



One Size Does Not Fit All: Developing Common Standards for Outcomes in Early-Phase Clinical Trials of Sound-, Psychology-, and Pharmacology-Based Interventions for Chronic Subjective Tinnitus in Adults

Trends in Hearing
Volume 23: 1–16
© The Author(s) 2019
DOI: 10.1177/2331216518824827
journals.sagepub.com/home/tia


Deborah A. Hall^{1,2,3,4} , Alice Hibbert^{1,2}, Harriet Smith^{1,2}, Haúla F. Haider⁵, Alain Londero⁶, Birgit Mazurek⁷ and Kathryn Fackrell^{1,2}; for the Core Outcome Measures in Tinnitus (COMiT) initiative

Abstract

Good practice in clinical trials advocates common standards for assessing and reporting condition-specific complaints (“outcome domains”). For tinnitus, there is no common standard. The Core Outcome Measures in Tinnitus International Delphi (COMiT>ID) study created recommendations that are relevant to the most common intervention approaches for chronic subjective tinnitus in adults using consensus methods. Here, the objectives were to examine why it is important to tailor outcome domain selection to the tinnitus intervention that is being evaluated in the clinical trial and to demonstrate that the COMiT>ID recommendations are robust. The COMiT>ID study used an online three-round Delphi method with three separate surveys for sound-, psychology-, and pharmacology-based interventions. Survey data were analyzed to assess quality and confidence in the consensus achieved across surveys and stakeholder groups and between survey rounds. Results found participants were highly discriminatory in their decision-making. Of the 34 outcome domains reaching the prespecified consensus definition in the final round, 17 (50%) were unique to one intervention, while only 12 (35%) were common to all three. Robustness was demonstrated by an acceptable level of agreement across and within stakeholder groups, across survey rounds, across medical specialties (for the health-care practitioners), and across health-care users with varying tinnitus duration. There were few dissenting voices, and results showed no attrition bias. In conclusion, there is compelling evidence that one set of outcomes does not fit all therapeutic aims. Our analyses evidence robust decisions by the electronic Delphi process, leading to recommendations for three unique intervention-specific outcome domain sets. This provides an important starting point for standardization.

Keywords

assessment, patient-reported outcome measures, treatment effectiveness, stakeholder agreement

Date received: 27 June 2018; revised: 11 December 2018; accepted: 21 December 2018

¹National Institute for Health Research Nottingham Biomedical Research Centre, Ropewalk House, UK

²Hearing Sciences, Division of Clinical Neuroscience, School of Medicine, University of Nottingham, UK

³Nottingham University Hospitals National Health Service Trust, Queens Medical Centre, UK

⁴University of Nottingham Malaysia, Semenyih, Malaysia

⁵ENT Department, Hospital Cuf Infante Santo – Nova Medical School, Lisbon, Portugal

⁶Service ORL et CCF, Consultation Acouphène et Hyperacousie, Hôpital Européen G. Pompidou, Paris, France

⁷Tinnitus Center, Charite University Hospital, Berlin, Germany

Corresponding Author:

Deborah A. Hall, National Institute for Health Research Nottingham Biomedical Research Centre, Ropewalk House, 113 The Ropewalk, Nottingham, NG1 5DU, UK.

Email: deborah.hall@nottingham.ac.uk

Introduction

Subjective tinnitus describes the conscious perception of an auditory sensation that can be experienced only by the individual and is in the absence of a corresponding external stimulus. Subjective tinnitus is a common, yet very heterogeneous condition whose perceptual characteristics and impacts can vary greatly from person to person (Hall et al., 2018). Tinnitus has no singularly effective treatment, and hence, a number of clinical specialties have responsibility for provision of patient care including, but not restricted to, general practitioners, otologists, audiologists, clinical psychologists, neurologists, physical therapists, and psychiatrists. Through these health-care professionals, patients are able to access a wide range of therapeutic interventions including drug medications, sound therapies, talking-based therapies, relaxation techniques, neuromodulation, physical therapy, and complementary and alternative therapies (Baguley, McFerran, & Hall, 2013). The most common intervention approaches evaluated in clinical trials to date are the first three listed involving medications, sound therapy, and psychology-based strategies (Hall et al., 2016).

Randomized trials, and systematic reviews of such trials, provide the most reliable evidence about the effects of existing health-care interventions in terms of how they compare against one another and how new interventions compare with existing ones (Higgins et al., 2011). Well-conducted clinical trials can make a significant impact on patient care by influencing regulatory body decisions, development of clinical guidelines, and commissioning of health-care provision (Higgins et al., 2011; Tunkel et al., 2014; U.S. Department of Health and Human Services Food and Drug Administration, 1998). For these organizations, confirmatory evidence helps to establish a definitive answer to the question of clinical efficacy. Confirmatory evidence requires a predefined hypothesis about the expected treatment benefit and a trial design that provides highly reliable and statistically strong evidence of an important clinical benefit (Higgins et al., 2011; U.S. Department of Health and Human Services Food and Drug Administration, 1998). For example, in the case of tinnitus, the hypothesis should consider what tinnitus-related complaints are to be alleviated by the intervention of interest and how they should be measured quantitatively so that the sample size can be adequately powered and a statistical analysis can be conducted.

To this end, our study focused on outcomes for assessing the efficacy of three families of interventions (sound, psychology, and pharmacology). We chose these not only because they are in common practice but also because each has a different therapeutic rationale, and so clinical outcomes might reasonably differ across interventions according to how the interventions is supposed

to be working and their likelihood of showing a treatment-related change. The study focused on outcome domains that refer to any aspect of tinnitus that is or can be experienced by a patient. Examples include tinnitus loudness, tinnitus annoyance, the ability to concentrate, sense of control, or impact on work. The aim of the study was to identify a minimum set of outcome domains that are considered important for each of the three families of interventions, and this should inform the choice of measurement instruments used in clinical trials.

For tinnitus, the outcome domains and measurement instruments reported in clinical trials are numerous and diverse (Hall et al., 2016), and this precludes comparisons across interventions, as well as pooling the evidence for one type of intervention across studies (e.g., Martinez-Devesa, Perera, Theodoulou, & Waddell, 2010). One of the biggest barriers to good clinical trial design in tinnitus is the insufficient evidence base for choosing which outcomes should be assessed in clinical trials (e.g., Landgrebe et al., 2012; Londero & Hall, 2017; Tyler, Oleson, Noble, Coelho, & Ji, 2007). The current diversity and lack of agreed standards impedes the ability to confidently select the most valid and best performing measurement instrument for quantifying expected treatment-related change for a tinnitus intervention (see Prinsen et al., 2018; Terwee et al., 2018).

The most popular measurement instruments are those that assess tinnitus as a composite multidomain construct, and these are used somewhat interchangeably to test a wide range of tinnitus interventions (Hall et al., 2016). The Tinnitus Handicap Inventory (Newman, Jacobsen, & Spitzer, 1996), Tinnitus Questionnaire (TQ; Hallam, 2009), and Tinnitus Handicap Questionnaire (THQ; Kuk, Tyler, Russell, & Jordan, 1990) are just three examples. However, the tinnitus domains captured by each can dramatically differ across instruments, and few conventions apply. To illustrate this point, we briefly compare two multidomain instruments in common usage. On one hand, the TQ has 52 items covering distress, intrusiveness, hearing difficulties, sleep disturbance, and somatic complaints (Hallam, 2009), while on the other, the THQ has 27 items covering social, emotional, and behavioral effects; hearing difficulties; and outlook on tinnitus (Kuk et al., 1990). Both the domains and content of the items differ substantially from one to the other (Kennedy, Wilson, & Stephens, 2004). For example, the TQ asks about pain in the ear or head and tension in the head or neck muscles, while the THQ does not address these somatic complaints at all. Moreover, despite some common domains across the TQ and THQ, the weighting of items differs substantially from one to the other (Kennedy et al., 2004). For example, the TQ has proportionately more items asking about emotional distress than does the

THQ (37% vs. 22%), while the THQ has proportionately more items asking about hearing than does the TQ (19% vs. 13%).

The Tinnitus Guideline Development Group acting on behalf of the American Academy of Otolaryngology Head and Neck Surgery has appealed for further research to “determine which questionnaire is most useful for assessing relevant treatment effects” (Tunkel et al., 2014, p. S32). A first step toward creating the evidence base to determine such answers is to identify which outcome domains are most relevant for covering the wide range of therapeutic interventions that are available for tinnitus. At least then, tinnitus investigators would know if any of the common measurement instruments cover all of the tinnitus-related complaints that are considered to be critically important for clinical trials of their intervention of interest.

This article forms a companion to Hall et al. (2018), also in this Special Collection, reporting the Core Outcome Measures in Tinnitus International Delphi (COMiTID) study. The COMiTID study has made specific sets of recommendations about core outcome domains that differ according to whether a clinical trial is testing sound-, psychology-, or pharmacology-based interventions for tinnitus. Because these recommendations are new and challenge current practice, this article examines the basis for the recommendations, reporting a series of in-depth analyses of the COMiTID study data. The main objectives in this article are to explain why it is important to tailor outcome domain selection to the tinnitus intervention that is being evaluated in the clinical trial and to demonstrate that the outcome domain recommendations achieved by the COMiTID study electronic Delphi (e-Delphi) consensus method are robust.

Methods

Study design followed best practice recommendations set out in the Core Outcome Measures for Effectiveness in Trials (COMET) handbook v1.0 (Williamson et al., 2017). The COMiTID study considered sound-, psychology-, and pharmacology-based interventions for tinnitus because these three approaches are in most frequent usage across clinical practice internationally (Hall et al., 2011, 2016). The study comprised three phases: (a) three e-Delphi surveys to prioritize outcome domains for each family of interventions, (b) three structured face-to-face meetings to reduce the set of outcome domains to a number feasible for a clinical trial, and (c) an electronic vote on the final recommendations. Phases 2 and 3 followed the COMET handbook recommendation that “representatives of key stakeholder groups have the opportunity for discussion of the results of the surveys to agree a final core set and undertake additional

voting if required before a final COS is agreed” (Williamson et al., 2017, p. 24). This article focuses only on the first phase because this was a substantive part of the COMiTID study, created the first wave of recommendations, and had the broadest input from the international tinnitus community. Further information on the methods in Phase 1 can be found in the published protocol (Fackrell et al., 2017) and in the companion article (Hall et al., 2018), but here, we provide a summary of the key design features.

Participants

The study team took a number of steps to safeguard the relevant expertise of participants. First, recruitment was targeted in a purposeful manner through invitation. For professionals, the study team created a contacts list of 592 named individuals who were personally nominated by the members of TINnitus NETwork outcome measurement working group, identified as authors of relevant tinnitus conference proceedings in the past 3 years, authors of the clinical trials of tinnitus included in our previous systematic review (Hall et al., 2016), authors of systematic reviews of tinnitus interventions published in the past 5 years (Cochrane or otherwise), or editors of scholarly journals in Audiology or Otolaryngology. Health-care users were targeted using planned recruitment routes that included designated clinical centers, as well as a number of national and international professional networks and organizations.

Second, as part of the study enrolment process at the point of obtaining informed consent, all participants were required to sign a self-declaration statement confirming that they met one of the following eligibility criteria:

1. Health-care users were required to have experienced tinnitus for a minimum of 3 months and have current or past experience with, or be considering trying in the future, a sound-, psychology-, or pharmacology-based tinnitus intervention.
2. Health-care professionals were required to be clinically qualified and actively working within a health-care institution providing a service to adults with tinnitus, specifically those who receive a sound-, psychology-, or pharmacology-based intervention.
3. Clinical researchers were required to be academically qualified and actively working within a research organization, either currently conducting or having recently conducted research regarding clinical efficacy of a sound-, psychology-, or pharmacology-based intervention for tinnitus.
4. Commercial representatives were required to work for a company that develops, manufactures, or sells sound- or pharmacology-based products for tinnitus.

5. Funders were required to work for an organization that had recently funded relevant tinnitus research.

Commercial representatives and funders were pooled in the same stakeholder group, as per protocol (Fackrell et al., 2017), because anticipated numbers for each were smaller than for other groups. There was no planned group for commercial representatives and funders in the survey on psychology-based interventions. Table 1 displays the number of participants in each of the four stakeholder groups at each round of the three e-Delphi surveys. All participants gave informed consent, and the study was approved by the Solihull Research Ethics Committee and Health Research Authority (ref: 17/WM/0095, March 2017).

e-Delphi Surveys

The starting point for the e-Delphi was a long list of 66 candidate outcome domains. This was created via three sources: (a) systematic review of outcome domains used in clinical trials of tinnitus treatment in adults (Hall et al., 2016), (b) narrative synthesis of tinnitus-related complaints reported by patients (Hall et al., 2018), and (c) content analysis of outcome domains assessed by items in commonly reported tinnitus questionnaires. The process identified 123 unique outcome domains that were then refined to 66 through a series of health-

care user-led decisions considering which were specific to tinnitus, distinct in construct, and not associated with *how to measure* the construct (Fackrell et al., 2017; Smith et al., 2018). For each outcome domain, a plain language concept definition was coproduced with health-care users so that all participants could understand its meaning. The labeling given to each outcome domain was also conducted with patient input. The final version is published as Additional File 1 (Fackrell et al., 2017).

There were three independent e-Delphi surveys; one for sound-, psychology-, and pharmacology-based interventions, respectively. The same list of 66 outcome domains was presented in a fixed order in all three surveys, according to the following categories: behavior, body structures and functions, cognition (thought processes), coping and acceptance, effects of tinnitus on hearing, emotions, factors related to the treatment being tested, health-related quality of life, negative thoughts, perceptions of the tinnitus sound, physical health, state of mind, and support and knowledge. Participants were asked to think about each tinnitus outcome domain with respect to how important it would be to measure when deciding if a sound-, psychology-, or pharmacology-based tinnitus treatment is working. They were asked to consider its relevance and likelihood to show a treatment-related change for all interventions within that family. Participants scored each individual outcome domain using a 1 to 9 scale, whereby 1 to 3 indicated that the domain was *not important*, 4 to

Table 1. Number of Participants in Each Stakeholder Group and for Each Intervention Type That Consented and Participated in Each Round of the e-Delphi Survey.

Stakeholder group	Consented	e-Delphi Round 1	e-Delphi Round 2	e-Delphi Round 3	Attrition (%)
Sound-based interventions					
Health-care user	199	182	160	142	22.0
Health-care practitioner	79	70	60	57	18.6
Clinical researchers	36	35	35	34	2.9
Commercial reps and funders	24	21	19	19	9.5
Psychology-based interventions					
Health-care user	118	114	97	89	21.9
Health-care practitioner	63	61	57	50	18.0
Clinical researchers	39	39	37	36	7.7
Commercial reps and funders	4	4	4	3	N/A
Pharmacology-based interventions					
Health-care user	67	62	48	41	33.9
Health-care practitioner	51	47	40	37	21.3
Clinical researchers	20	17	14	13	23.5
Commercial reps and funders	19	18	15	12	33.3
Total	719^a	670	586	533^b	20.4

Note. Attrition refers to the percent who withdrew or dropped out between Rounds 1 and 3.

^aNote some individuals consented to participate in more than one study, and so when those duplicates have been accounted for, the 719 comprises 641 unique individuals.

^bFor those participating in more than one study, when duplicates have been accounted for, the 533 comprises 472 unique individuals.

6 indicated it was *important but not critical*, and 7 to 9 indicated that it was *critically important* in deciding whether a tinnitus intervention is effective (Guyatt et al., 2011). Following the COMET handbook v1.0, the study design also included an *unable to score* category to allow for the fact that some participants may not have the level of expertise to score certain outcomes. Over the three e-Delphi surveys, 1% to 2% of all responses were of this type. As expected, most often it was health-care users who chose to use this option, and least often it was clinical researchers. In Round 1, there were also open-text boxes for adding comments beside each outcome domain.

In Round 1, participants could nominate additional outcome domains that they felt had been missed from the initial list of 66, and 8 new outcome domains were added in total (2 for sound, 4 for psychology, and 2 for pharmacology). Frequency of occurrence of tinnitus episodes' was added to both sound and pharmacology, hence one outcome was repeated. In Rounds 2 and 3, participants received (numerical and graphical) feedback on the distribution of scores for the 66 outcome domains, and the new outcome domains were presented for scoring. Round 2 enabled participants to reflect on their scores in light of the distribution of scores for their own stakeholder group and to score the outcomes again. Round 3 enabled participants to reflect on their scores in light of the distribution of scores for all stakeholder groups presented separately and to score the outcomes again. DelphiManager software was used for online administration and data management (see Williamson et al., 2017).

Definition of What Constitutes Agreement About a Common Standard

All outcome domains were retained across all three rounds of the e-Delphi surveys so that participants could be free to change their scores across rounds. The voting threshold was defined, according to recommendation (Williamson et al., 2012, 2017), as at least 70% of the participants in all stakeholder groups scoring 7 to 9 and fewer than 15% in any stakeholder group scoring 1 to 3. The reason for this definition is that consensus requires agreement by the majority regarding the critical importance of the outcome domain, with only a small minority considering it to have little or no importance.

Analysis Methods

Individual survey responses were carefully evaluated in a series of analyses that used descriptive statistics. We first collated all those outcome domains that reached the pre-specified criteria for recommendation as a common standard for each intervention type and compared across the three intervention types. We examined the percentage of participants in each stakeholder group scoring 7 to 9

(in favor) and scoring 1 to 3 (against) in the final round of the e-Delphi surveys and the percentage of respondents who gave a score of 7 to 9 broken down by stakeholder group and by round of the e-Delphi survey.

Individual scores in Round 3 were also analyzed statistically to determine the degree of interrater agreement, within each stakeholder group and in each survey. A weighted kappa statistic (Fleiss, 1971) was chosen because this accounts for the meaningful order of the outcome domain scores and handles study designs with more than two raters. The 1 to 9 scale was transformed into the units relevant to the e-Delphi survey (i.e., 1–3 = *not important*, 4–6 = *important but not critical*, 7–9 = *critically important*). Following Landis and Koch (1977), weighted kappa statistics were interpreted as follows: fair agreement ($K=0.21-0.40$) and moderate agreement ($K=0.41-0.60$).

The similarities and differences of opinion between health-care practitioners coming from different medical specialties were explored by examining the average scores in Round 3, broken down post hoc by specialty subgroups. Likewise, the similarities and differences of opinion across health-care users with different durations of chronic tinnitus were explored using the same approach.

The final analysis considered attrition bias, which occurs when the participants who do not respond in subsequent rounds have different views from their stakeholder group peers who continue to participate (Williamson et al., 2017). The potential for attrition bias was investigated across the three e-Delphi survey rounds according to methods used by Bruce et al. (2015) in which a graphical representation is created to compare the response distributions of withdrawn and completing participants.

Results

The final outcome domains identified by the COMITID study can be seen in Figure 1. These make up three intervention-specific core outcome domain sets, recommended for use in all clinical trials of readily available interventions for chronic subjective tinnitus in adults. Supplementary Table S1 provides full details of the percent of participants who scored each outcome domain as *critically important* (i.e., scores 7–9). For the full findings from all three phases of the COMITID study, please see the companion article (Hall et al., 2018). Here, the results focus on addressing the two issues of interest pertaining to Phase 1, which was the e-Delphi consensus process.

Importance of Tailoring Outcome Domains to the Intervention Being Evaluated in the Clinical Trial

Two pieces of evidence support the conclusion that intervention-specific outcomes are needed when designing clinical trials in chronic subjective tinnitus.

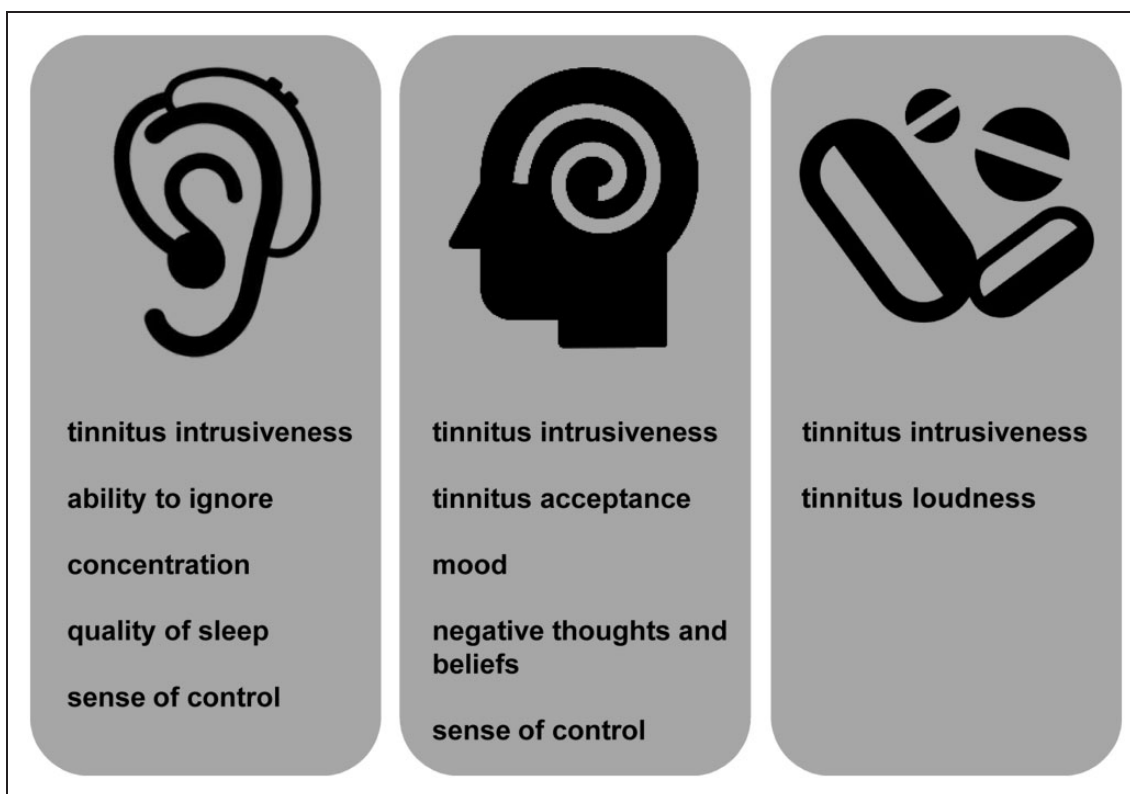


Figure 1. Graphic illustrating the COMiT'ID recommendations for core outcome domain sets for each family of interventions widely available for chronic subjective tinnitus in adults.

Unique outcome domains were considered important to each intervention type. First, for an outcome domain to be recommended as a common standard, all four stakeholder groups needed to reach the 70% threshold rating it as critically important (scores 7–9), and here, we observed discriminatory choices according to whether the intervention of interest was sound, psychology, or pharmacology based. Table 2 lists all the outcome domains that reached the prespecified criteria for recommendation as a common standard, while Supplementary Table S1 gives more details on the individual scoring. From the overall set of 34 recommended outcome domains, 17 of them (i.e., 50%) were unique to only one intervention approach. The uniquely relevant outcome domains for each intervention approach are summarized as follows. Agreed to be of critical importance only to the sound-based interventions such as hearing aids and cochlear implants were effects of tinnitus on hearing (listening, conversations); perceptions of the tinnitus sound (frequency of occurrence of tinnitus episodes, tinnitus awareness); and physical health (ability to relax). In contrast, agreed to be of critical importance only to pharmacological interventions were loudness as a perceptual characteristic of tinnitus, cognition and thought processes (confusion), and potential side effects of drug-taking (adverse reaction). Finally, agreed to be of critical

importance only for psychological therapies (e.g., talking or thinking strategies that are aimed at helping people deal with how tinnitus makes them think and feel) were cognition and thought processes associated with tinnitus (tinnitus-related thoughts, negative thoughts/beliefs, suicidal thoughts, catastrophizing); emotions associated with tinnitus (worries/concerns, fear, mood, irritable); and health-related quality of life (impact on relationships).

In contrast to the 17 outcome domains that were voted in to just one family of interventions, only 12 outcome domains (35%) reached the voting threshold for all three. These were ability to ignore, concentration, annoyance, anxiety, depressive symptoms, difficulties getting to sleep, quality of sleep, coping, tinnitus intrusiveness and impacts on individual activities, and social life and work.

Professional opinion differentiated which domains were deemed important, according to each intervention type. Second, members of the professional stakeholders groups revealed themselves to be highly discriminatory in their decision-making according to the intervention type of interest. The COMiT'ID study revealed many examples, but just three are given here for illustrative purposes. At Round 3, a much greater percentage of health-care

Table 2. All Outcome Domains That Reached the Prespecified Consensus Definition Based On the e-Delphi Round 3 Voting.

Sound-based interventions	Psychology-based interventions	Pharmacology-based interventions
Ability to ignore	Ability to ignore	Ability to ignore
Ability to relax		
Acceptance of tinnitus	Acceptance of tinnitus	Adverse reaction
		Annoyance
Annoyance	Annoyance	Anxiety
Anxiety	Anxiety	
	Catastrophizing	
Concentration	Concentration	Concentration
Conversations		
		Confusion
Coping	Coping	Coping
Depressive symptoms	Depressive symptoms	Depressive symptoms
Difficulties getting to sleep	Difficulties getting to sleep	Difficulties getting to sleep
	Fear	
Frequency of occurrence of tinnitus episodes		
Helplessness (lack of control)	Helplessness (lack of control)	
Impact on individual activities	Impact on individual activities	Impact on individual activities
	Impact on relationships	
Impact on social life	Impact on social life	Impact on social life
Impact on work	Impact on work	Impact on work
	Irritable	
Listening		
	Mood	
	Negative thoughts/beliefs	
Quality of sleep	Quality of sleep	Quality of sleep
Sense of control	Sense of control	
	Suicidal thoughts	
Tinnitus awareness		
Tinnitus intrusiveness	Tinnitus intrusiveness	Tinnitus intrusiveness
		Tinnitus loudness
	Tinnitus-related thoughts	
Tinnitus unpleasantness		Tinnitus unpleasantness
Treatment satisfaction		Treatment satisfaction
	Worries/concerns	

Note. Outcome domains presented in bold are unique to only one of the intervention types.

practitioners considered “sense of control” to be critically important for testing the efficacy of sound and psychology interventions (88% and 96%, respectively) than for pharmacological interventions (35%). Second, many more clinical researchers considered “tinnitus pitch” to be critically important for testing pharmacological interventions (77%) than for sound- or psychology-based interventions (35% and 19%, respectively). Similar patterns of views were often held across professional stakeholder groups. For example >75% of the clinical researchers and commercial representatives and funders considered pharmacokinetics to be critically

important for trials testing the efficacy of medications, but <10% considered it to be so when testing the efficacy of the other interventions. It is worth noting that health-care practitioners self-selected into each e-Delphi survey according to their expertise in the particular intervention strategy, and so this could have contributed to the discriminatory opinions expressed across the three surveys. For example, as shown in Table 3, otologists were the predominant health-care specialty in the survey considering pharmacological interventions, whereas clinical psychologists put themselves forward with expertise in psychological interventions alone, and audiologists

Table 3. Health-Care Practitioners Who Consented to Participate in the e-Delphi Survey and Then Who Completed Round 3.

	Individuals who consented	Experts in sound-based interventions	Experts in psychology-based interventions	Experts in pharmacology-based interventions
General practitioner	1	0	1	0
Otologist	53	15	13	29
Audiovestibular physician	3	1	1	2
Audiologist	57	30	14	3
Hearing aid technician	2	2	0	0
Hearing therapist	24	6	8	0
Clinical psychologist	14	0	11	0
Psychiatrist	1	1	0	0
Neurologist	1	0	0	1
Psychotherapist	1	0	0	0
Phoniatrician	1	0	0	0
Unknown	6	2	2	2
Total	164	57	50	37

Note. Some experts are represented in more than one e-Delphi survey, especially otologists.

identified themselves as experts in sound- and psychology-based interventions.

The Recommended Intervention-Specific Outcome Domains Are Robust

Seven pieces of evidence support the conclusion that the recommended intervention-specific outcome domains are robust.

High agreement across stakeholder groups. First, we observed that for every recommended outcome domain, there was a high degree of agreement across stakeholder groups participating in that e-Delphi survey. Figure 2 illustrates this point. The solid lines plotted in the four colors (one for each stakeholder group) reveal a predominance of Round 3 results positioned around the outer rim of the three radar plots; a pattern that indicates that a high percentage of participants in every group expressed strong views that those outcome domains were critically important.

High agreement across survey rounds. The radar plots (Figure 2) also illustrate the second piece of evidence which is that, for each e-Delphi survey, the scores for the recommended outcome domains were consistently high across all three e-Delphi rounds. In Figure 2, the successive rounds are denoted by dotted, dashed, and solid lines. For the psychology-based tinnitus interventions, the 70% voting threshold was reached even at Round 1 for the majority of recommended outcome domains. In this round, participants were blinded to the scores given by others. By Round 3, after participants had the opportunity to reflect on the views of

their peers and of others, most of the recommended outcome domains were being scored at the top of the scale (8 or 9) with 70% agreement. This same pattern was broadly repeated for the sound- and pharmacology-based interventions. Voting tended to be in favor of inclusion, but the numbers of participants scoring at the top of the scale for the recommended outcome domains steadily rose as the survey progressed through Rounds 1 to 3 and as opinions were shared among participants.

High agreement within stakeholder groups. Third, we observed that when all the outcome domains were considered, weighted kappa statistics indicated acceptable (fair to moderate) agreement in the Round 3 scores across the participants in each stakeholder group (Table 4). Exceptions were for health-care users and clinical researchers in the pharmacology-based survey. It is not exactly clear why these had poor agreement as Supplementary Table S1 confirms a high percent scoring *critically important*, but we note that clinical researchers comprised a relatively small group (Table 1) and greater amounts of data for health-care users were excluded because scores were in the *unable to score* category. Both of these factors might compromise the reliability of the *K* statistic.

High agreement across health-care disciplines. The health-care practitioners who participated in the survey spanned a diverse range of clinical disciplines (Table 3). The fourth piece of evidence was that these health-care practitioners were in reasonably good agreement with one another, even when they had expertise in different medical specialties. Figure 3 gives an illustrative example

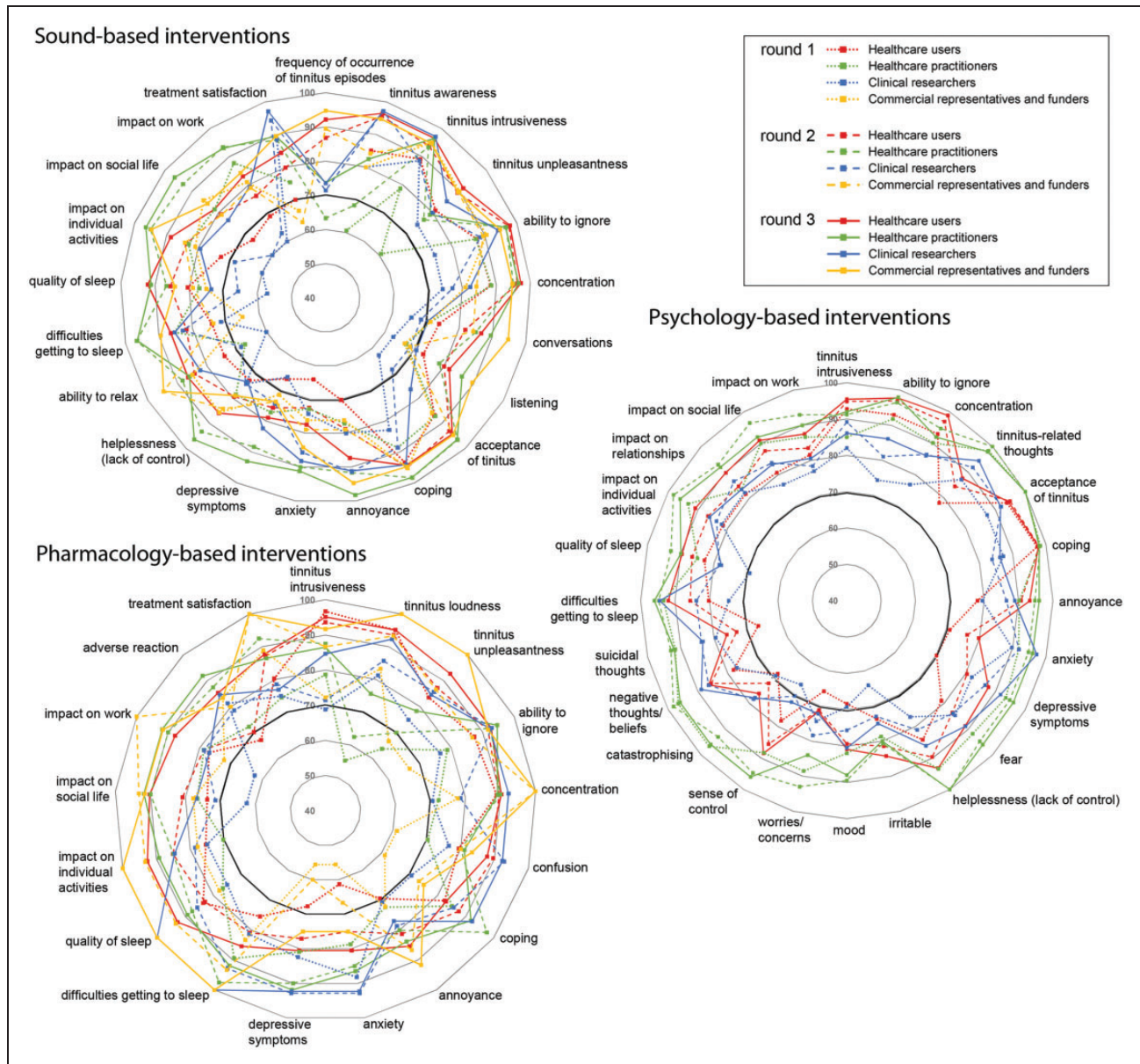


Figure 2. Percentage of respondents who gave a score of 7 to 9 for the recommended outcome domains in each family of tinnitus interventions. Percentages are shown separately for each stakeholder group and for each successive round of the e-Delphi survey. The percentages at Round 3 determined which outcome domains reached consensus. At this point, the voting threshold was defined as at least 70% of the participants in all stakeholder groups scoring 7 to 9. This 70% cutoff is indicated by the solid black line.

representing the average scores in Round 3 on the 70 outcome domains for psychology-based interventions. A high level of agreement in the pattern was found across the participating otologists, audiologists, hearing therapists, and clinical psychologists. Where differences were observed, it was due to clinical psychologists generally scoring all outcome domains lower than the other health-care professionals, not due to any substantive divergence of opinion on specific outcome domains. This supports our conclusion that, despite having differing expertise and potentially differing vantage points and

priorities for tinnitus, the range of health-care practitioners who participated in the e-Delphi surveys generally held the same opinion about what outcome domains should be measured in clinical trials of tinnitus for each intervention type. A similar pattern in results was observed for the two other e-Delphi surveys.

High agreement across health-care users, regardless of tinnitus duration. Fifth, a question asked to health-care users in the e-Delphi survey enabled us to classify duration of their chronic tinnitus experience into five time periods,

Table 4. Interrater Agreement on the Round 3 Scores Given by Each Stakeholder Group in Each e-Delphi Survey (K = Weighted Kappa Statistic; Fleiss, 1971).

e-Delphi survey	Stakeholder group	K	95% Confidence interval	
			Lower	Upper
Sound-based interventions	Health-care users	0.23*	0.17	0.28
	Health-care practitioners	0.42**	0.34	0.51
	Clinical researchers	0.38*	0.30	0.47
	Commercial representatives and funders	0.34*	0.26	0.42
Psychology-based interventions	Health-care users	0.24*	0.18	0.31
	Health-care practitioners	0.46**	0.38	0.54
	Clinical researchers	0.35*	0.28	0.42
Pharmacology-based interventions	Health-care users	0.16	0.10	0.22
	Health-care practitioners	0.32*	0.26	0.39
	Clinical researchers	0.17	0.11	0.22
	Commercial representatives and funders	0.53**	0.44	0.62

Note. *Fair agreement ($K = 0.21-0.40$) and **Moderate agreement ($K = 0.41-0.60$).

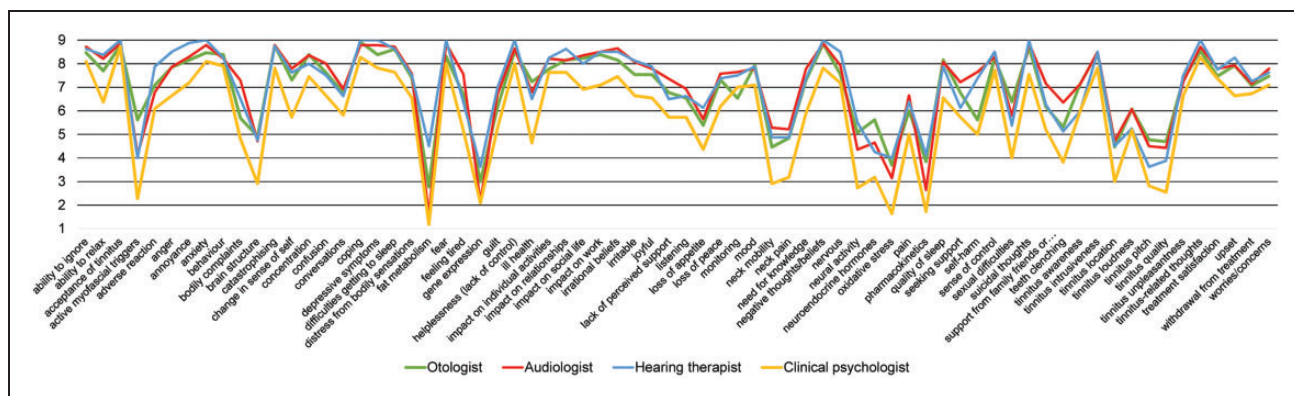


Figure 3. Average scores given by health-care practitioners in Round 3 of the e-Delphi survey for psychology-based interventions. Participating specialties were selected for illustrative purposes because of their reasonably balanced size; otologists ($n = 13$), audiologists ($n = 14$), hearing therapists ($n = 8$), and clinical psychologists ($n = 11$). The pattern for the sound- and pharmacology-based interventions is similar, but subgroup sizes are uneven.

from 3 months to more than 10 years. A high level of agreement was observed across health-care users at all tinnitus durations. For example, the average scores in Round 3 for the 68 candidate outcome domains for sound-based interventions showed similar patterns across all five durations (Figure 4). As with health-care practitioners from different medical specialties, where differences were observed, they were most often a result of one tinnitus duration group consistently rating all outcome domains lower rather than any significant divergence in scoring individual outcome domains. Noticeably, those who had experienced tinnitus for more than 10 years consistently rated all outcome domains slightly lower for importance, perhaps demonstrating a greater acceptance or resignation to the

tinnitus and less of a strong reactionary response to the outcome domains. This supports our conclusion that the recommended outcome domains are appropriate to all individuals with chronic subjective tinnitus, regardless of how long they have experienced it.

Low disagreement across stakeholder groups. Sixth, by definition, any recommendation for a common standard had to have fewer than 15% of participants in any stakeholder group holding the dissenting opinion that the outcome domain was *not important* (i.e., scoring 1–3). In fact, there were very few such opposing voices. Across those outcome domains listed in Table 2 (reaching 70% agreement), the median percentage of scores in the *not important* category was 0 (range 0%–14%), again

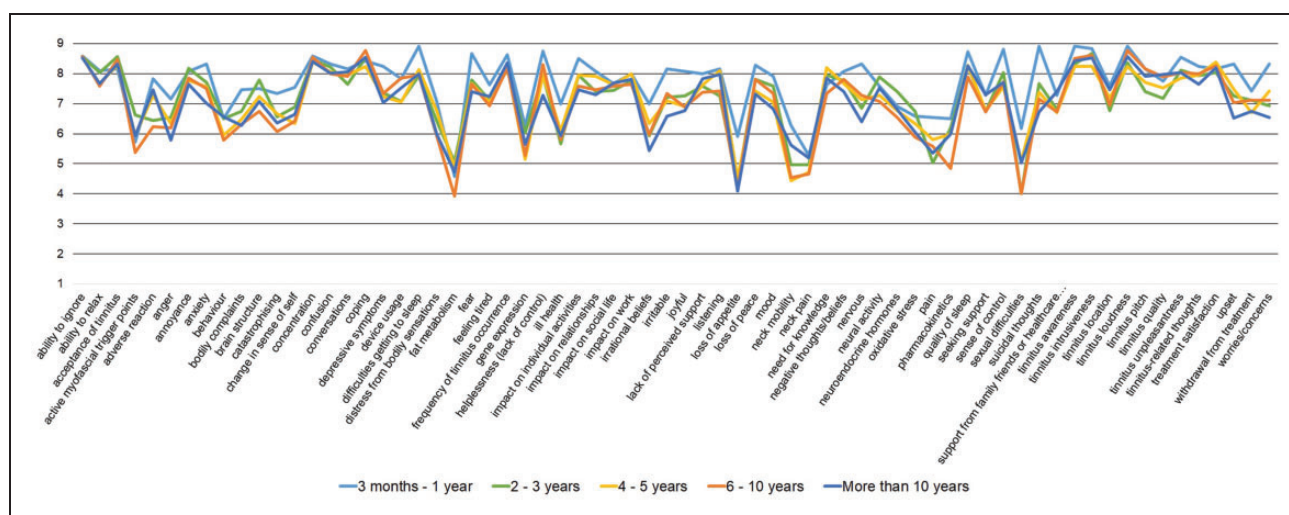


Figure 4. Illustrative example of average scores given by health-care users in Round 3 of the e-Delphi survey for sound-based interventions. Although the number of participants within each time period were not equal, all bands had a reasonable number; 3 months to 1 year ($n = 12$), 2 to 3 years ($n = 28$), 4 to 5 years ($n = 21$), 6 to 10 years ($n = 27$), and more than 10 years ($n = 58$). The pattern for the psychology- and pharmacology-based interventions is similar, and again subgroup sizes are uneven.

indicating the high level of agreement on the importance of these particular outcome domains for each intervention approach. Supplementary Table S2 gives more details on the individual scoring. From the comments submitted in the survey, we can say that when health-care users scored low, it was often because the outcome domain was not personally relevant. For example, with respect to “quality of sleep,” two people said “I get 7 hours of sleep because of my audio books” and “I get disturbed sleep for the amount of times I visit the toilet, not through tinnitus.” Professionals tended not to use the comments box to explain their scores.

Withdrawal or dropout did not appear to affect the final recommendations. Finally, the retention rate of participants across the three e-Delphi survey rounds was good (average = 80%), but consensus-based decision-making can be particularly sensitive to attrition bias because withdrawal from later rounds can sometimes be due to holding extreme views not shared by the majority of their stakeholder peers (see Williamson et al., 2017). Rarely do Core Outcome Set studies look for and evaluate such potential biases (but see Bruce et al., 2015). Our attrition analysis results clearly demonstrate that the average scores for the withdrawn and dropped out participants who completed only Round 1 (blue bars, Figure 5) or only Rounds 1 and 2 (blue bars, Supplementary Figure S1) are well contained within the average scores of those completing the corresponding successive rounds (white bars). In other words, on average, participants who completed the study scored the outcome domains similarly to those who withdrew or

dropped out from the study, indicating that attrition bias is unlikely to have affected the outcome domain recommendations.

Discussion

This article presents an in-depth exploration of the COMiTID study data for the separate e-Delphi surveys which addressed the three most commonly used types of tinnitus treatment. The intervention-specific differences in the consensus-based decisions illustrate how the tinnitus community recognizes a need to tailor outcome domains to the specific intervention being evaluated in the design of future clinical trials of chronic subjective tinnitus in adults. The outcome domain recommendations achieved using this consensus method are robust across stakeholders, including people with chronic tinnitus of all durations and tinnitus professionals working in all relevant disciplines. Hence, we are confident that the views expressed and the subsequent recommendations are representative of the whole tinnitus community.

One of the strengths of the current study methods is the value that it has placed in the opinions of people with lived experience of chronic subjective tinnitus. Involving health-care users in research and explicitly taking into account their perspectives is recognized as current best practice to be confident that the outcomes measured in a clinical trial are relevant, appropriate, and of importance in the real-world clinical setting. Published methods for selecting outcome instruments recognize that “patients are regarded the primary experts regarding patient-reported outcome measures” (Terwee et al., 2018, p. 1165). We designed

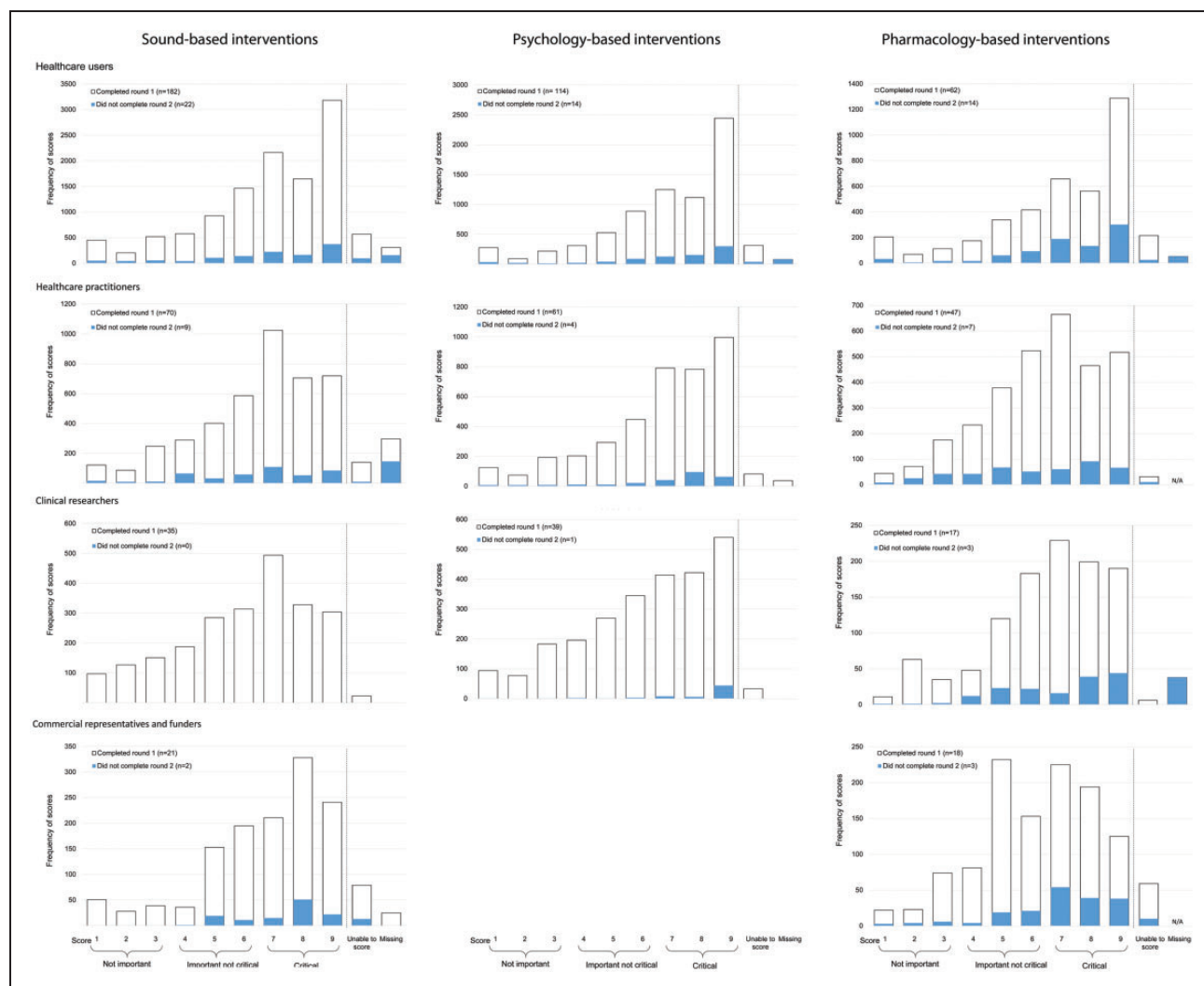


Figure 5. Round I average scores across all outcomes by stakeholder group (health-care users, health-care practitioners, clinical researchers, and commercial representatives and funders). Blue bars represent those who provided scores in Round I only; open bars represent those scoring in both Rounds I and 2.

the consensus process to meet these best practice requirements, ensuring that the recommended outcome domains are considered to be important and critical to all stakeholders alike, including health-care users alongside clinical researchers and health-care practitioners.

Another strength concerns the expertise and commitment of the members in the professional stakeholder groups; the health-care practitioners, clinical researchers, commercial representatives, and funders. Of particular note, 67% (i.e., 191/283) of those enrolled participants were nominated and invited because of their reputation and research activity. Nevertheless, we do acknowledge that the degree of knowledge is unknown, and in the case of the health-care user group, there are likely to be self-selection biases. The good retention rates across survey

rounds demonstrate the willingness of these professionals to engage in the process, a factor that is just as important for participant selection as is degree of knowledge (Hsu & Sandford, 2007).

One of the limitations of using a threshold-based criterion is that some outcome domains narrowly missed inclusion because the votes from just one stakeholder group failed to reach the threshold by a narrow margin (within 10%, i.e., 63%–69%). Two outcome domains are worth highlighting for the impact that this threshold approach had. Tinnitus unpleasantness and treatment satisfaction both reached consensus in the sound- and pharmacology-based surveys but just missed inclusion in the psychology-based survey because only 64% and 67% of clinical researchers scored 7 to 9, respectively (Supplementary Table S1). We suggest that if

investigators wish to add more outcome domains to their clinical trial designs, then tinnitus unpleasantness and treatment satisfaction would be worthy of further consideration.

The differences between the methodological requirements for good clinical trials versus good clinical research can be somewhat confusing in terms of the specification of appropriate outcomes. Clinical research can afford to be more exploratory. For example, questions may seek to shed light on the mechanism of action of the intervention, find an optimal dose, or characterize which subgroup of patients would be most responsive to a treatment. In contrast, clinical trials must have a clear and predefined hypothesized treatment benefit because they seek to provide a definitive answer to the question of whether an intervention for tinnitus is effective (Higgins et al., 2011). The recommendations arising from our e-Delphi consensus process are therefore most applicable to clinical trials, and they can direct decisions about how to measure expected therapeutic benefits. In clinical trials, there are three important reasons why it is essential for investigators to clearly define outcome domains prior to selecting what questionnaires or tests to use.

Concept Definition

Many patient-reported questionnaires intend to measure complex and unobservable concepts, and so it is important for all investigators to understand the exact nature of the concept(s) being measured. Indeed, many of the outcome domains considered in the COMiT²ID study are of this type. For each outcome domain, the study management team engaged health-care users and health-care practitioners in choosing the wording for the name and in providing a description of what it meant (Fackrell et al., 2017; Smith et al., 2018) to flesh out the concept as a more precisely defined construct. The COMET handbook v1.0 (Williamson et al., 2017) highlights this as an important step to avoid ambiguity of language, and other projects have done likewise (Bruce et al., 2015). Different questionnaires for tinnitus may intend and claim to be assessing the same concept, but without a clearly defined common construct as a starting point, the exact ways in which they each operationalize and define the concept with specific questionnaire items may result in somewhat different constructs which are not directly equivalent or comparable. This can mean that across multidomain questionnaires containing subscales that purport to measure the same outcome domain, the subscale items might actually cover entirely different concepts. Tinnitus intrusiveness is a good example. For the COMiT²ID study, our working description was “noticing the sound of tinnitus is there and it is invading your life or your personal space.” Important

for our health-care users was that this concept captures the very negative aspects of the experience, with tinnitus seen to be an unwanted presence that impedes everyday functions and activities. Comparing with existing tinnitus questionnaires, the intrusiveness subscale of the TQ seems to go some way toward capturing the same personal meaning through items such as “I feel I can never get away from the noises” (p.12) and “The noises never ‘let up’” (p.13) (Hallam, 2009). While the intrusiveness subscale of the Tinnitus Functional Index asks only about quantifying the degree of awareness, loudness, and annoyance of tinnitus on a numerical scale (Meikle et al., 2012). Interestingly, these three concepts were identified as three distinct candidate outcome domains separate from tinnitus intrusiveness in our COMiT²ID study, and they did not all reach the voting threshold to be recommended alongside tinnitus intrusiveness.

The concept descriptions generated during the COMiT²ID study provide an invaluable resource so that ambiguity of language is minimized, and all investigators precisely understand the concepts that have been recommended. Because the descriptions were always visible to participants in the e-Delphi survey, we can be confident that the scoring was conducted on this basis.

Content Validity

This is the most important measurement property of any patient-reported questionnaire so that investigators can be confident in selecting the most valid instrument for quantifying the expected treatment-related change (Prinsen et al., 2018; Terwee et al., 2018). Content validity is emphasized by regulatory authorities, such as the U.S. Food and Drug Administration (U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, & Center for Devices and Radiological Health, 2009) and the European Medicines Agency (Committee for Medicinal Products for Human Use, 2005). To have content validity, a questionnaire should contain items that are all relevant to the concept within a specific population and context of use, it should be comprehensive with respect to patient concerns with no important aspects missing, and it should be understood by patients as intended.

The work presented here is important with respect to content validity because it defines the scope of the concepts that are considered to be critically important when assessing whether or not an intervention of interest has worked. For example, effects of tinnitus on the ability to understand somebody talking (e.g., TV and radio) and on the ability to listen, understand, and take part in conversations were agreed to be of critical importance only to the sound-based interventions. Conversely, questions

about “listening” and “conversations” are not necessary to ask when assessing psychology- and pharmacology-based interventions.

Responsiveness

Greater knowledge about the sensitivity of any instrument with respect to its ability to measure treatment-related change is important so that investigators can be confident in selecting the best performing measurement for use as an end point in a clinical trial. Such knowledge can also help to understand what size of change to expect so that the clinical trial can be adequately powered to detect that benefit (Jones, Carley, & Harrison, 2003). Our recommendations for tailored outcome domains are relevant to both of these considerations. It is reasonable to assume that those outcome domains that have been carefully chosen by tinnitus experts for their suitability as outcomes in a clinical trial could point to specific scales or instruments that have a high likelihood of being responsive. Numerous multidomain tinnitus questionnaires already exist that contain items corresponding to some of the outcome domains identified by the consensus process. However, while multidomain questionnaires may meet clinical needs, such as for selecting or categorizing patients, they are less useful as tools for assessing outcome in clinical trials. Unless all the items and subscales in an instrument are relevant to the intervention being tested, then information can be lost (Hallam, 2009) or difficult to interpret (Johnston et al., 2013) when a composite global score is employed as a measure of change over time. An alternative approach is to select a number of measurement instruments, each one corresponding to one outcome domain, and this is the approach that we advocate. For example, we have explored trial data to show that a patient-reported measure of tinnitus loudness using a Likert scale is probably able to detect only large changes (at least 3.5 points out of 11; Hall, Mehta, & Fackrell, 2017).

Terwee et al. (2018) have cautioned that content validity does not necessarily mean that the instrument is responsive. For example, one might measure the incomplete or incorrect concept very reliably, and a real change in the concept of interest may be over or underestimated due to irrelevant or missing items. Our previous examination of tinnitus loudness measures indicates that an investigator-administered test of loudness matching failed to measure a sensation that was meaningful to people who lived with the experience of tinnitus and hence probably has poor content validity (Hall et al., 2017).

The current work focuses on the selection of outcomes to provide confirmatory evidence of clinical efficacy. Nevertheless, there can be much to gain in a clinical trial by collecting qualitative data that are not constrained by a priori determination. For example, the UK Medical Research Council framework recommends that both

qualitative and quantitative methods are important when evaluating complex interventions, that is, those that contain several interacting therapeutic components (Craig et al., 2008). Qualitative data could help to access the thoughts and feelings of participants, enabling a richer understanding of the meaning that people ascribe to their experiences of the tinnitus intervention. In particular, we would advocate qualitative data to explore the way in which an intervention is implemented, to shed light on why an intervention might have failed or had unexpected consequences or to understand why a successful intervention worked and how it could be further optimized.

Conclusions and Next Steps

Developing and employing a common standard for clinical trials of chronic subjective tinnitus would aid the global tinnitus community in developing a shared understanding of the concepts underpinning each tinnitus-related domain, in creating a convention for labeling those domain concepts, and in recommending which domains are most relevant to be measured as clinical trial outcomes that might provide confirmatory evidence for clinical efficacy. The COMiTID study and resulting core outcome domain set recommendations are intended to provide this framework for greater comparability across clinical trials. The next step is gaining wider awareness and endorsement of the core outcome domains across the global tinnitus community; hence, the importance of reporting the e-Delphi process and fully explaining the findings to ensure misunderstanding does not inadvertently become a barrier to implementation and uptake. From this article, it should be clear that the outcome domain recommendations are robust, supported by and representative of the wide variety of tinnitus stakeholders, and also that the split between sound-, psychology-, and pharmacology-based interventions is justified.

It is premature to recommend any existing instrument as *the preferred* instrument as this will require rigorous evaluation of content validity for the recommended outcome domains and health-care users' interpretation and against the standards required by the regulatory authorities, commissioners of health-care services, and health insurers. However, the COMiTID study takes us one step closer. We strongly urge investigators seeking to use the best available evidence base to select outcomes in a tinnitus clinical trial to consider the COMiTID recommendations.

Acknowledgments

We thank Professor Paula Williamson, University of Liverpool, for commenting on the study protocol; Richard Crew, University of Liverpool, for providing access to the DelphiManager software and for managing the data; and Felix Meyer for analyzing the data for attrition bias. The COMiTID

team acknowledges the support of the National Institute for Health Research (NIHR) Clinical Research Network in participant recruitment, and our public research partners—Brian Thacker and Veronica Colley—for their input into study design.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was primarily funded through the NIHR Nottingham Biomedical Research Centre and European Cooperation in Science and Technology Action (BM1306). D. A. H. is an NIHR senior investigator. Small research grants were awarded by Action on Hearing Loss to purchase relevant software licenses and to create the Introduction video described in this article and by British Tinnitus Association to support Public Research Partner Involvement. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article. The views expressed in this article are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health and Social Care.

ORCID iD

Deborah A. Hall  <http://orcid.org/0000-0002-3804-1452>

Supplemental Material

Supplemental material for this article is available online.

References

- Baguley, D., McFerran, D., & Hall, D. (2013). Tinnitus. *The Lancet*, 382(9904), 1600–1607. doi:10.1016/S0140-6736(13)60142-7
- Bruce, I., Harman, N., Williamson, P., Tierney, S., Callery, P., Mohiuddin, S., . . . O'Brien, K. (2015). The management of Otitis Media with Effusion in children with cleft palate (mOMent): A feasibility study and economic evaluation. *Health Technology Assessment*, 19(68), 1. doi: 10.3310/hta19680.
- Committee for Medicinal Products for Human Use. (2005). *Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products* (Doc. Ref. EMEA/CHMP/EWP/139391/2004). London, England: European Medicines Agency.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*, 337(2008), a1655. doi: 10.1136/bmj.a1655.
- Fackrell, K., Smith, H., Colley, V., Thacker, B., Horobin, A., Haider, H. F., . . . Hall, D. A. (2017). Core Outcome Domains for early phase clinical trials of sound-, psychology-, and pharmacology-based interventions to manage chronic subjective tinnitus in adults: The COMIT'ID study protocol for using a Delphi process and face-to-face meetings to establish consensus. *Trials*, 18(1), 388. doi:10.1186/s13063-017-2123-0
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378. doi:10.1037/h0031619
- Guyatt, G. H., Oxman, A. D., Kunz, R., Atkins, D., Brozek, J., Vist, G., . . . Schünemann, H. J. (2011). GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of Clinical Epidemiology*, 64(4), 395–400. doi:10.1016/j.jclinepi.2010.09.012
- Hall, D. A., Fackrell, K., Li, A. B., Thavayogan, R., Smith, S., Kennedy, V., . . . Haider, H. F. (2018). A narrative synthesis of research evidence for tinnitus-related complaints as reported by patients and their significant others. *Health and Quality of Life Outcomes*, 16(1), 61. doi:10.1186/s12955-018-0888-9
- Hall, D. A., Haider, H., Szczepek, A. J., Lau, P., Rabau, S., Jones-Diette, J., . . . Fuller, T. (2016). Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults. *Trials*, 17(1), 270. doi:10.1186/s13063-016-1399-9
- Hall, D. A., Láinez, M. J., Newman, C. W., Sanchez, T. G., Egler, M., Tennigkeit, F., . . . Langguth, B. (2011). Treatment options for subjective tinnitus: Self reports from a sample of general practitioners and ENT physicians within Europe and the USA. *BMC Health Services Research*, 11(1), 302. doi:10.1186/1472-6963-11-302
- Hall, D. A., Mehta, R. L., & Fackrell, K. (2017). How to choose between measures of tinnitus loudness for clinical research? A report on the reliability and validity of an investigator-administered test and a patient-reported measure using baseline data collected in a phase IIa drug trial. *American Journal of Audiology*, 26(3), 338–346. doi:10.1044/2017_AJA-16-0129
- Hall, D. A., Smith, H., Hibbert, A., Colley, V., Haider, H. F., Horobin, A., . . . Fackrell, K. (2018). The COMIT'ID study: Developing core outcome domains sets for clinical trials of sound-, psychology-, and pharmacology-based interventions for chronic subjective tinnitus in adults. *Trends in Hearing* 22: 2331216518814384. doi: 10.1177/2331216518814384
- Hallam, R. S. (2009). *TQ manual of the tinnitus questionnaire, revised and updated*. London, England: Polypress Press.
- Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., . . . Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, d5928. doi:10.1136/bmj.d5928
- Hsu, C. C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation*, 12(10), 1–8. Retrieved from <http://pareonline.net/getvn.asp?v=12&n=10>
- Johnston, B. C., Patrick, D. L., Busse, J. W., Schünemann, H. J., Agarwal, A., Guyatt, G. H. (2013). Patient-reported outcomes in meta-analyses – Part 1: Assessing risk of bias and combining outcomes. *Health and Quality of Life Outcomes*, 11(1), 109. doi:10.1186/1477-7525-11-109
- Jones, S. R., Carley, S., & Harrison, M. (2003). An introduction to power and sample size estimation. *Emergency Medicine Journal*, 20(5), 453–458. doi:10.1136/emj.20.5.453

- Kennedy, V., Wilson, C., & Stephens, D. (2004). Quality of life and tinnitus. *Audiological Medicine*, 2(1), 29–40. doi:10.1080/16513860410027349
- Kuk, F. K., Tyler, R. S., Russell, D., & Jordan, H. (1990). The psychometric properties of a tinnitus handicap questionnaire. *Ear and Hearing*, 11(6), 434–445.
- Landgrebe, M., Azevedo, A., Baguley, D., Bauer, C., Cacace, A., Coelho, C.,...van de Heyning, P. (2012). Methodological aspects of clinical trials in tinnitus: A proposal for an international standard. *Journal of Psychosomatic Research*, 73(2), 112–121. doi:10.1016/j.jpsychores.2012.05.002
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. doi:10.2307/2529310
- Londero, A., & Hall, D. A. (2017). Call for an evidence-based consensus on outcome reporting in tinnitus intervention studies. *Frontiers in Medicine*, 4, 42. doi:10.3389/fmed.2017.00042
- Martinez-Devesa, P., Perera, R., Theodoulou, M., & Waddell, A. (2010). Cognitive behavioural therapy for tinnitus. *Cochrane Database of Systematic Reviews*, (9). CD005233. doi:10.1002/14651858.CD005233.pub3
- Meikle, M. B., Henry, J. A., Griest, S. E., Stewart, B. J., Abrams, H. B., McArdle, R.,...Folmer, R. L. (2012). The tinnitus functional index: Development of a new clinical measure for chronic, intrusive tinnitus. *Ear and Hearing*, 33(2), 153–176. doi:10.1097/AUD.0b013e31822f67c0
- Newman, C. W., Jacobson, G. P., & Spitzer, J. B. (1996). Development of the tinnitus handicap inventory. *Archives of Otolaryngology – Head & Neck Surgery*, 122(2), 143–148. doi:10.1001/archotol.1996.01890140029007.
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147–1157. doi:10.1007/s11136-018-1798-3
- Smith, H., Horobin, A., Fackrell, K., Colley, V., Thacker, B., & Hall, D. A. (2018). Defining and evaluating novel procedures for involving patients in core outcome set research: Creating a meaningful long list of candidate outcome domains. *Research Involvement and Engagement*, 4(1), 8. doi:10.1186/s40900-018-0091-5
- Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J.,...Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, 27(5), 1159–1170. doi:10.1007/s11136-018-1829-0
- Tunkel, D. E., Bauer, C. A., Sun, G. H., Rosenfeld, R. M., Chandrasekhar, S. S., Cunningham, E. R. Jr.,...Henry, J. A. (2014). Clinical practice guideline: Tinnitus. *Otolaryngology – Head and Neck Surgery*, 151(2 Suppl), S1–S40. doi:10.1177/0194599814545325
- Tyler, R. S., Oleson, J., Noble, W., Coelho, C., & Ji, H. (2007). Clinical trials for tinnitus: Study populations, designs, measurement variables, and data analysis. *Progress in Brain Research*, 166, 499–509. doi:10.1016/S0079-6123(07)66048-8
- U.S. Department of Health and Human Services Food and Drug Administration. (1998). *Guidance for industry on providing clinical evidence of effectiveness for human drugs and biological products*. Rockville, MD: Food and Drug Administration.
- U.S. Department of Health, Human Services Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, & Center for Devices and Radiological Health. (2009). *Guidance for industry patient-reported outcome measures: Use in medical product development to support labeling claims*. Rockville, MD: Food and Drug Administration.
- Williamson, P. R., Altman, D. G., Bagley, H., Barnes, K. L., Blazeby, J. M., Brookes, S. T.,...Kirkham, J. J. (2017). The COMET handbook: Version 1.0. *Trials*, 18(3), 280. doi:10.1186/s13063-017-1978-4
- Williamson, P. R., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., & Tugwell, P. (2012). Developing core outcome sets for clinical trials: Issues to consider. *Trials*, 13(1), 132. doi:10.1186/1745-6215-13-132