over the waiting period for clients referred to a primary care psychology service. Prim Care Mental Health 2006;4:113–119.
6. Steiner C, Stadler K, Stark R, Leichsenring F. The Effects of waiting for treatment: a meta-analysis of wait-list control groups in randomized controlled trials for social anxiety disorder [published online July 22, 2016]. Clin Psychol Psychother doi: 10.1002/cpp.2032. Hesser H, Weise C, Rief W, Andersson G. The effect of waiting: a meta-analysis of wait-list control groups in trials for tinnitus distress. J Psychosom Res 2013;70:376–398.
7. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2706.
8. Group DFP, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2088–2092.
9. Centre for Reviews and Dissemination. Systematic Reviews. CRD’s Guidance for Undertaking Reviews in Health Care. York, United Kingdom: York Publishing Services; 2008.
10. Hallam RS, Jakes SC, Hinchcliffe R. Cognitive variables in tinnitus annoyance. Br J Clin Psychol 1988;27:213–222.
11. Hall DA, Haider H, Szczepek A, et al. Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults. Trials 2016;17:270.
12. Higgins JP, Altman DG, Gøtzsche PC; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.
13. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook of Systematic Reviews of Interventions. West Sussex, United Kingdom: John Wiley; 2008:187–241.
14. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. Psychol Methods 1998;3:486–504.
15. Andersson G, Stromgren T, Strom L, Lyttkens L. Randomized controlled trial of internet-based cognitive behavior therapy for distress associated with tinnitus. Psychosom Med 2002;64:810–816.
16. Andersson G, Persson D, Persson J, et al. Tinnitus re-training therapy in the elderly: a controlled trial of cognitive behavior therapy. Int J Audiol 2005;44:671–675.
17. Biesinger E, Kipman U, Schatz S, Leichsenring F. The Effects of waiting for treatment: a meta-analysis of wait-list control groups in randomized controlled trials for social anxiety disorder. J Affect Disord 2016;202:270–277.
18. Hall DA, Haider H, Szczepek A, et al. Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults. Trials 2016;17:270.
19. Higgins JP, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook of Systematic Reviews of Interventions. West Sussex, United Kingdom: John Wiley; 2008:187–241.
20. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. Psychol Methods 1998;3:486–504.
21. Andersson G, Stromgren T, Strom L, Lyttkens L. Randomized controlled trial of internet-based cognitive behavior therapy for distress associated with tinnitus. Psychosom Med 2002;64:810–816.
22. Andersson G, Persson D, Persson J, et al. Tinnitus re-training therapy in the elderly: a controlled trial of cognitive behavior therapy. Int J Audiol 2005;44:671–675.
23. Biesinger E, Kipman U, Schatz S, Leichsenring F. The Effects of waiting for treatment: a meta-analysis of wait-list control groups in randomized controlled trials for social anxiety disorder. J Affect Disord 2016;202:270–277.
24. Hall DA, Haider H, Szczepek A, et al. Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults. Trials 2016;17:270.
25. Higgins JP, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook of Systematic Reviews of Interventions. West Sussex, United Kingdom: John Wiley; 2008:187–241.
26. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. Psychol Methods 1998;3:486–504.
randomized controlled pilot study. **BMC Complement Altern Med** 2012;12:235–242.

30. Krick CM, Grapp M, Daneshvar-Talebi J, Reith W, Plinkert PK, Bolay HV. Cortical reorganization in recent-onset tinnitus patients by the Heidelberg Model of Music Therapy. **Front Neurosci** 2015;9:49.

31. Kroner-Herwig B, Frenzel A, Fritsche G, Schilkowsky G, Esser G. The management of chronic tinnitus—comparison of an outpatient cognitive-behavioral group training to minimal-contact interventions. **J Psychosom Res** 2003;54:381–389.

32. Maleff J, Noble W, Schutte NS, Bhullar N. The effectiveness of bibliotherapy in alleviating tinnitus-related distress. **J Psychosom Res** 2010;68:245–251.

33. Neri G, De Stefano A, Baffa C et al. Treatment of central and sensorineural tinnitus with orally administered melatonin and sulodexide: personal experience from a randomized controlled study. **Acta Otorhinolaryngol Ital** 2009;29:86–91.

34. Ross UH, Lange O, Unterrainer J, Laszig R. Ericksonian hypnosis in tinnitus therapy: effects of a 28-day inpatient multimodal treatment concept measured by Tinnitus-Questionnaire and Health Survey SF-36. **Eur Arch Otorhinolaryngol** 2007;264:483–488.

35. Sadlier M, Stephens SDG, Kennedy V. Tinnitus rehabilitation: a mindfulness meditation cognitive behavioural therapy approach. **J Laryngol Otol** 2008;122:31–37.

36. Seydel C, Haupt H, Szczepak AJ, Klapp BP, Mazzucchi B. Long-term improvement in tinnitus after modified tinnitus retraining therapy enhanced by a variety of psychological approaches. **Audiol Neurotol** 2010;15:69–80.

37. Simm M, Kulahl I, Vural A, Sahin MI, Aydin M. The evaluation of ozone and betahistine in the treatment of tinnitus. **Eur Arch Otorhinolaryngol** 2012;270:1999–2006.

38. Weise C, Heinecke K, Rief W. Biofeedback-based behavioral treatment for chronic tinnitus: results of a randomized controlled trial. **J Consult Clin Psychol** 2008;76:1046–1057.

39. Westin VZ, Schultin M, Hesser H, et al. Acceptance and commitment therapy versus tinnitus retraining therapy in the treatment of tinnitus: a randomized controlled trial. **Behav Res Ther** 2011;49:737–747.

40. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. **Acta Psychiatr Scand** 1983;67:361–370.

41. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. **Arch Gen Psychiatry** 1961;4:561–571.

42. Reiss S, Peterson RA, Gurley DM, McNally RJ. Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. **Behav Res Ther** 1986;34:1–8.

43. Henrich G, Herrschbach P. Questions on Life Satisfaction (FLZ)—a short questionnaire for assessing subjective quality of life. **Euro J Psychol Assess** 2000;16:150–159.

44. Frisch MB, Cornell J, Villanueva M, Retzlaff PJ. Clinical validation of the Quality Of Life Inventory: a measure of life satisfaction for use in treatment planning and outcome assessment. **Psychol Assess** 1992;4:92e101.