Otology & Neurotology

How do we know that our patients have benefitted from our ENT/Audiological interventions? Presented at the Annual Meeting of ADANO 2016 in Berlin

--Manuscript Draft--

Manuscript Number:	ON-17-853R1
Full Title:	How do we know that our patients have benefitted from our ENT/Audiological interventions? Presented at the Annual Meeting of ADANO 2016 in Berlin
Article Type:	Special Feature (With Editor's Approval)
Keywords:	questionnaire; Hearing loss; measures; mild-to-moderate sensorineural hearing loss; single sided deafness; tinnitus
Corresponding Author:	Deborah Hall, PhD National Institute for Health Research Nottingham, UNITED KINGDOM
Corresponding Author's Institution:	National Institute for Health Research
First Author:	Deborah Hall, PhD
First Author Secondary Information:	
Order of Authors:	Deborah Hall, PhD
	Pádraig Kitterick, PhD
	Eithne Heffernan, PhD
	Kathryn Fackrell, PhD
	Laura Lucas, Msc
	Melanie Ferguson, PhD
Abstract:	This short review article gives an introduction to some of the fundamental concepts and challenges facing measurement in hearing healthcare practice and research. The impact of hearing loss almost always extends beyond the sensory impairment itself, even when the measured degree of audiometric loss is mild. Yet, going beyond audibility, into the realm of measuring impact, takes us into a much more complex and less well-defined space. How does one therefore best measure the therapeutic benefit for evaluating efficacy or for clinical practice audit? Three case studies illustrate approaches to overcome such challenges. Each example highlights the importance of thinking critically about what it is one is seeking trying to measure, rather than selecting a questionnaire instrument based simply on its popularity or accessibility. We conclude by highlighting the important role that clinicians can play in collecting clinical data about their preferred instruments so that we have some evidence to inform decisions about good practice (content validity etc). We would also strongly support open data sharing as we believe that this is one of the best ways to make the most rapid progress the field.
Additional Information:	
Question	**Response**
Please provide the Word Count of your manuscript text. Include only the main body of text (exclude abstract, references, figures, and table legends).	3713
Please provide the number of main figures in your submissions. Do not include figure parts (ex. a, b, c) or supplemental figures in the count.	2
Please provide the number of tables in your submission.	1

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Response to Reviewers' Comments

Reviewer #1:

1. The short paper would benefit from footnoting the citations from the literature in the journal style.

We have amended the citations in the manuscript so that they adhere to the journal style.

Reviewer #2:

1. In this paper the authors outline their approach to development of the SPaRQ, which seems good. The SPaRQ has not been applied widely though so there are no data across different centres that support the tool. I guess the idea is to present it and hope that it will be studied further in the future.

To address this comment, we have clarified in the manuscript our reasons for choosing to outline the development of the SPaRQ, despite the fact that it is a new tool and thus has not yet been widely used (Page 3, Lines 131-136). Specifically, the purpose of presenting the SPaRQ was to provide a case study of the utilisation of best practice questionnaire development techniques. The use of such techniques, particularly qualitative research with patients and Rasch analysis, to develop questionnaires remains rare in the field of hearing research. Therefore, the SPaRQ is a unique example of the multi-stage, multi-method process necessary for designing a high quality questionnaire.

2. One must wonder whether another questionnaire tool is really needed. Rather than increasing knowledge, it is quite possible that adding more tools just dilutes the applicability of the whole field, but that remains to be seen.

We have amended the manuscript by acknowledging that a limitation of this research is that it adds another questionnaire to the range of existing questionnaires that are currently being used in hearing research. In addition, we have stated that it is sometimes necessary to develop new measures, like the SPaRQ, in order to address the lack of gold standard measures in this field. Finally, we have proposed that what is needed is guidance for researchers and clinicians to help them choose a suitable measure from the range available to them (Page 4, Lines 179-185).

EDITORIAL COMMENTS:

1. Minimal revisions are required for this paper, but the references cited within the text do need to be updated to the journal's reference style. Within the text, each reference should be cited using its corresponding number from the main reference list.

We have amended the references cited in the manuscript so that they adhere to the journal style.

Title page

Full title: How do we know that our patients have benefitted from our ENT/Audiological interventions? Presented at the Annual Meeting of ADANO 2016 in Berlin

Short running head: Measuring patient benefit and harms

Authors: Deborah A Hall PhD [1,2], Pádraig Kitterick PhD [1,2], Eithne Heffernan PhD[1,2], Kathryn Fackrell PhD [1,2], Laura Lucas MSc [1,2], Melanie Ferguson PhD [1,3]

Affiliations:

1. NIHR Nottingham Biomedical Research Centre, Ropewalk House, 113 The Ropewalk, Nottingham, NG1 5DU. UK

2. Otology and Hearing Group, Division of Clinical Neuroscience, School of Medicine, University of Nottingham, Nottingham, NG7 2UH

3. Nottingham University Hospitals NHS Trust. Queens Medical Centre, Derby Road, Nottingham, NG7 2UH

Conflicts of interest and sources of funding: This article presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the National Health Service, the NIHR or the Department of Health. None of the authors have any conflicts of interest to declare.

Correspondence address: NIHR Nottingham Biomedical Research Centre, Ropewalk House, 113 The Ropewalk, Nottingham, NG1 5DU. UK

Correspondence telephone: +44 (0) 115 823 2600

Correspondence e-mail: deborah.hall@nottingham.ac.uk

REVIEW ARTICLE

How do we know that our patients have benefitted from our ENT/Audiological interventions?

ABSTRACT

This short review article gives an introduction to some of the fundamental concepts and challenges facing measurement in hearing healthcare practice and research. The impact of hearing loss almost always extends beyond the sensory impairment itself, even when the measured degree of audiometric loss is mild. Yet, going beyond audibility, into the realm of measuring impact, takes us into a much more complex and less well-defined space. How does one therefore best measure the therapeutic benefit for evaluating efficacy or for clinical practice audit? Three case studies illustrate approaches to overcome such challenges. Each example highlights the importance of thinking critically about what it is one is seeking trying to measure, rather than selecting a questionnaire instrument based simply on its popularity or accessibility. We conclude by highlighting the important role that clinicians can play in collecting clinical data about their preferred instruments so that we have some evidence to inform decisions about good practice (content validity etc). We would also strongly support open data sharing as we believe that this is one of the best ways to make the most rapid progress the field.

INTRODUCTION

The purpose of this short article is to introduce the reader to some of the fundamental concepts and challenges facing measurement in healthcare practice and research. The concept of measurement will perhaps be most familiar to the reader in the context of the audiogram. The audiogram plots air conduction threshold for tones presented to either ear and is useful for determining hearing sensitivity. Pure-tone averages can be interpreted with respect to standard category boundaries, such as mild hearing loss (26-40 dB A) (1). The impact of hearing loss almost always extends beyond the sensory impairment itself, even when the measured degree of audiometric loss is mild. It is well known that residual hearing is not related in any straightforward way to the burden of disability experienced by a person with hearing loss (2). Going beyond audibility, into the realm of measuring impact, takes us into a much more complex and less well-defined space. For example, mild-to-moderate hearing loss has been reported by patients to interfere with hearing environmental sounds, listening, communicating, speaking, and it can negatively affect family life, social relationships, and ability to work. On a personal level, the negative stigma can affect personal identity, promote a sense of isolation, negative emotions such as frustration, distress and depression. Hearing loss can also increase the effort required for listening and communicating causing fatigue (3). The impact of hearing-related problems, such as tinnitus, similarly spans a wide array of psychological and social dimensions (4).

No gold standard measure Instead of clinician-administered tests, the impacts of hearing loss and tinnitus are often assessed using a patient-reported instrument such as a questionnaire. However, there are no gold standards. This is reflected in the lack of consensus in the selection of questionnaires for hearing studies (4, 5).

Diversity of patient complaints Given the diversity of reported complaints, every patient presents with a complex array of symptoms and functional impacts. Moreover, any clinician or researcher who has worked with people with a hearing-related problem will appreciate that every individual's experience is a very personal one. In practical terms, while one person's primary motivation for seeking medical help might be because their hearing-related problem means that they no longer enjoy socialising with friends

down the pub, for another it may be because their ability to play in an orchestra is hindered, while for another hearing loss might make it difficult watching television comfortably with their spouse. The impact of hearing loss is therefore a construct that is very individualised and personal.

Practical challenges This situation presents the ENT/Audiology professional with two major practical challenges; the first concerns how to comprehensively assess a patient for a precise clinical diagnosis, and the second concerns how to measure the therapeutic benefit for evaluating efficacy or for clinical practice audit. With some degree of success, the challenge for clinical diagnosis has been resolved by creating multi-attribute questionnaire instruments whose scores can be used to discriminate between individuals. For example, the Hearing Handicap Inventory for the Elderly (HHIE) asks 25 questions about the emotional consequences of hearing impairment, social and situational effects (6), with pre-defined cut-offs for determining "no handicap", "mild to moderate handicap" and "significant handicap".

However, the solution to the first challenge tends to be incompatible with evaluating therapeutic benefit. This is because questionnaire items that discriminate well between different patients at the diagnostic appointment are not necessarily sensitive to evaluating changes over time within the same patient (7). And it is difficult to design a questionnaire instrument that is both discriminative *and* evaluative. To illustrate this with an example, tinnitus-related emotional distress and auditory difficulties might both discriminate one patient from another, but only one of these might be responsive to treatment (e.g. hearing aids should reduce auditory difficulties, but might not reduce distress). Averaging the benefit scores for these components could therefore compromise the sensitivity of an aggregated score to measuring treatment-related change. As a general rule, questionnaire instruments that successfully measure therapeutic benefit in different situations tend to be those with good statistical properties that enable the clinician or investigator to interpret specific complaints rather than a global non-specific construct like "severity" or "handicap" (8).

In this short review, we present three case studies which illustrate approaches to overcome the challenges of evaluating therapeutic benefit. These examples highlight the need to think critically about what it is one is seeking trying to measure, rather than selecting a questionnaire instrument based simply on its popularity or accessibility.

Measuring psychosocial functioning of adults with mild-to-moderate hearing loss

The International Classification of Functioning, Disability, and Health (ICF) is a biopsychosocial framework designed to standardise the description, measurement, clinical assessment, and teaching of functioning, disability, and health for researchers, clinicians, clinical educators, and policy-makers around the world (9). The ICF consists of three primary domains of patient burden: (1) physical impairments, or deficits in body functions or body structures, (2) activity limitations, or problems executing tasks and actions, and (3) participation restrictions, or problems with involvement in life situations. These domains are influenced by both environmental factors and personal factors (9, 10). The ICF also includes a comprehensive taxonomy of categories of functioning (e.g. listening, education, self-care). The categories of functioning most relevant to hearing loss have been identified by a large, cross-cultural, mixed-methods study (10). Therefore, the ICF could be used in the future to standardise the measurement of individuals with hearing loss in clinical practice or in research.

The domain of participation restrictions is thought to be the most difficult of the ICF domains to measure (11). One obstacle is that the conceptualisation of participation restrictions is imprecise and inconsistent (12). The WHO (2001) definition above is rather

broad, which means that it is difficult to distinguish participation restrictions from related constructs, such as activity limitations and quality of life (13). Also, there is no consensus regarding the categories of functioning that should be included in a participation restrictions measurement instrument (14). Another obstacle is that different people participate in different ways, depending on their personal preferences and circumstances. It is difficult to capture such a highly individual construct in one standardised tool (13). One solution is to develop different questionnaire instruments for different subgroups (15). However, this can impede comparisons across groups and across studies. Another solution is to create patient-generated measurement tools that permit respondents to personalise their content. However, personalised instruments may not be well suited to the grouping of scores or comparisons across time periods and across individuals. Also, they can be difficult for some respondents to understand and complete (16). Another approach is to obtain counts of social interaction frequency or social network size (17). However, such measures fail to acknowledge that the quality of social contacts can be more important for wellbeing than quantity of social contacts (18).

**** insert Table 1 about here ****

Here, we provide a case study of the utilisation of best practice techniques to develop a hearing-specific measure of participation. Best practice techniques, which include qualitative methodologies (e.g. cognitive interviews) and modern psychometric analysis (e.g. Rasch analysis), are necessary for the creation of gold standard measures. However, to date, these techniques have seldom been employed in the development of hearing-specific measures.

The questionnaire we developed, entitled the Social Participation Restrictions Questionnaire (SPaRQ), was designed to serve as an outcome measure in either research or clinical practice. The SPaRQ consists of a 9-item subscale measuring social behaviours (e.g. difficulties with social interactions) and a 10-item subscale measuring social perceptions (e.g. feelings of isolation). It uses an 11-point response scale (0=completely disagree, 10=completely agree) because a broad range of response options are considered to enhance responsiveness (19). The SPaRQ was designed by conducting a series of qualitative and quantitative studies (see Table 1) in accordance with internationally-recognised guidelines from the questionnaire development literature (20, 21). Our aim was to ensure that the measurement properties of the SPaRQ met the standards required of outcome measures used in clinical practice and in clinical trials (21).

The first step was to create a precise conceptual model of hearing-related participation restrictions and to determine the categories of functioning that should be included in the measure by (1) reviewing the literature, including existing questionnaire instruments and the ICF, and (2) interviewing adults with hearing loss and hearing healthcare professionals (22). The second step was to evaluate the content validity of the SPaRQ, including its relevance, clarity, acceptability, and potential responsiveness, by (1) conducting cognitive interviews with adults with hearing loss and (2) surveying hearing healthcare professionals. Qualitative research with key stakeholders is an often overlooked but essential component of questionnaire development, at it ensures that the instrument adequately captures the respondents‘ experiences, uses everyday language, and is easy to administer and complete (23). The third step was to assess the psychometric properties of the SPaRQ by applying (1) Rasch analysis and (2) traditional (i.e. Classical Test Theory) psychometric analysis to data collected from adults with hearing loss. Whilst most hearing-specific questionnaires have been developed using traditional psychometric analysis alone, a modern approach (i.e. Rasch analysis or Item Response Theory) should also be applied because it enables all the relevant psychometric properties (e.g. unidimensionality, differential item functioning) to be adequately assessed (24). The outcome of this rigorous development process was the production of

a questionnaire that possesses an array of good measurement properties. For instance, each subscale was found to be unidimensional, which means that all of the items within a subscale measure the same construct, and well-targeted, which means that the subscales have high measurement precision and capture a wide range of participation restrictions. There was also evidence to support the convergent validity of the subscales with each one displaying strong, positive correlations with a hearing-specific disability measure and moderate, positive correlations with a generic disability measure and a mental health screening tool. Responsiveness of the SPaRQ is yet to be examined, but this is planned for future research.

One limitation of this research is that it adds another questionnaire to the range of existing questionnaires that are currently being used in hearing research (5). However, it is sometimes necessary to develop new measures in accordance with the latest best practice recommendations in order to address the lack of gold standard measures in the field. In the future, researchers and clinicians would benefit from the introduction of guidelines to help them to identify high quality measures that are appropriate for their purposes.

Relevance of existing questionnaires for assessing burden of single-sided deafness (SSD)

At face value, single-sided deafness (SSD) would appear to be a form of hearing loss where the task of determining whether or not a patient has benefitted from an intervention should be relatively straightforward. Lack of hearing on one side of the head would be expected to hinder access to acoustic information in that hemifield and disrupt the ability to segregate information from different sources (25). One might also be tempted to assume that relevant interventions for this patient group are those that address these impaired listening skills, and benefit should be measured in terms of the extent to which they have restored or improved such skills. However, some of the earliest published observations about these patients remarked on the unexpected degree of burden that impairments to these listening skills impose on the patient. Harford and Barry noted "the persistence and earnestness of reports from unilaterally hearing-impaired individuals stating serious difficulty encountered in many common listening situations" (26). Early work also suggested a breadth and depth of burden that one might not predict from these functional difficulties. Giolas and Wark noted that a majority of patients reported strong negative emotions that included embarrassment and helplessness (27). The extent of these feelings was such that they recommended they should be addressed actively as part of their clinical management, an approach that is still recommended almost 50 years later (28).

The incongruence between the fact that SSD patients still have access to one 'good' hearing ear and the chronic and complex burden that they report is perhaps why there is an increasing focus on the surgical restoration of hearing in their deaf ear (29) rather than traditional interventions that re-route sound between the ears (30). Early-phase trials have suggested that cochlear implantation is capable of restoring bilateral input and addressing, at least in part, the functional impairments of SSD (31, 32). However, as the field moves beyond demonstrations of clinical efficacy in the form that can be measured using controlled listening tests in the clinic or laboratory, increasing emphasis will inevitably be placed on conducting trials to measure broader impacts on quality of life to demonstrate the additional benefits to health it provides over currently available treatments.

In designing these trials, one must first ask whether the intervention addresses one or more aspects of burden that are relevant to SSD patients, and what specific aspects of burden are being targeted. Such knowledge would ideally be supported by evidence from

early-phase trials so that the mechanism through which the intervention works is well understood. The choice of outcomes that are being measured would also need to be examined to ask whether they are considered by patients to be important for their health and wellbeing. Finally, outcome measures should be selected based evidence for their validity to measure those outcomes in these patients. Here we describe a research process that has been designed to address these questions in the field of SSD and to lay the groundwork for the development of a Core Outcome Set (Figure 1).

**** insert Figure 1 about here ****

Fundamental to addressing many of these issues is a comprehensive understanding of the health condition itself. Little if any qualitative work around the burden imposed by SSD has been conducted since Giolas and Wark applied the Critical Incident Technique to study the functional consequences of SSD (27). This technique structures the interview around events that the patient recognises were affected by their hearing loss. Patient interviews were therefore conducted using a similar methodology to construct a hierarchical model of burden (33) based on patient-reported incidents and emerging themes from the transcripts. This qualitative approach provided a comprehensive characterisation of the impact of the health condition (34) and was initially used to assess whether interventions targeted aspects of health that are impaired by SSD. A systematic review identified those interventions and concluded that studies have focussed almost exclusively on intervening to improve functional impairments to speech perception and spatial hearing (35). However, the wide range and inconsistent use of patient-reported questionnaire instruments as outcome measures in existing trials meant that there is considerable uncertainty over what outcomes if any beyond the direct functional impairments to hearing were being targeted by these interventions (36). To address this uncertainty, a second systematic review is underway to identify what studies say they are trying to measure and to map those outcomes onto their use of specific measurement instruments (37). The content of the questionnaire instruments will be compared with the model of patient burden to assess whether they are targeting domains of health which are considered relevant by this patient group (23). The analysis will examine how successful these instruments are at targeting specific domains of health and therefore their suitability for use as outcome measures in the context of clinical trials (38).

Relevance of existing questionnaires for assessing benefits of tinnitus treatments

There is a substantial literature concerning self-assessment questionnaires for scaling the negative impacts of tinnitus. This literature shows that many different tinnitus-specific questionnaires have been used to assess treatment-related changes in tinnitus. For example, our review of clinical trials from 2006 to 2015 identified at least 78 different outcome instruments used in 228 trials; with 24 of those being different tinnitus-specific questionnaires. These were predominantly the Tinnitus Handicap Inventory (THI) (39) and the Tinnitus Questionnaire (TQ) (4, 40, 41). But even these two most popular instruments were used in only a minority of clinical research since we noted that usage was 15% and 7% out of 228 studies, for the THI and TQ respectively. We also note that these questionnaire instruments have predominantly been designed for screening and diagnostic purposes, not for measuring benefit from ENT/Audiological interventions. In particular, they measure multiple domains of patient burden.

The tinnitus community widely acknowledges that a standard is needed to ensure that therapeutic benefit is measured much more consistently across studies, and that benefit is quantified using a measurement instrument that is fit for the purpose of outcome measurement (e.g. 42). A first attempt by Langguth and 28 other colleagues in 2006 sought to develop a set of international recommendations on choice of instruments for

assessing the outcome from an intervention for tinnitus (43). The recommendations by this working group suggested four questionnaires; namely the Tinnitus Handicap Inventory (THI) (39), the Tinnitus Handicap Questionnaire (THQ) (44), the Tinnitus Questionnaire (TQ) (41) and the Tinnitus Reaction Questionnaire (TRQ) (45). These instruments were developed in diverse patient populations across the USA, UK, and Australia, but were not all developed for the same applications. In particular, while the THI, TQ and TRQ focus on aspects of psychological distress, the THQ was created to comprehensively measure the perceived degree of broad handicaps attributed to tinnitus (see Fackrell et al. (46) for a review). Nevertheless, they were chosen as they were the most widely used at the time, and had been translated for use in different languages and cultures. Their questions also broadly span the emotional impact of tinnitus, disability and handicap.

**** insert Figure 2 about here ****

In making their interim recommendation, Langguth and colleagues commented that the THI, THQ, TQ and TRQ also share a common feature in that they attempt to quantify a combination of tinnitus-related distress, disability and handicap resulting in a large overlap of their items (43). Conceptual similarity is supported by statistical evidence for a high convergent validity between the global scores. For example, pairwise correlations between the THI, THQ, TQ and TRQ range from 0.74 to 0.89 (see Fackrell et al. (46) for a review). To explore conceptual equivalence in more detail we have conducted a fine-grained evaluation of each individual questionnaire item to specify exactly what health concepts form the ingredients of each instrument. The findings from this evaluation are illustrated in Figure 2. The black cells indicate where the instrument contains at least one item that we judge to be asking about the corresponding tinnitus-related complaint. All questionnaire instruments contain items that ask about a diverse range of tinnitus-related complaints covering all the major high-level categories of impact on everyday life, such as emotional impacts or activities and relationships. However, the patchwork highlights clear differences between instruments in terms of their specific item-level content. Some of these detailed differences could be clinically important for some individuals with critical gaps where an instrument entirely misses out questions on a particular type of complaint. For example, the impact of tinnitus on physical health is explored only in the TQ ('bodily complaints') and the THQ ('ill health'). We have not yet compared the content of the instruments with available information about patient burden to assess whether they are targeting domains of health which are considered relevant by people with tinnitus (23). This analysis is planned. It will tell us how successful these instruments might be at targeting specific domains of health and therefore their suitability for use as outcome measures in the context of clinical trials of tinnitus, especially under certain circumstances (e.g. with a specific patient subtype, or for evaluating the outcome from a specific intervention).

Langguth et al. (43) appreciated some of these limitations with the THI, THQ, TQ and TRQ and so the working group agreed that in the future, a "better" questionnaire was required. Since that time, a multi-item tinnitus questionnaire has been developed in the USA using a method to select items that optimized the overall responsiveness of the outcome score to treatment-related change (47). The resulting Tinnitus Functional Index (TFI) asks 25 questions about the intrusive of tinnitus, reduced sense of control, reduced quality of life, sleep disturbance, auditory difficulties, cognitive interference, interference with relaxation, and emotional distress, with pre-defined cut-offs for determining "not a problem", "small problem", "moderate problem", "big problem", and "very big problem". When opting to use the TFI in other countries and cultures, it would be advisable to explore the content validity and severity grading in the new target population.

DISCUSSION

These three examples illustrate different approaches to overcome the challenges of evaluating therapeutic benefit. In common, they all highlight the need to think critically about what it is one is seeking trying to measure. We end our review with some concluding remarks:

- We have previously argued that it would be helpful to step away from using terms such as 'handicap' or 'severity' when naming a questionnaire instrument. These terms are not helpful to clinicians and researchers because they do not meaningfully describe exactly what health-related construct is being measured by the instrument (4). The development of the SPaRQ by Heffernan et al. provides a good example where the questionnaire name describes exactly what aspect of health the instrument claims to measure (22).

- Although often questionnaire developers typically present psychometric validations of a questionnaire instrument, the word 'validation' is quite emotive. Validity is not a fixed property, but varies across populations and cultures. Its good practice therefore to keep an open mind and to evaluate any questionnaire instrument the first time its going to be used for a particular purpose and in a particular patient population.

- At the end of the questionnaire evaluation, we might end up by failing to find any instruments which meet stringent contemporary standards of performance for outcome measures in clinical trials of SSD and tinnitus. What then? Clearly new research will be needed to modify an existing instrument, or create a new one from scratch. But what should we do in the meantime? Well, just because an instrument is not perfect does not necessarily mean that it should not be used. In this situation, clinicians can play an important role by collecting clinical data about their preferred instruments so that we have some evidence to inform decisions about good practice (content validity etc). We would also strongly support open data sharing as we believe that this is one of the best ways to make the most rapid progress the field.

REFERENCES

1. World Health Organization. Accessed June 1, 2017. [http://www.who.int/pbd/deafness/hearing_impairment_grades/en/]
2. Granberg S, Pronk M, Swanepoel DW, et al. The ICF core sets for hearing loss project: Functioning and disability from the patient perspective. *Int J Audiol* 2014;53:777-86.
3. Pichora-Fuller MK, Kramer SE, Eckert MA, et al. Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear* 2016;37(Suppl 1):5-27S.
4. Hall DA, Haider H, Szczepek AJ, et al. Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults. *Trials* 2016;17(1):270.
5. Granberg S, Dahlström J, Möller C, et al. The ICF Core Sets for hearing loss - researcher perspective. Part I : Systematic review of outcome measures identified in audiological research. International Journal of Audiology 2014;53(2):65-76.
6. Ventry IM, Weinstein BE. The hearing handicap inventory for the elderly: a new tool. *Ear Hear* 1982;3(3):128-34.
7. Guyatt G, Kirshner B, Jaeschke R: Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992;45(12):1341–5.
8. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials* 2016;17(1):449.

9. World Health Organization. The International Classification of Functioning, Disability and Health (ICF). 2001; Geneva, WHO. Accessed June 1, 2017. [http://www.who.int/classifications/icf/en/].
10. Danermark B, Granberg S, Kramer SE, et al. The creation of a comprehensive and a brief core set for hearing loss using the International Classification of Functioning, Disability and Health. *Am J Audiol* 2013;22(2):323-8.
11. Whiteneck G, Dijkers MP. Difficult to measure constructs: conceptual and methodological issues concerning participation and environmental factors. *Arch Phys Med Rehabil* 2009;90(Suppl 11):S22-35.
12. Heinemann AW, Tulsky D, Dijkers M et al. Issues in participation measurement in research and clinical applications. *Arch Phys Med Rehabil* 2010;91(9):S72-6.
13. Dijkers M. Issues in the conceptualization and measurement of participation: an overview. *Arch Phys Med Rehabil* 2010;91(Suppl 9):S5-16.
14. Eyssen IC, Steultjens MP, Dekker J et al. A systematic review of instruments assessing participation: challenges in defining participation. *Arch Phys Med Rehabil* 2011;92(6):983-97.
15. Dijkers M, Whiteneck G, El-Jaroudi R. Measures of social outcomes in disability research. *Arch Phys Med Rehabil* 2000;81(Suppl 2):S63-80.
16. Patel KK, Veenstra DL, Patrick DL. A review of selected patient-generated outcome measures and their application in clinical trials. *Value Health* 2003;6(5):595-603.
17. Glass TA, De Leon CFM, Bassuk SS et al. Social engagement and depressive symptoms in late life longitudinal findings. *J Aging Health* 2006;18(4):604-28.
18. Pinquart M, Sörensen S. Influences of socioeconomic status, social network, and competence on subjective well-being in later life: a meta-analysis. *Psychol Aging* 2000;15(2):187-224.
19. Stewart BJ, Archbold PG. Nursing intervention studies require outcome measures that are sensitive to change: Part Two. *Res Nurs Health* 1993;16(1):77-81.
20. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10(1):22.
21. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34-42.
22. Heffernan E, Coulson NS, Henshaw H, et al. Understanding the psychosocial experiences of adults with mild-moderate hearing loss: An application of Leventhal's self-regulatory model. *Int J Audiol* 2016; 55(S3):S3-12.
23. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res* 2009;18(9):1263-1278.
24. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009;13(12):1-202.
25. Hawley ML, Litovsky RY, Culling JF. The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *J Acoust Soc Am* 2004;115(2):833-43.
26. Harford E, Barry J. A rehabilitative approach to the problem of unilateral hearing impairment: The contralateral routing of signals (CROS). *J Speech Hear Disord* 1965:30:121-38.
27. Giolas T, Wark D. Communication problems with unilateral hearing loss. *J Speech Hear Disord* 1967;32:336-43.
28. Knappett R. Audiological and psychological consequences of single-sided deafness. *ENT & Audiology News* 2015;24:77-8.
29. Kitterick PT, O'Donoghue GM, Edmondson-Jones M, et al. Comparison of the benefits of cochlear implantation versus contra-lateral routing of signal hearing aids in adult patients with single-sided deafness: study protocol for a prospective within-subject longitudinal trial. *BMC Ear Nose Throat Disord* 2014;14(1):7.

30. Harford E, Dodds E. The clinical application of CROS: A hearing aid for unilateral deafness. *Arch Otolaryngol* 1966;83(5):455-64.
31. Arndt S, Aschendorff A, Laszig R, et al. Comparison of pseudobinaural hearing to real binaural hearing rehabilitation after cochlear implantation in patients with unilateral deafness and tinnitus. *Otol Neurotol* 2011;32(1):39-47.
32. Vermeire K, Van de Heyning P. Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus. *Audiol Neurootol* 2009;14(3):163-71.
33. Buchbinder R, Batterham R, Elsworth G, et al. A validity-driven approach to the understanding of the personal and societal burden of low back pain: development of a conceptual and measurement model. *Arthritis Res Ther* 2011;13(5):R152.
34. Flanagan JC. The critical incident technique. *Psychol Bull* 1954; 51:327.
35. Kitterick PT, Smith SN, Lucas L. Hearing instruments for unilateral severe-to-profound sensorineural hearing loss in adults: a systematic review and meta-analysis. *Ear Hear* 2016;37(5):495.
36. Kitterick PT, Lucas L, Smith SN. Improving health-related quality of life in single-sided deafness: a systematic review and meta-analysis. *Audiol Neurootol* 2015;20(Suppl 1):79-86.
37. Kitterick PT, Lucas L, Smith SN. Systematic review and content validity analysis of patient-reported outcome measures for assessing the effects of hearing instruments in adults with single-sided (unilateral) deafness. PROSPERO 2017:CRD42017056989 [Available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42017056989]
38. Walton MK, Powers JH, Hobart J. et al. Clinical outcome assessments: conceptual foundation—report of the ispor clinical outcomes assessment–emerging good practices for outcomes research task force. *Value Health* 2015;18(6):741–52.
39. Newman CW, Jacobson GP, Spitzer JB. Development of the Tinnitus Handicap Inventory. *Arch Otolaryngol Head Neck Surg* 1996;122(2):143-8.
40. Hiller W, Goebe G. A psychometric study of complaints in chronic tinnitus. *J Psychosom Res* 1992;36(4):337–48.
41. Hallam RS, Jakes SC, Hinchcliffe R. Cognitive variables in tinnitus annoyance. *Br J Clin Psychol* 1988;27(Pt 3):213-22.
42. Londero A, Hall DA. Call for an Evidence-Based Consensus on Outcome Reporting in Tinnitus Intervention Studies. *Frontiers in Medicine Family Medicine and Family Care* 2017;4:42.
43. Langguth B, Goodey R, Azevedo A, et al. Consensus for tinnitus patient assessment and treatment outcome measurement: Tinnitus Research Initiative meeting, Regensburg, July 2006. *Prog Brain Res* 2007;166:525-36.
44. Kuk FK, Tyler RS, Russell D, et al. The psychometric properties of a tinnitus handicap questionnaire. *Ear Hear* 1990;11(6):434-45.
45. Wilson PH, Henry J, Bowen M, et al. Tinnitus reaction questionnaire: psychometric properties of a measure of distress associated with tinnitus. *J Speech Hear Res* 1991;34(1):197-201.
46. Fackrell K, Hall DA, Barry JG et al. Tools for tinnitus measurement: Development and validity of questionnaires to assess handicap and treatment effects. In: Tinnitus: Causes, Treatment and Short & Long-Term Health Effects. F Signorelli and F Turjman (eds). New York:Nova Science Publishers Inc. 2014;13-60.
47. Meikle MB, Henry JA, Griest SE, et al. The tinnitus functional index: development of a new clinical measure for chronic, intrusive tinnitus. *Ear Hear* 2012;33(2):153-76.

Figure 1. Process for evaluating the choice of interventions and outcomes in clinical trials of single-sided deafness (SSD) and assessing the content validty of outcome measures.

Figure 2. Item analysis of five tinnitus-specific questionnaires that have been used in clinical trials as instruments for measuring therapeutic outcomes. Black cells indicate that the questionnaire contains at least one item asking patients about that specific complaint.

REVIEW ARTICLE

How do we know that our patients have benefitted from our ENT/Audiological interventions?

ABSTRACT

This short review article gives an introduction to some of the fundamental concepts and challenges facing measurement in hearing healthcare practice and research. The impact of hearing loss almost always extends beyond the sensory impairment itself, even when the measured degree of audiometric loss is mild. Yet, going beyond audibility, into the realm of measuring impact, takes us into a much more complex and less well-defined space. How does one therefore best measure the therapeutic benefit for evaluating efficacy or for clinical practice audit? Three case studies illustrate approaches to overcome such challenges. Each example highlights the importance of thinking critically about what it is one is seeking trying to measure, rather than selecting a questionnaire instrument based simply on its popularity or accessibility. We conclude by highlighting the important role that clinicians can play in collecting clinical data about their preferred instruments so that we have some evidence to inform decisions about good practice (content validity etc). We would also strongly support open data sharing as we believe that this is one of the best ways to make the most rapid progress the field.

INTRODUCTION

The purpose of this short article is to introduce the reader to some of the fundamental concepts and challenges facing measurement in healthcare practice and research. The concept of measurement will perhaps be most familiar to the reader in the context of the audiogram. The audiogram plots air conduction threshold for tones presented to either ear and is useful for determining hearing sensitivity. Pure-tone averages can be interpreted with respect to standard category boundaries, such as mild hearing loss (26-40 dB A) (1). The impact of hearing loss almost always extends beyond the sensory impairment itself, even when the measured degree of audiometric loss is mild. It is well known that residual hearing is not related in any straightforward way to the burden of disability experienced by a person with hearing loss (2). Going beyond audibility, into the realm of measuring impact, takes us into a much more complex and less well-defined space. For example, mild-to-moderate hearing loss has been reported by patients to interfere with hearing environmental sounds, listening, communicating, speaking, and it can negatively affect family life, social relationships, and ability to work. On a personal level, the negative stigma can affect personal identity, promote a sense of isolation, negative emotions such as frustration, distress and depression. Hearing loss can also increase the effort required for listening and communicating causing fatigue (3). The impact of hearing-related problems, such as tinnitus, similarly spans a wide array of psychological and social dimensions (4).

No gold standard measure Instead of clinician-administered tests, the impacts of hearing loss and tinnitus are often assessed using a patient-reported instrument such as a questionnaire. However, there are no gold standards. This is reflected in the lack of consensus in the selection of questionnaires for hearing studies (4, 5).

Diversity of patient complaints Given the diversity of reported complaints, every patient presents with a complex array of symptoms and functional impacts. Moreover, any clinician or researcher who has worked with people with a hearing-related problem will appreciate that every individual's experience is a very personal one. In practical terms, while one person's primary motivation for seeking medical help might be because their hearing-related problem means that they no longer enjoy socialising with friends

down the pub, for another it may be because their ability to play in an orchestra is hindered, while for another hearing loss might make it difficult watching television comfortably with their spouse. The impact of hearing loss is therefore a construct that is very individualised and personal.

Practical challenges This situation presents the ENT/Audiology professional with two major practical challenges; the first concerns how to comprehensively assess a patient for a precise clinical diagnosis, and the second concerns how to measure the therapeutic benefit for evaluating efficacy or for clinical practice audit. With some degree of success, the challenge for clinical diagnosis has been resolved by creating multi-attribute questionnaire instruments whose scores can be used to discriminate between individuals. For example, the Hearing Handicap Inventory for the Elderly (HHIE) asks 25 questions about the emotional consequences of hearing impairment, social and situational effects (6), with pre-defined cut-offs for determining "no handicap", "mild to moderate handicap" and "significant handicap".

However, the solution to the first challenge tends to be incompatible with evaluating therapeutic benefit. This is because questionnaire items that discriminate well between different patients at the diagnostic appointment are not necessarily sensitive to evaluating changes over time within the same patient (7). And it is difficult to design a questionnaire instrument that is both discriminative *and* evaluative. To illustrate this with an example, tinnitus-related emotional distress and auditory difficulties might both discriminate one patient from another, but only one of these might be responsive to treatment (e.g. hearing aids should reduce auditory difficulties, but might not reduce distress). Averaging the benefit scores for these components could therefore compromise the sensitivity of an aggregated score to measuring treatment-related change. As a general rule, questionnaire instruments that successfully measure therapeutic benefit in different situations tend to be those with good statistical properties that enable the clinician or investigator to interpret specific complaints rather than a global non-specific construct like "severity" or "handicap" (8).

In this short review, we present three case studies which illustrate approaches to overcome the challenges of evaluating therapeutic benefit. These examples highlight the need to think critically about what it is one is seeking trying to measure, rather than selecting a questionnaire instrument based simply on its popularity or accessibility.

Measuring psychosocial functioning of adults with mild-to-moderate hearing loss

The International Classification of Functioning, Disability, and Health (ICF) is a biopsychosocial framework designed to standardise the description, measurement, clinical assessment, and teaching of functioning, disability, and health for researchers, clinicians, clinical educators, and policy-makers around the world (9). The ICF consists of three primary domains of patient burden: (1) physical impairments, or deficits in body functions or body structures, (2) activity limitations, or problems executing tasks and actions, and (3) participation restrictions, or problems with involvement in life situations. These domains are influenced by both environmental factors and personal factors (9, 10). The ICF also includes a comprehensive taxonomy of categories of functioning (e.g. listening, education, self-care). The categories of functioning most relevant to hearing loss have been identified by a large, cross-cultural, mixed-methods study (10). Therefore, the ICF could be used in the future to standardise the measurement of individuals with hearing loss in clinical practice or in research.

The domain of participation restrictions is thought to be the most difficult of the ICF domains to measure (11). One obstacle is that the conceptualisation of participation restrictions is imprecise and inconsistent (12). The WHO (2001) definition above is rather

broad, which means that it is difficult to distinguish participation restrictions from related constructs, such as activity limitations and quality of life (13). Also, there is no consensus regarding the categories of functioning that should be included in a participation restrictions measurement instrument (14). Another obstacle is that different people participate in different ways, depending on their personal preferences and circumstances. It is difficult to capture such a highly individual construct in one standardised tool (13). One solution is to develop different questionnaire instruments for different subgroups (15). However, this can impede comparisons across groups and across studies. Another solution is to create patient-generated measurement tools that permit respondents to personalise their content. However, personalised instruments may not be well suited to the grouping of scores or comparisons across time periods and across individuals. Also, they can be difficult for some respondents to understand and complete (16). Another approach is to obtain counts of social interaction frequency or social network size (17). However, such measures fail to acknowledge that the quality of social contacts can be more important for wellbeing than quantity of social contacts (18).

**** insert Table 1 about here ****

Here, we provide a case study of the utilisation of best practice techniques to develop a hearing-specific measure of participation. Best practice techniques, which include qualitative methodologies (e.g. cognitive interviews) and modern psychometric analysis (e.g. Rasch analysis), are necessary for the creation of gold standard measures. However, to date, these techniques have seldom been employed in the development of hearing-specific measures.

The questionnaire we developed, entitled the Social Participation Restrictions Questionnaire (SPaRQ), was designed to serve as an outcome measure in either research or clinical practice. The SPaRQ consists of a 9-item subscale measuring social behaviours (e.g. difficulties with social interactions) and a 10-item subscale measuring social perceptions (e.g. feelings of isolation). It uses an 11-point response scale (0=completely disagree, 10=completely agree) because a broad range of response options are considered to enhance responsiveness (19). The SPaRQ was designed by conducting a series of qualitative and quantitative studies (see Table 1) in accordance with internationally-recognised guidelines from the questionnaire development literature (20, 21). Our aim was to ensure that the measurement properties of the SPaRQ met the standards required of outcome measures used in clinical practice and in clinical trials (21).

The first step was to create a precise conceptual model of hearing-related participation restrictions and to determine the categories of functioning that should be included in the measure by (1) reviewing the literature, including existing questionnaire instruments and the ICF, and (2) interviewing adults with hearing loss and hearing healthcare professionals (22). The second step was to evaluate the content validity of the SPaRQ, including its relevance, clarity, acceptability, and potential responsiveness, by (1) conducting cognitive interviews with adults with hearing loss and (2) surveying hearing healthcare professionals. Qualitative research with key stakeholders is an often overlooked but essential component of questionnaire development, at it ensures that the instrument adequately captures the respondents‘ experiences, uses everyday language, and is easy to administer and complete (23). The third step was to assess the psychometric properties of the SPaRQ by applying (1) Rasch analysis and (2) traditional (i.e. Classical Test Theory) psychometric analysis to data collected from adults with hearing loss. Whilst most hearing-specific questionnaires have been developed using traditional psychometric analysis alone, a modern approach (i.e. Rasch analysis or Item Response Theory) should also be applied because it enables all the relevant psychometric properties (e.g. unidimensionality, differential item functioning) to be adequately assessed (24). The outcome of this rigorous development process was the production of

a questionnaire that possesses an array of good measurement properties. For instance, each subscale was found to be unidimensional, which means that all of the items within a subscale measure the same construct, and well-targeted, which means that the subscales have high measurement precision and capture a wide range of participation restrictions. There was also evidence to support the convergent validity of the subscales with each one displaying strong, positive correlations with a hearing-specific disability measure and moderate, positive correlations with a generic disability measure and a mental health screening tool. Responsiveness of the SPaRQ is yet to be examined, but this is planned for future research.

One limitation of this research is that it adds another questionnaire to the range of existing questionnaires that are currently being used in hearing research (5). However, it is sometimes necessary to develop new measures in accordance with the latest best practice recommendations in order to address the lack of gold standard measures in the field. In the future, researchers and clinicians would benefit from the introduction of guidelines to help them to identify high quality measures that are appropriate for their purposes.

Relevance of existing questionnaires for assessing burden of single-sided deafness (SSD)

At face value, single-sided deafness (SSD) would appear to be a form of hearing loss where the task of determining whether or not a patient has benefitted from an intervention should be relatively straightforward. Lack of hearing on one side of the head would be expected to hinder access to acoustic information in that hemifield and disrupt the ability to segregate information from different sources (25). One might also be tempted to assume that relevant interventions for this patient group are those that address these impaired listening skills, and benefit should be measured in terms of the extent to which they have restored or improved such skills. However, some of the earliest published observations about these patients remarked on the unexpected degree of burden that impairments to these listening skills impose on the patient. Harford and Barry noted "the persistence and earnestness of reports from unilaterally hearing-impaired individuals stating serious difficulty encountered in many common listening situations" (26). Early work also suggested a breadth and depth of burden that one might not predict from these functional difficulties. Giolas and Wark noted that a majority of patients reported strong negative emotions that included embarrassment and helplessness (27). The extent of these feelings was such that they recommended they should be addressed actively as part of their clinical management, an approach that is still recommended almost 50 years later (28).

The incongruence between the fact that SSD patients still have access to one 'good' hearing ear and the chronic and complex burden that they report is perhaps why there is an increasing focus on the surgical restoration of hearing in their deaf ear (29) rather than traditional interventions that re-route sound between the ears (30). Early-phase trials have suggested that cochlear implantation is capable of restoring bilateral input and addressing, at least in part, the functional impairments of SSD (31, 32). However, as the field moves beyond demonstrations of clinical efficacy in the form that can be measured using controlled listening tests in the clinic or laboratory, increasing emphasis will inevitably be placed on conducting trials to measure broader impacts on quality of life to demonstrate the additional benefits to health it provides over currently available treatments.

In designing these trials, one must first ask whether the intervention addresses one or more aspects of burden that are relevant to SSD patients, and what specific aspects of burden are being targeted. Such knowledge would ideally be supported by evidence from

early-phase trials so that the mechanism through which the intervention works is well understood. The choice of outcomes that are being measured would also need to be examined to ask whether they are considered by patients to be important for their health and wellbeing. Finally, outcome measures should be selected based evidence for their validity to measure those outcomes in these patients. Here we describe a research process that has been designed to address these questions in the field of SSD and to lay the groundwork for the development of a Core Outcome Set (Figure 1).

**** insert Figure 1 about here ****

Fundamental to addressing many of these issues is a comprehensive understanding of the health condition itself. Little if any qualitative work around the burden imposed by SSD has been conducted since Giolas and Wark applied the Critical Incident Technique to study the functional consequences of SSD (27). This technique structures the interview around events that the patient recognises were affected by their hearing loss. Patient interviews were therefore conducted using a similar methodology to construct a hierarchical model of burden (33) based on patient-reported incidents and emerging themes from the transcripts. This qualitative approach provided a comprehensive characterisation of the impact of the health condition (34) and was initially used to assess whether interventions targeted aspects of health that are impaired by SSD. A systematic review identified those interventions and concluded that studies have focussed almost exclusively on intervening to improve functional impairments to speech perception and spatial hearing (35). However, the wide range and inconsistent use of patient-reported questionnaire instruments as outcome measures in existing trials meant that there is considerable uncertainty over what outcomes if any beyond the direct functional impairments to hearing were being targeted by these interventions (36). To address this uncertainty, a second systematic review is underway to identify what studies say they are trying to measure and to map those outcomes onto their use of specific measurement instruments (37). The content of the questionnaire instruments will be compared with the model of patient burden to assess whether they are targeting domains of health which are considered relevant by this patient group (23). The analysis will examine how successful these instruments are at targeting specific domains of health and therefore their suitability for use as outcome measures in the context of clinical trials (38).

Relevance of existing questionnaires for assessing benefits of tinnitus treatments

There is a substantial literature concerning self-assessment questionnaires for scaling the negative impacts of tinnitus. This literature shows that many different tinnitus-specific questionnaires have been used to assess treatment-related changes in tinnitus. For example, our review of clinical trials from 2006 to 2015 identified at least 78 different outcome instruments used in 228 trials; with 24 of those being different tinnitus-specific questionnaires. These were predominantly the Tinnitus Handicap Inventory (THI) (39) and the Tinnitus Questionnaire (TQ) (4, 40, 41). But even these two most popular instruments were used in only a minority of clinical research since we noted that usage was 15% and 7% out of 228 studies, for the THI and TQ respectively. We also note that these questionnaire instruments have predominantly been designed for screening and diagnostic purposes, not for measuring benefit from ENT/Audiological interventions. In particular, they measure multiple domains of patient burden.

The tinnitus community widely acknowledges that a standard is needed to ensure that therapeutic benefit is measured much more consistently across studies, and that benefit is quantified using a measurement instrument that is fit for the purpose of outcome measurement (e.g. 42). A first attempt by Langguth and 28 other colleagues in 2006 sought to develop a set of international recommendations on choice of instruments for

assessing the outcome from an intervention for tinnitus (43). The recommendations by this working group suggested four questionnaires; namely the Tinnitus Handicap Inventory (THI) (39), the Tinnitus Handicap Questionnaire (THQ) (44), the Tinnitus Questionnaire (TQ) (41) and the Tinnitus Reaction Questionnaire (TRQ) (45). These instruments were developed in diverse patient populations across the USA, UK, and Australia, but were not all developed for the same applications. In particular, while the THI, TQ and TRQ focus on aspects of psychological distress, the THQ was created to comprehensively measure the perceived degree of broad handicaps attributed to tinnitus (see Fackrell et al. (46) for a review). Nevertheless, they were chosen as they were the most widely used at the time, and had been translated for use in different languages and cultures. Their questions also broadly span the emotional impact of tinnitus, disability and handicap.

**** insert Figure 2 about here ****

In making their interim recommendation, Langguth and colleagues commented that the THI, THQ, TQ and TRQ also share a common feature in that they attempt to quantify a combination of tinnitus-related distress, disability and handicap resulting in a large overlap of their items (43). Conceptual similarity is supported by statistical evidence for a high convergent validity between the global scores. For example, pairwise correlations between the THI, THQ, TQ and TRQ range from 0.74 to 0.89 (see Fackrell et al. (46) for a review). To explore conceptual equivalence in more detail we have conducted a fine-grained evaluation of each individual questionnaire item to specify exactly what health concepts form the ingredients of each instrument. The findings from this evaluation are illustrated in Figure 2. The black cells indicate where the instrument contains at least one item that we judge to be asking about the corresponding tinnitus-related complaint. All questionnaire instruments contain items that ask about a diverse range of tinnitus-related complaints covering all the major high-level categories of impact on everyday life, such as emotional impacts or activities and relationships. However, the patchwork highlights clear differences between instruments in terms of their specific item-level content. Some of these detailed differences could be clinically important for some individuals with critical gaps where an instrument entirely misses out questions on a particular type of complaint. For example, the impact of tinnitus on physical health is explored only in the TQ ('bodily complaints') and the THQ ('ill health'). We have not yet compared the content of the instruments with available information about patient burden to assess whether they are targeting domains of health which are considered relevant by people with tinnitus (23). This analysis is planned. It will tell us how successful these instruments might be at targeting specific domains of health and therefore their suitability for use as outcome measures in the context of clinical trials of tinnitus, especially under certain circumstances (e.g. with a specific patient subtype, or for evaluating the outcome from a specific intervention).

Langguth et al. (43) appreciated some of these limitations with the THI, THQ, TQ and TRQ and so the working group agreed that in the future, a "better" questionnaire was required. Since that time, a multi-item tinnitus questionnaire has been developed in the USA using a method to select items that optimized the overall responsiveness of the outcome score to treatment-related change (47). The resulting Tinnitus Functional Index (TFI) asks 25 questions about the intrusive of tinnitus, reduced sense of control, reduced quality of life, sleep disturbance, auditory difficulties, cognitive interference, interference with relaxation, and emotional distress, with pre-defined cut-offs for determining "not a problem", "small problem", "moderate problem", "big problem", and "very big problem". When opting to use the TFI in other countries and cultures, it would be advisable to explore the content validity and severity grading in the new target population.

DISCUSSION

These three examples illustrate different approaches to overcome the challenges of evaluating therapeutic benefit. In common, they all highlight the need to think critically about what it is one is seeking trying to measure. We end our review with some concluding remarks:

- We have previously argued that it would be helpful to step away from using terms such as 'handicap' or 'severity' when naming a questionnaire instrument. These terms are not helpful to clinicians and researchers because they do not meaningfully describe exactly what health-related construct is being measured by the instrument (4). The development of the SPaRQ by Heffernan et al. provides a good example where the questionnaire name describes exactly what aspect of health the instrument claims to measure (22).

- Although often questionnaire developers typically present psychometric validations of a questionnaire instrument, the word 'validation' is quite emotive. Validity is not a fixed property, but varies across populations and cultures. Its good practice therefore to keep an open mind and to evaluate any questionnaire instrument the first time its going to be used for a particular purpose and in a particular patient population.

- At the end of the questionnaire evaluation, we might end up by failing to find any instruments which meet stringent contemporary standards of performance for outcome measures in clinical trials of SSD and tinnitus. What then? Clearly new research will be needed to modify an existing instrument, or create a new one from scratch. But what should we do in the meantime? Well, just because an instrument is not perfect does not necessarily mean that it should not be used. In this situation, clinicians can play an important role by collecting clinical data about their preferred instruments so that we have some evidence to inform decisions about good practice (content validity etc). We would also strongly support open data sharing as we believe that this is one of the best ways to make the most rapid progress the field.

REFERENCES

1. World Health Organization. Accessed June 1, 2017. [http://www.who.int/pbd/deafness/hearing_impairment_grades/en/]
2. Granberg S, Pronk M, Swanepoel DW, et al. The ICF core sets for hearing loss project: Functioning and disability from the patient perspective. *Int J Audiol* 2014;53:777-86.
3. Pichora-Fuller MK, Kramer SE, Eckert MA, et al. Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear* 2016;37(Suppl 1):5-27S.
4. Hall DA, Haider H, Szczepek AJ, et al. Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults. *Trials* 2016;17(1):270.
5. Granberg S, Dahlström J, Möller C, et al. The ICF Core Sets for hearing loss - researcher perspective. Part I : Systematic review of outcome measures identified in audiological research. International Journal of Audiology 2014;53(2):65-76.
6. Ventry IM, Weinstein BE. The hearing handicap inventory for the elderly: a new tool. *Ear Hear* 1982;3(3):128-34.
7. Guyatt G, Kirshner B, Jaeschke R: Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992;45(12):1341–5.
8. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials* 2016;17(1):449.

9. World Health Organization. The International Classification of Functioning, Disability and Health (ICF). 2001; Geneva, WHO. Accessed June 1, 2017. [http://www.who.int/classifications/icf/en/].
10. Danermark B, Granberg S, Kramer SE, et al. The creation of a comprehensive and a brief core set for hearing loss using the International Classification of Functioning, Disability and Health. *Am J Audiol* 2013;22(2):323-8.
11. Whiteneck G, Dijkers MP. Difficult to measure constructs: conceptual and methodological issues concerning participation and environmental factors. *Arch Phys Med Rehabil* 2009;90(Suppl 11):S22-35.
12. Heinemann AW, Tulsky D, Dijkers M et al. Issues in participation measurement in research and clinical applications. *Arch Phys Med Rehabil* 2010;91(9):S72-6.
13. Dijkers M. Issues in the conceptualization and measurement of participation: an overview. *Arch Phys Med Rehabil* 2010;91(Suppl 9):S5-16.
14. Eyssen IC, Steultjens MP, Dekker J et al. A systematic review of instruments assessing participation: challenges in defining participation. *Arch Phys Med Rehabil* 2011;92(6):983-97.
15. Dijkers M, Whiteneck G, El-Jaroudi R. Measures of social outcomes in disability research. *Arch Phys Med Rehabil* 2000;81(Suppl 2):S63-80.
16. Patel KK, Veenstra DL, Patrick DL. A review of selected patient-generated outcome measures and their application in clinical trials. *Value Health* 2003;6(5):595-603.
17. Glass TA, De Leon CFM, Bassuk SS et al. Social engagement and depressive symptoms in late life longitudinal findings. *J Aging Health* 2006;18(4):604-28.
18. Pinquart M, Sörensen S. Influences of socioeconomic status, social network, and competence on subjective well-being in later life: a meta-analysis. *Psychol Aging* 2000;15(2):187-224.
19. Stewart BJ, Archbold PG. Nursing intervention studies require outcome measures that are sensitive to change: Part Two. *Res Nurs Health* 1993;16(1):77-81.
20. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10(1):22.
21. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34-42.
22. Heffernan E, Coulson NS, Henshaw H, et al. Understanding the psychosocial experiences of adults with mild-moderate hearing loss: An application of Leventhal's self-regulatory model. *Int J Audiol* 2016; 55(S3):S3-12.
23. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res* 2009;18(9):1263-1278.
24. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009;13(12):1-202.
25. Hawley ML, Litovsky RY, Culling JF. The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *J Acoust Soc Am* 2004;115(2):833-43.
26. Harford E, Barry J. A rehabilitative approach to the problem of unilateral hearing impairment: The contralateral routing of signals (CROS). *J Speech Hear Disord* 1965:30:121-38.
27. Giolas T, Wark D. Communication problems with unilateral hearing loss. *J Speech Hear Disord* 1967;32:336-43.
28. Knappett R. Audiological and psychological consequences of single-sided deafness. *ENT & Audiology News* 2015;24:77-8.
29. Kitterick PT, O'Donoghue GM, Edmondson-Jones M, et al. Comparison of the benefits of cochlear implantation versus contra-lateral routing of signal hearing aids in adult patients with single-sided deafness: study protocol for a prospective within-subject longitudinal trial. *BMC Ear Nose Throat Disord* 2014;14(1):7.

30. Harford E, Dodds E. The clinical application of CROS: A hearing aid for unilateral deafness. *Arch Otolaryngol* 1966;83(5):455-64.
31. Arndt S, Aschendorff A, Laszig R, et al. Comparison of pseudobinaural hearing to real binaural hearing rehabilitation after cochlear implantation in patients with unilateral deafness and tinnitus. *Otol Neurotol* 2011;32(1):39-47.
32. Vermeire K, Van de Heyning P. Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus. *Audiol Neurootol* 2009;14(3):163-71.
33. Buchbinder R, Batterham R, Elsworth G, et al. A validity-driven approach to the understanding of the personal and societal burden of low back pain: development of a conceptual and measurement model. *Arthritis Res Ther* 2011;13(5):R152.
34. Flanagan JC. The critical incident technique. *Psychol Bull* 1954; 51:327.
35. Kitterick PT, Smith SN, Lucas L. Hearing instruments for unilateral severe-to-profound sensorineural hearing loss in adults: a systematic review and meta-analysis. *Ear Hear* 2016;37(5):495.
36. Kitterick PT, Lucas L, Smith SN. Improving health-related quality of life in single-sided deafness: a systematic review and meta-analysis. *Audiol Neurootol* 2015;20(Suppl 1):79-86.
37. Kitterick PT, Lucas L, Smith SN. Systematic review and content validity analysis of patient-reported outcome measures for assessing the effects of hearing instruments in adults with single-sided (unilateral) deafness. PROSPERO 2017:CRD42017056989 [Available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42017056989]
38. Walton MK, Powers JH, Hobart J. et al. Clinical outcome assessments: conceptual foundation—report of the ispor clinical outcomes assessment–emerging good practices for outcomes research task force. *Value Health* 2015;18(6):741–52.
39. Newman CW, Jacobson GP, Spitzer JB. Development of the Tinnitus Handicap Inventory. *Arch Otolaryngol Head Neck Surg* 1996;122(2):143-8.
40. Hiller W, Goebe G. A psychometric study of complaints in chronic tinnitus. *J Psychosom Res* 1992;36(4):337–48.
41. Hallam RS, Jakes SC, Hinchcliffe R. Cognitive variables in tinnitus annoyance. *Br J Clin Psychol* 1988;27(Pt 3):213-22.
42. Londero A, Hall DA. Call for an Evidence-Based Consensus on Outcome Reporting in Tinnitus Intervention Studies. *Frontiers in Medicine Family Medicine and Family Care* 2017;4:42.
43. Langguth B, Goodey R, Azevedo A, et al. Consensus for tinnitus patient assessment and treatment outcome measurement: Tinnitus Research Initiative meeting, Regensburg, July 2006. *Prog Brain Res* 2007;166:525-36.
44. Kuk FK, Tyler RS, Russell D, et al. The psychometric properties of a tinnitus handicap questionnaire. *Ear Hear* 1990;11(6):434-45.
45. Wilson PH, Henry J, Bowen M, et al. Tinnitus reaction questionnaire: psychometric properties of a measure of distress associated with tinnitus. *J Speech Hear Res* 1991;34(1):197-201.
46. Fackrell K, Hall DA, Barry JG et al. Tools for tinnitus measurement: Development and validity of questionnaires to assess handicap and treatment effects. In: Tinnitus: Causes, Treatment and Short & Long-Term Health Effects. F Signorelli and F Turjman (eds). New York:Nova Science Publishers Inc. 2014;13-60.
47. Meikle MB, Henry JA, Griest SE, et al. The tinnitus functional index: development of a new clinical measure for chronic, intrusive tinnitus. *Ear Hear* 2012;33(2):153-76.

Figure 1. Process for evaluating the choice of interventions and outcomes in clinical trials of single-sided deafness (SSD) and assessing the content validty of outcome measures.

Figure 2. Item analysis of five tinnitus-specific questionnaires that have been used in clinical trials as instruments for measuring therapeutic outcomes. Black cells indicate that the questionnaire contains at least one item asking patients about that specific complaint.

Figure 1

Click here to download Figure (TIFF, PDF, Word Doc, PPT, or EPS files are acceptable) Figure 1.tif

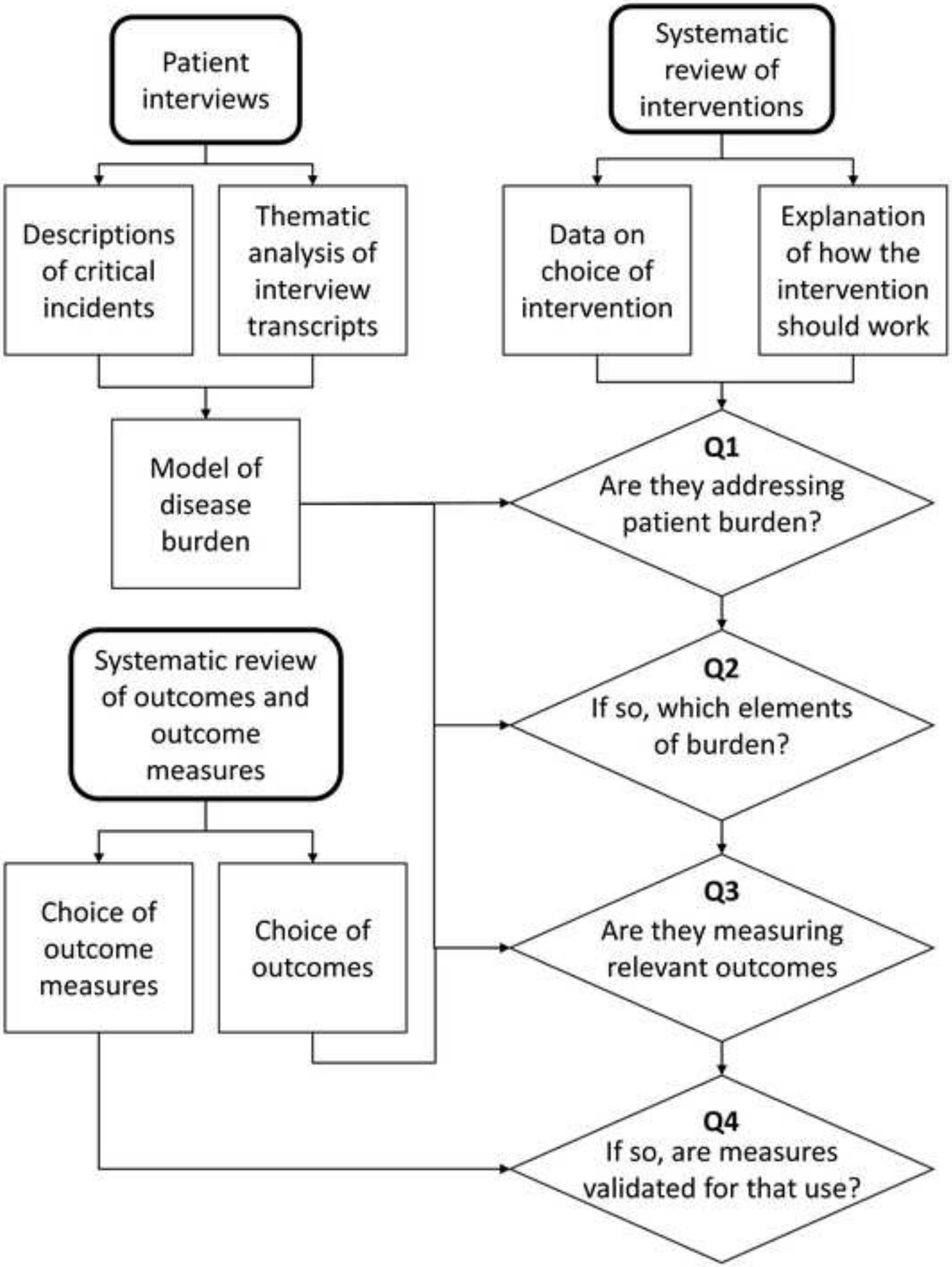

Figure 2

Click here to download Figure (TIFF, PDF, Word Doc, PPT, or EPS files are acceptable) Figure 2.tif

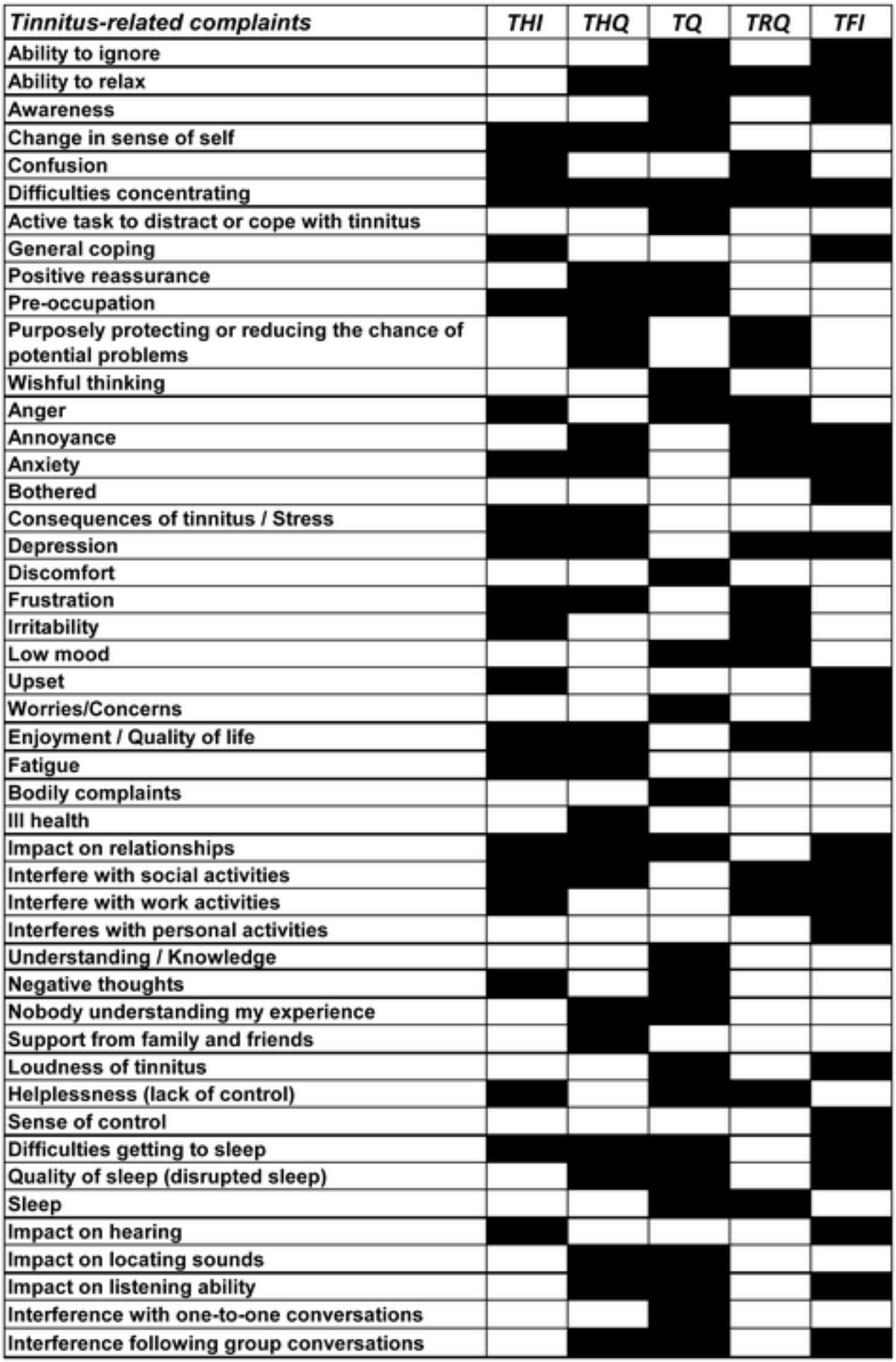

Table 1. Development of the Social Participation Restrictions Questionnaire (SPaRQ)

Study	Main Aim	Method	Data Analysis
1	Conceptualise participation restrictions in adults with hearing loss Generate content for the first SPaRQ prototype	Semi-structured interviews with 25 adults with hearing loss and 9 hearing healthcare professionals	Deductive thematic analysis
2	Evaluate the content validity of the first SPaRQ prototype	Cognitive interviews 14 adults with hearing loss Online survey of 20 hearing healthcare professionals	Deductive analysis using a taxonomy of respondent problems Descriptive statistics
3	Assess the psychometric properties of the second SPaRQ prototype	Questionnaire completed by 279 adults with hearing loss	Rasch analysis
4	Assess the psychometric properties of the finalised SPaRQ	Questionnaire completed by 102 adults with hearing loss	Traditional psychometric analysis