The Tinnitus Functional Index: Development of a New Clinical Measure for Chronic, Intrusive Tinnitus

Mary B. Meikle,[1] James A. Henry,[1,2] Susan E. Griest,[1,2] Barbara J. Stewart,[1] Harvey B. Abrams,[3] Rachel McArdle,[3] Paula J. Myers,[4] Craig W. Newman,[5] Sharon Sandridge,[5] Dennis C. Turk,[6] Robert L. Folmer,[1,2] Eric J. Frederick,[7] John W. House,[8] Gary P. Jacobson,[9] Sam E. Kinney,[5] William H. Martin,[1] Stephen M. Nagler,[10] Gloria E. Reich,[1] Grant Searchfield,[11] Robert Sweetow,[12] and Jack A. Vernon[1]

Editor's Note: The first author of this article, Dr. Mary B. Meikle, passed away on February 5, 2011. Her more than 40-year career in hearing research focused specifically on the diagnosis and clinical care of patients with tinnitus. This publication, presented as a collaborative research effort with coauthors from across the United States and from New Zealand, proposes a new tool for establishing a baseline measurement of tinnitus and its treatment outcomes. It is Dr. Meikle's final scientific publication.

Objectives: Chronic subjective tinnitus is a prevalent condition that causes significant distress to millions of Americans. Effective tinnitus treatments are urgently needed, but evaluating them is hampered by the lack of standardized measures that are validated for both intake assessment and evaluation of treatment outcomes. This work was designed to develop a new self-report questionnaire, the Tinnitus Functional Index (TFI), that would have documented validity both for scaling the severity and negative impact of tinnitus for use in intake assessment and for measuring treatment-related changes in tinnitus (responsiveness) and that would provide comprehensive coverage of multiple tinnitus severity domains.

Design: To use preexisting knowledge concerning tinnitus-related problems, an Item Selection Panel (17 expert judges) surveyed the content (175 items) of nine widely used tinnitus questionnaires. From those items, the Panel identified 13 separate domains of tinnitus distress and selected 70 items most likely to be responsive to treatment effects. Eliminating redundant items while retaining good content validity and adding new items to achieve the recommended minimum of 3 to 4 items per domain yielded 43 items, which were then used for constructing TFI Prototype 1.

Prototype 1 was tested at five clinics. The 326 participants included consecutive patients receiving tinnitus treatment who provided informed consent—constituting a convenience sample. Construct validity of Prototype 1 as an outcome measure was evaluated by measuring responsiveness of the overall scale and its individual items at 3 and 6 mo follow-up with 65 and 42 participants, respectively. Using a predetermined list of criteria, the 30 best-functioning items were selected for constructing TFI Prototype 2.

Prototype 2 was tested at four clinics with 347 participants, including 155 and 86 who provided 3 and 6 mo follow-up data, respectively. Analyses were the same as for Prototype 1. Results were used to select the 25 best-functioning items for the final TFI.

Results: Both prototypes and the final TFI displayed strong measurement properties, with few missing data, high validity for scaling of tinnitus severity, and good reliability. All TFI versions exhibited the same eight factors characterizing tinnitus severity and negative impact. Responsiveness, evaluated by computing effect sizes for responses at follow-up, was satisfactory in all TFI versions.

In the final TFI, Cronbach's alpha was 0.97 and test–retest reliability 0.78. Convergent validity ($r = 0.86$ with Tinnitus Handicap Inventory [THI]; $r = 0.75$ with Visual Analog Scale [VAS]) and discriminant validity ($r = 0.56$ with Beck Depression Inventory-Primary Care [BDI-PC]) were good. The final TFI was successful at detecting improvement from the initial clinic visit to 3 mo with moderate to large effect sizes and from initial to 6 mo with large effect sizes. Effect sizes for the TFI were generally larger than those obtained for the VAS and THI. After careful evaluation, a 13-point reduction was considered a preliminary criterion for meaningful reduction in TFI outcome scores.

Conclusions: The TFI should be useful in both clinical and research settings because of its responsiveness to treatment-related change, validity for scaling the overall severity of tinnitus, and comprehensive coverage of multiple domains of tinnitus severity.

(Ear & Hearing 2011;32;1–●)

INTRODUCTION

Chronic subjective tinnitus (ringing or other sounds audible only to the affected individual) is a prevalent condition affecting millions of Americans, many of whom experience significant distress as a result (Hoffman & Reed 2004). The disabling effects of severe tinnitus resemble many effects associated with chronic pain (Moller 2007), typically including sleep interference; cognitive difficulties (particularly with concentration); difficulties at work, at home, and in social relationships; and negative emotional reactions including anxiety, frustration, anger, and depression (Tyler & Baker 1983; Stouffer & Tyler 1990; Axelsson 1992; Meikle 1992; Dobie 2004b).

Despite many efforts to provide relief for tinnitus over the past three decades, there is little agreement concerning the relative merits of the various treatments used (Noble & Tyler 2007). Evaluating the efficacy of tinnitus treatments is hampered by the lack of standardized outcome measures (Axelsson 1992; Meikle 1992; Dobie 2004a; Kamalski et al. 2010). Clinical trials of proposed treatment efforts would benefit greatly from standardization of tinnitus measures. Standardization would improve comparability of treatment effects between different treatment

[1]Oregon Health & Science University, Portland, Oregon; [2]VA National Center for Rehabilitative Auditory Research, Portland, Oregon; [3]Bay Pines VA Healthcare System, Bay Pines, Florida; [4]James A. Haley Veterans' Hospital, Tampa, Florida; [5]Cleveland Clinic, Cleveland, Ohio; [6]University of Washington, Seattle, Washington; [7]Balance and Hearing Center Northwest, Portland, Oregon; [8]The House Ear Institute, Los Angeles, California; [9]Vanderbilt University, Nashville, Tennessee; [10]Atlanta Medical Consultants, Atlanta, Georgia; [11]University of Auckland, Auckland, New Zealand; and [12]University of California at San Francisco, San Francisco, California.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

0196/0202/11/3205-0001/0 • Ear & Hearing • Copyright © 2011 by Lippincott Williams & Wilkins • Printed in the U.S.A.

TABLE 1. Nine widely used tinnitus questionnaires and the corresponding clinical samples on which they were based

Authors and Year of Publication	Questionnaire Title	No. of Patients Providing Data
Coles et al. (1992)	Tinnitus Severity Grading	1121
Halford & Anderson (1991)	Subjective Tinnitus Severity Scale	142
Hallam et al. (1988)	Tinnitus Questionnaire	179
Jastreboff & Jastreboff (1999)	Tinnitus Retraining Therapy Initial Interview	*
Kuk et al. (1990)	Tinnitus Handicap Questionnaire	375
Meikle et al. (1995)	Tinnitus Severity Index	3119
Newman et al. (1996)	Tinnitus Handicap Inventory	150
Sweetow & Levy (1990)	Tinnitus Severity Scale	24
Wilson et al. (1991)	Tinnitus Reaction Questionnaire	156
Total		5266

To our knowledge, information on patient base has not been published.

centers, facilitate meta-analyses involving multiple treatment studies, and provide a more consistent basis for recruiting comparable patients into different treatment groups (Meikle & Griest 2002; Turk 2002; Newman & Sandridge 2004; Meikle et al. 2007).

The subjective nature of tinnitus necessitates the use of self-report questionnaires for characterizing patients' status at intake, and numerous questionnaires have been developed for scaling the negative impact (severity) of tinnitus (see reviews by Meikle 1992; Tyler 1993; Newman & Sandridge 2004; Meikle et al. 2008). At least nine English-language questionnaires are widely known and used, constituting a valuable resource concerning the major negative impacts of tinnitus. Table 1 lists the nine questionnaires and the sample size used in each. These questionnaires were developed between 1988 and 1999, collectively using data from more than 5000 tinnitus patients.

Although the existing questionnaires have proven useful for measuring individual differences regarding tinnitus severity, an important limitation is that none were designed to maximize responsiveness (i.e., sensitivity for measuring treatment-related changes in tinnitus). In a recent review of tinnitus-specific health-related quality of life instruments used to assess treatment outcomes in clinical trials, Kamalski et al. (2010, p. 181) concluded that the "instruments currently used in tinnitus trials appear not to be validated to measure effectiveness of interventions" and that their responsiveness is not known.

It is likely that responsiveness was not explicitly evaluated in the tinnitus questionnaires listed in Table 1 because familiarity with the concept of responsiveness was not widespread in the 1980s and 1990s. The few early articles describing methods to improve responsiveness often were published in journals that might not have been read by tinnitus researchers (e.g., Guyatt 1988).* Now, more than 20 years since the first tinnitus questionnaires were developed, there is extensive research literature on responsiveness and measurement sensitivity for intervention studies.

Development of responsive outcome measures should above all emphasize content validity (Nunnally 1978), identifying those aspects of an attribute that are likely to undergo clinically important change as a result of the treatment (Kirshner & Guyatt 1985). Outcome measures that pertain to a specific disease or condition exhibit greater responsiveness than those that measure generic quality of life (Wiebe et al. 2003). Such results highlight the benefits of extensive content validity evaluation and argue for including items that capture in detail the negative impact of tinnitus.

Evidence of an outcome measure's responsiveness commonly involves computing the effect size detected by the measure in an intervention trial (Lipsey 1990). An effect size quantifies observed treatment effects in terms of SD units of the item or scale in question (Cohen 1988). Maximizing effect sizes of outcome measures can increase the statistical power of clinical trials (Lipsey 1990; Stewart & Archbold 1992, 1993). Measures with inadequate effect size not only require larger sample sizes to detect significant treatment effects (thus increasing research costs) but also result in failure to detect real treatment effects even when they have occurred. For that reason, the present study was undertaken to develop a tinnitus questionnaire designed specifically to maximize effect sizes in response to treatment-related improvements in tinnitus.

Responsive outcome measures are also characterized by a potential distribution of scores that will allow detection of change, for example, through the use of fine-grained measurement intervals and avoidance of items with potential floor or ceiling effects (Lipsey 1990). With respect to reliability, measures with high reliability may work well for intervention trials, but low internal consistency or low test–retest reliability may not be a problem (Nunnally 1978). For example, Carver (1974) described an ideal outcome measure as one where the pretreatment scores are all at the "poor" end of the score range and posttreatment scores are all at the "good" end of the score range. Both pre- and posttreatment scores exhibit small SDs, reflecting a restriction of range that leads to low internal consistency or test–retest reliability at pretreatment and at posttreatment, and yet such a measure is highly responsive to treatment effects. Thus, high reliability, by itself, is not a convincing evidence that a measure will be responsive in

*Starting in the 1980s, researchers began to emphasize the importance of measurement sensitivity and responsiveness for program evaluation and health-related interventions (e.g., Kirshner & Guyatt 1985; Lipsey 1983). Guyatt et al. (1987) began to report on responsiveness in measures of health-related quality of life. Lipsey's much-cited 1990 book, *Design Sensitivity*, provided some guidance on selecting measures for intervention studies that would be sensitive to change (Lipsey 1990). Despite this progress, Lipsey and Cordray (2000) reported in their review article that "the characteristics that make a measure sensitive to individual differences on a construct of interest are not necessarily the ones that make them sensitive to change on that construct over time … Although there is general recognition in the field that outcome measures must be sensitive to change, there has been surprisingly little systematic analysis of the ways sensitive measures can be identified and how sensitivity can be enhanced" (pp. 355–356).

detecting treatment effects (Stewart & Archbold 1993; Puhan et al. 2005).

In summary, important advances in measurement science regarding the development of responsive outcome measures for intervention trials have made it obligatory to apply such methods to patient-reported outcomes for tinnitus treatment. Accordingly, the new Tinnitus Functional Index (TFI) was designed to have strong content validity with fine-grained measurement intervals while avoiding floor and ceiling effects to maximize its effect sizes.

The goal of this research was to develop a new self-report questionnaire, the TFI, to measure the construct defined as the severity and negative impact of tinnitus. One intended purpose of the TFI was for use as an outcome measure with documented responsiveness for assessing treatment-related changes in tinnitus. A second intended purpose of the TFI was as a measure of individual differences for intake evaluation. A third intended purpose of the TFI was to provide reliable and valid measurement of the multiple domains of tinnitus severity.

OVERVIEW OF STUDY

The approach used to develop the TFI used observational rather than experimental methodology. It is important to emphasize that this project was not intended as an evaluation of any particular tinnitus treatment and therefore did not involve both treated and untreated groups of participants. Instead, because its aim was to determine which among a number of questionnaire items performed best in evaluating treatment-related improvements in tinnitus, the work was conducted with a large group of participants all of whom were seeking clinical interventions for their tinnitus. At 3 and 6 mo follow-up, comparisons using the new TFI were made between those who reported their tinnitus as improved versus those who reported their tinnitus as unchanged or worse. Figure 1 shows a flowchart summarizing the development of the TFI.

This study used the general model used to develop the Pain Outcomes Questionnaire-VA (Clark et al. 2003), which had been evaluated at six Veterans Affairs (VA) pain centers to obtain sufficient numbers of cases for valid statistical evaluation. In that study, no specific treatment protocol was required. Instead, each center provided treatment according to its usual standards of care. Combining observations from separate clinics also increased the diversity of participant groups, strengthening the generalizability of results.

In the present study, a similar multisite approach was used, with two data collection sites in Florida, two in Oregon, and one in Ohio. In addition, to minimize measurement bias that might be introduced if investigators from only one clinic selected the questionnaire content, input was obtained from the 21 investigators listed as authors, who jointly represented 10 tinnitus treatment centers and 1 pain treatment center. The Oregon Health & Science University (OHSU) served as the administrative site.

Timetable

The present research was conducted over the period July 2004 through June 2008, using an iterative process to develop and evaluate successively smaller questionnaire versions. For each version, evidence was obtained for the measurement properties of responsiveness (effect sizes), internal structure (factor analysis), and ability to scale overall tinnitus severity for intake assessment (internal consistency, test–retest reliability, and convergent and discriminant validity). The process was completed in three stages:

- Stage 1 (6 mo): The focus was on content validity evaluation and selection of 43 items for TFI Prototype 1.
- Stage 2 (18 mo): TFI Prototype 1 underwent testing with 326 tinnitus patients at clinical sites. Results regarding measurement properties were used to select the 30 best-performing items for TFI Prototype 2.
- Stage 3 (24 mo): TFI Prototype 2 was tested with 347 patients and results were used to select the 25 best-performing items for the final version of the TFI.

Summary of Major Design Goals

Throughout this work, emphasis was consistently placed on selecting questionnaire items that individually exhibited high sensitivity to change while jointly providing comprehensive coverage of the major negative functional impacts from tinnitus plus high construct validity for scaling of tinnitus severity. To minimize respondent and examiner burden, the final version of the TFI was designed to be as brief as possible while satisfying these goals. Table 2 summarizes the major considerations taken into account for designing the item content of the TFI.

STAGE 1: ITEM SELECTION AND DESIGN OF TFI PROTOTYPE 1

Methods

Item Selection Panel • To maximize construct validity, selection of items for the new questionnaire followed recommendations of Haynes et al. (1995) to use multiple judges of content validity and to quantify their judgments using formalized scaling procedures. The Item Selection Panel included 17 of the 21 authors of this article, distributed across tinnitus treatment centers located in Oregon, California, Ohio, Tennessee, Georgia, Florida, and New Zealand. These 17 Panel judges included audiologists, otologists, hearing scientists, and other health researchers. Fifteen of them had substantial direct experience with tinnitus patients and two had more than 25 yr of experience each with patient reports about their tinnitus. Nine Panel judges also had previous experience in development of tinnitus measures.

Item Selection and Content Validity • A website was created to solicit Panel judges' evaluations concerning the qualifications of each of 175 items contained in the nine previously published tinnitus questionnaires listed in Table 1. On a separate web page for each item, each judge made two ratings: (1) domain identification (content relevance) for that item and (2) its expected responsiveness to treatment-related improvement. Supplemental Digital Content 1, Figure A1 (http://links.lww.com/EANDH/A53), Supplemental Digital Content 2, Figure A2 (http://links.lww.com/EANDH/A54), and Supplemental Digital Content 3, Figure A3 (http://links.lww.com/EANDH/A55) show an example of one of the 175 items and the format for judges to rate relevant tinnitus domains and expected item responsiveness. The Panel judges could access the various web pages in any desired order and, if they wished, review and correct their previous responses. Each judge's

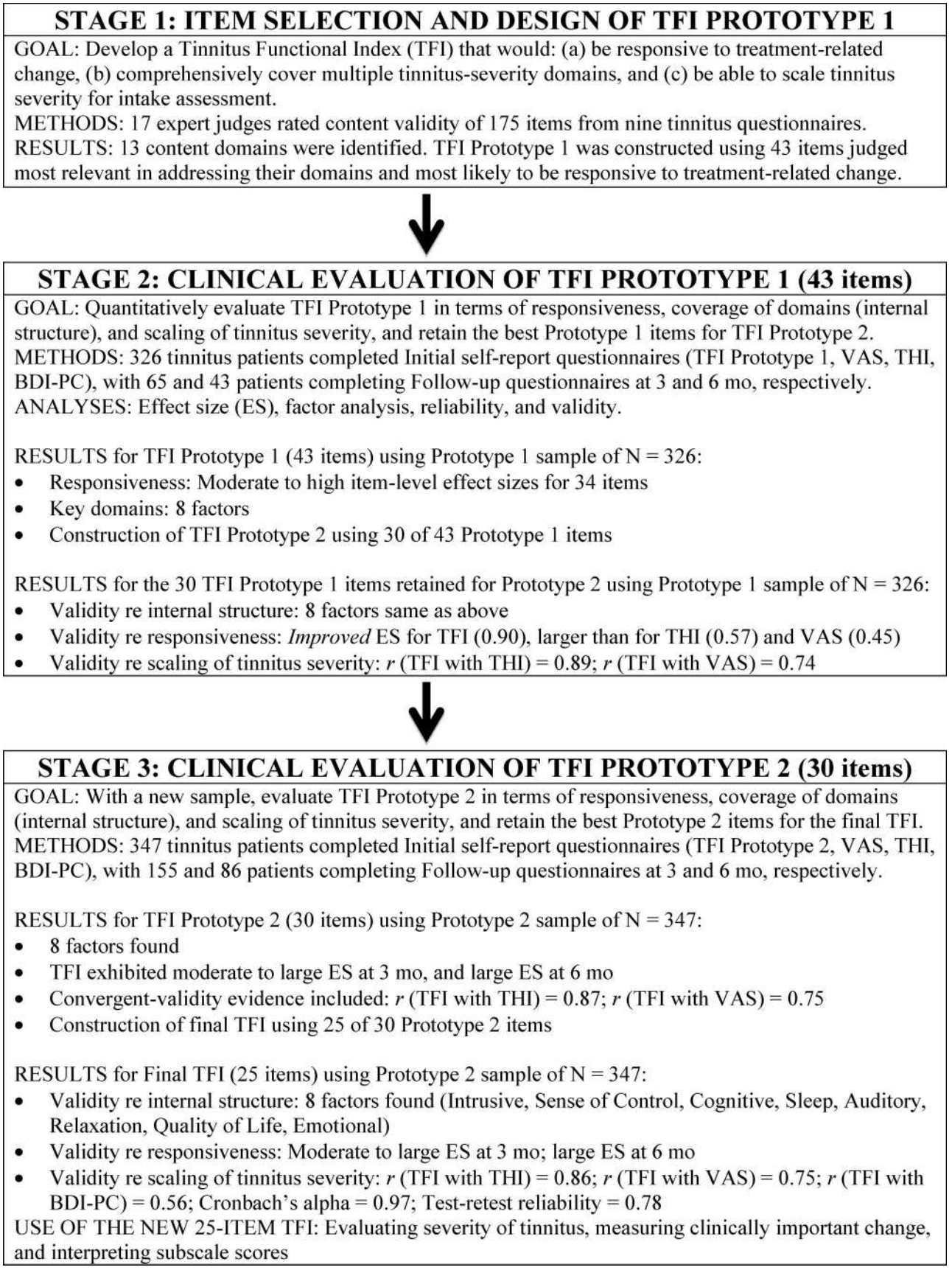

Fig. 1. Flowchart of the development of the TFI. BDI-PC, Beck Depression Inventory-Primary Care; ES, effect size; TFI, Tinnitus Functional Index; THI, Tinnitus Handicap Inventory; VAS, Visual Analog Scale.

TABLE 2. Important design considerations in constructing the Tinnitus Functional Index

	Requirement	Explanation
1	Responsiveness	Include only those items that are demonstrated to have moderate to high sensitivity to treatment-related change in tinnitus.
2	High construct validity for scaling of tinnitus severity	Each item should contribute to the overall effectiveness of the questionnaire in detecting individual differences in tinnitus severity.
3	Comprehensive coverage	To strengthen content validity, items, when taken together, should address all domains of tinnitus distress that have been represented in the majority of preexisting tinnitus questionnaires.
4	Brevity	Limit the questionnaire to 25 or fewer items if possible, but must be consistent with Item 3 above (comprehensiveness requirement).
5	Quantitative scaling	Likert-type scales preferred for all items; response options should provide high resolution without being conceptually complex.
6	Ease of use for patient	Wording of items should minimize reading difficulty and avoid ambiguity.
7	Ease of use for examiner	Scoring of items and of overall questionnaire should be simple, avoiding scale reversals and complex numerical calculations.
8	Avoidance of overly negative ideation	Avoid suggesting overly negative thoughts in questionnaire items (e.g., suicidal thoughts, feeling victimized, feeling hopeless, feelings of despair, dread, suffering). Note: The criterion to minimize negative ideation was established by the Tinnitus Research Consortium, the agency funding this research, and accepted as appropriate by the various participating investigators.

responses were visible only to that individual and the principal investigator and were not seen by other judges.

Content Relevance and Domain Identification • Each judge used a list of 10 content domains to select the major domains of negative tinnitus impact relevant to each item. (See Supplemental Digital Content 2, Figure A2, http://links.lww.com/EANDH/A54, for the 10 domains, which were originally recommended by the Tinnitus Research Consortium in its Request for Proposals for this project.) Judges could vote for more than one entry when selecting the domains they believed were relevant to any given item, and they could also propose other domains. Each web page provided comment fields in which judges could explain their responses.

After all the judges had completed their evaluations, the study investigators tallied the responses and presented the grouped results to the judges via the research website. Each questionnaire item was assigned to the specific domain that had received the largest number of votes. The domains identified by this process thus made use of the judges' accumulated experience with patients in clinical and research settings, supplying an organizing structure on which to base construction of TFI Prototype 1. It was anticipated that the set of initially defined domains would later be modified when factor analysis was applied to the results of clinical evaluation of Prototype 1, as described in Stage 2.

Content Validity Regarding Item Responsiveness • Each judge used a three-level scale (High = 2, Moderate = 1, Low = 0) to rate each item's expected responsiveness. For each item, the Panel's ratings were summed, divided by the number of judges providing ratings for that item, and multiplied by 50. This procedure produced standard scores on a scale ranging from 0 to 100, representing the group's mean rating regarding the expected responsiveness of each item.

Other Item Selection Criteria • Several a priori criteria were also applied to the item selection process: avoiding overly negative items (e.g., referring to suicidal thoughts, crying, or feeling victimized or helpless), as requested by the Tinnitus Research Consortium; eliminating several published questionnaire items whose wording referred exclusively to hearing loss without mention of tinnitus; and avoiding use of any item referring simultaneously to multiple subtopics within a domain, to avoid ambiguous or equivocal responses from respondents.

To take advantage of items already shown to have large effect sizes for detecting treatment effects, the selection process also made use of preexisting information on item effect sizes obtained during a clinical trial employing four of the nine preexisting questionnaires (those of Kuk et al. 1990; Meikle et al. 1995; Newman et al. 1996; and Jastreboff & Jastreboff 1999). The effect sizes were available from unpublished data obtained during the course of a prospective controlled study conducted to compare two different tinnitus treatment modalities from baseline to 18 mo (Henry et al. 2006). We identified items that exhibited large versus negligible effect sizes not only when comparing the two treatments but also when examining change within each treatment from baseline to 3, 6, 12, and 18 mo after treatment began. At the time the item selection process began, we tried but were unable to find comparable data on effect sizes for the other five previously published questionnaires that provided topics for inclusion in the TFI.

Results

Domains Identified • The Panel judges seemed to have little difficulty in assigning each of the 175 questionnaire items to the various domains they considered relevant. In most cases, a clear majority of votes identified the specific domain relevant to each item (e.g., Appendix A, Supplemental Digital Content 4, Figure A4, http://links.lww.com/EANDH/A56). By far, the largest proportion of the original 175 items—totaling 52 items—was assigned to the Emotional distress domain. In a few cases, there were equivocal results (e.g., Supplemental Digital Content 5, Figure A5, http://links.lww.com/EANDH/A57). In such cases, the judges were contacted again and asked to give input on which of the various candidate domains was the most

appropriate or whether some new domain could be identified which would be better.

The Panel identified several questionnaire items that did not belong to any of the 10 domains presented initially, resulting in the addition of three other domains of tinnitus impact: Cognitive interference from tinnitus, reduced Sense of control, and overall Quality of life. All three of the new domains were considered important dimensions of tinnitus impact, without which the content of Prototype 1 would be inadequate. The Panel review thus resulted in a total of 13 domains for constructing Prototype 1: Intrusiveness, Persistence, Emotional distress, Social distress or impact, Work interference, Leisure interference, Disturbance of sleep & rest, Disturbance of relaxation, Auditory perceptual difficulties attributable to tinnitus, Somatic & physical complaints due to tinnitus, Cognitive interference, Impaired quality of life, and Reduced sense of control. (As the study progressed, some domain labels were modified slightly.)

Responsiveness Ratings • All items that had been assigned to a given domain were listed in rank order of judges' ratings (0–100) of expected responsiveness. Supplemental Digital Content 6, Table A1, http://links.lww.com/EANDH/A58, shows the rank order listing of responsiveness of items in the Social distress domain.

Selection of Items for Prototype 1 • The rank-ordered responsiveness listings for the various domains formed the basis for selection of items for Prototype 1. The listings highlighted many duplications of content among the nine preexisting questionnaires; redundant items were excluded from further consideration. Items that were phrased in an overly negative way, or that were ambiguous, were also excluded, reducing the item set to a total of 35 nonduplicative items that addressed one or more of the 13 domains of tinnitus impact, with responsiveness ratings at or above the median values for the relevant domains.

Minimum of Three to Four Items for Each Domain • Published recommendations by measurement specialists have indicated that, for adequate reliability, any one major domain should be addressed by a minimum of three to four items, each preferably dealing with a single unique subtopic (Fabrigar et al. 1999; Moran et al. 2001). The set of 35 selected items did not provide the minimum of 3 to 4 items for all 13 domains identified by the judges; 8 items were added to meet this criterion. Two new items for the Work domain were adapted from the SF-36 questionnaire (Ware & Sherbourne 1992); two items were composed for the Sleep domain; one item each was composed for the Persistence, Leisure, and Cognitive domains; and one Somatic item (frequency of headaches related to tinnitus) was adapted from the Medical and Health History form used by the Tinnitus Clinic (Meikle et al. 2004). The resulting item set amounted to a total of 43 items for Prototype 1 (see Supplemental Digital Content 7, Table A2, http://links.lww.com/EANDH/A59, in which the topic addressed by each of the 43 items is briefly summarized).

Formatting and Preliminary Testing of Prototype 1 • The 43 items were formatted as questions, using a Likert-type response scale (0- to 10-point numeric rating scale) chosen because it provides good resolution for responsiveness (Lipsey 1990; Turk & Burwinkle 2005) and is familiar to many people and preferred over other response formats (Castle & Engberg 2004). Item-specific verbal anchors were supplied at scale extremes. The choice of a 0 to 10 response scale was also guided by the advice of Nunnally (1978) on properties of rating scales: Although there is a rapid increase in reliability going from 2 to 3 steps and so forth, the increase in reliability "tends to level off at about 7, and after about 11 steps there is little gain in reliability from increasing the number of steps" (p. 595).

On the basis of consultation with measurement experts, written instructions directing participants to describe their tinnitus "over the past week" were inserted as the lead-in phrase for each block of three to six items. Choice of recall interval is an important issue (U.S. Department of Health and Human Services 2006). A brief recall interval can help to minimize recall errors. Furthermore, for respondents whose tinnitus varies over time, a brief recall interval helps to minimize response variability.

The four-page Prototype 1 questionnaire received preliminary testing with 10 patients (5 in Ohio and 5 in Oregon), none of whom reported any problems responding to the questions. Prototype 1 was then submitted to and approved by the institutional review boards for the participating sites.

Conclusions

Content validity evaluation of items for TFI Prototype 1 accomplished the following: (1) Responses were obtained from the Panel judges for a large range of candidate items presented via the research website. (2) Quantitative methods were established for developing consensus among the judges. (3) A comprehensive set of items addressing 13 domains of tinnitus impact that are important to patients was identified. (4) Panel ratings of item responsiveness provided a rational basis for selecting the specific item content for each of the 13 domains. (5) Design of the 43-item Prototype 1 was accomplished using a consistent, user-friendly 0 to 10 metric and a recall interval of "Over the past week," both of which improve response reliability.

STAGE 2: CLINICAL EVALUATION OF TFI PROTOTYPE 1

Methods

Study Design • The goal of Stage 2 was to evaluate TFI Prototype 1 quantitatively in terms of responsiveness, underlying domains (internal structure), and ability to scale tinnitus severity. The best Prototype 1 items would be retained for TFI Prototype 2. Three classes of data were acquired: initial, retest, and follow-up. Participants were enrolled from patient populations at the five study sites shown in Table 3.

Initial Data • The initial questionnaire packets were mailed to all prospective patient participants several weeks before their initial clinic visit. The packets contained routine administrative papers as required by each participating site, in addition to the questionnaires for this study: a brief tinnitus history questionnaire developed at OHSU, the TFI Prototype 1, the Tinnitus Handicap Inventory (THI; Newman et al. 1996), and the Beck Depression Inventory-Primary Care (BDI-PC; Beck et al. 1997). Patients were asked to complete the forms at home and bring them to their clinic visit. At their visit, they were given the option of participating in this study. If they declined, their study-specific questionnaires constituted clinical data and were not used for this study.

TABLE 3. Number of participants providing initial and follow-up data (3 mo, 6 mo) at each site for TFI Prototypes 1 and 2

	Site					
	Oregon Health & Science University	Cleveland Clinic	Bay Pines Veterans Affairs Medical Center	James A. Haley Veterans' Hospital	Hearing & Speech Institute	Total
TFI Prototype 1 (43 items)						
Initial data	36	83	103	69	35	326
3 mo follow-up	*	28	17	20	*	65
6 mo follow-up	*	29	9	4	*	42
TFI Prototype 2 (30 items)						
Initial data	68	64	134	81	†	347
3 mo follow-up	46	44	32	33	†	155
6 mo follow-up	29	31	17	9	†	86

*Site did not participate in collection of Prototype 1 follow-up data.
†Site did not participate in collection of any Prototype 2 data.
TFI, Tinnitus Functional Index.

Retest Data • For a subset of participants, the TFI was administered a second time to obtain data for test–retest reliability. Participants who were eligible to provide Retest data were those who (a) had given informed consent to allow their initial data to be used in the study and (b) had filled out their initial questionnaires at home within the specified Retest interval of 7 to 30 days before their clinic visit. Eligible participants were asked whether, while waiting to see the audiologist, they would complete several additional questionnaires (without being informed of the Retest nature of the task). Before any contact with professional staff, consenting participants filled out what were described as "questionnaires similar to what you filled out at home," including the TFI Prototype 1.

Follow-Up Data • One of the major study goals (evaluating TFI responsiveness when used as an outcome measure after tinnitus-related interventions) required that we evaluate participants' tinnitus status at appropriate follow-up intervals. Follow-up data were scheduled for acquisition at 3, 6, and 9 mo after intake. Because follow-up responses at 9 mo numbered only 25 cases for TFI Prototype 1 and 27 cases for TFI Prototype 2, and thus fell short of the numbers needed for valid statistical analysis of responsiveness using subgroups, we excluded 9 mo data from this article.

Participants

Site Selection • Prior clinical experience had shown that the two tertiary care centers participating in this study—the OHSU Tinnitus Clinic and the Tinnitus Management Clinic at the Cleveland Clinic (CC)—attracted mainly patients with more severe tinnitus, with relatively few reporting tinnitus that was only a small problem. It was essential for this study to include patients with milder tinnitus, however, to evaluate the ability of the TFI to scale the severity of tinnitus over the widest possible range. Three additional sites were therefore included which were known to encompass the full range of tinnitus severity levels: Bay Pines Veterans Affairs Medical Center (BPVA) in Bay Pines, Florida; James A. Haley Veterans' Hospital (JHVH) in Tampa, Florida; and the Hearing & Speech Institute (HSI) in Oregon. The large majority of patients at the VA medical centers are male, most of whom receive tinnitus clinical services at little or no cost. At this time, there is no comparable healthcare resource for female tinnitus patients whose tinnitus is mild. To achieve the full range of tinnitus severity required for adequate evaluation of the TFI, this study accepted the tradeoff of oversampling male patients.

Initial and Retest Participants • At all study sites, consecutive patients reporting persistent tinnitus were invited to give informed consent to allow their initial (intake) data to be used in the study (constituting a convenience sample of those who consented). For patients who gave consent, no payment was provided for their initial data. As described earlier, a subset of those who provided initial data were invited to fill out additional questionnaires while waiting to see the audiologist, to obtain test–retest reliability data on the TFI. This occurred at two sites (OHSU and HSI). Participants received $10 for completing the retest questionnaires.

Follow-Up Participants • Patients who had consented to follow-up participation at their initial clinic visit were later mailed follow-up questionnaires (with preaddressed, stamped return envelopes). Follow-up data were acquired at 3 and 6 mo only at CC, BPVA, and JHVH; resources were not available at OHSU and HSI to collect follow-up data. At CC and BPVA, follow-up participants were paid $10 for each completed follow-up packet returned to the clinic (payment was neither available for, nor provided to, follow-up participants at JHVH).

Measures

Tinnitus History Questionnaire • The Tinnitus History questionnaire included the following:

a. Demographic items (age, gender, Hispanic origin, race).
b. A Visual Analog Scale (VAS) for tinnitus severity, which asked participants to place a mark on a horizontal 100-mm line, with the directions "On the line below, please place a mark to show HOW SEVERE your tinnitus has been over the past week." The line was anchored at the left with the label "No tinnitus present" and at the right with the label "The worst tinnitus you can imagine." A score was obtained by measuring in millimeters from the left end of the line to the point that the participant marked.
c. An item for scaling the global severity of tinnitus ("How much of a problem is your tinnitus?") with response options 1 (not a problem), 2 (a small problem), 3 (a moderate problem), 4 (a big problem), and 5 (a very big problem).
d. An item for self-report of hearing problems ("Are you having any PROBLEMS HEARING speech or other

sounds?") with response options 1 (no problem), 2 (a small problem), 3 (a moderate problem), 4 (a big problem), and 5 (a very big problem).

Tinnitus Handicap Inventory • Participants filled out the 25-item THI (Newman et al. 1996), a widely used questionnaire with established validity for scaling the severity of tinnitus. Response options for each THI item have three levels (yes = 4, sometimes = 2, no = 0). An overall THI score is computed by summing responses and has a potential range of 0 to 100. The THI contains three subscales (functional, 11 items; emotional, 9 items; catastrophic, 5 items).

Beck Depression Inventory • Participants completed the 7-item BDI-PC (Beck et al. 1997). Each item has a four-level response option that is tailored to the seven aspects of depression being measured (e.g., sadness, pessimism).

Retest and Follow-Up Data Packets • Retest data packets included the TFI Prototype 1, THI, and VAS. Follow-up data packets at 3 and 6 mo included the TFI Prototype 1; the THI and VAS; an abbreviated Tinnitus History questionnaire (omitting demographic questions); and a block of questions concerning treatments the participant may have pursued since the Initial visit.

Global Perception of Change • Follow-up forms also included an item concerning the participant's Global Perception of Change. For this item, participants were asked the following question: "All things considered, how is your overall tinnitus condition now, compared to your first visit to this clinic?" Participants rated the magnitude of any changes in their tinnitus using a seven-point scale: 1 = much improved, 2 = moderately improved, 3 = slightly improved, 4 = no change, 5 = slightly worse, 6 = moderately worse, 7 = much worse.

Treatment Considerations in Collection of Follow-Up Data • It should be emphasized that evaluating treatment efficacy was not an objective of this research. To maintain our primary focus on evaluating the responsiveness of the TFI, data acquisition concentrated on evaluating participants' self-reported change in overall tinnitus status at 3 and 6 mo follow-up without regard to the specific form of treatment that might have mediated such changes in status.

Tinnitus interventions varied widely between the five study sites, ranging from more intensive treatment (e.g., special counseling sessions; fitting of ear-level "maskers" and combination instruments [hearing aid and masker combined]; provision of tabletop sound generators; and medications for associated sleep disturbance, anxiety, depression) to less intensive treatment (e.g., fitting of hearing aids; provision of written information about tinnitus; and brief counseling in conjunction with audiology appointments). There were also many possible combinations of the various alternatives.

Scoring and Missing Data • Scoring of the overall TFI was accomplished by summing each participant's individual responses to the TFI items, dividing by the number of items answered, and then multiplying by 10 to achieve an overall scale of 0 to 100. Both the THI and the severity VAS also use 0 to 100 scales, facilitating direct comparisons between all three scales. Although 100% completion of a questionnaire is desirable for clinical use, there will always be items that some patients do not answer. One solution to the problem of unanswered items (if few in number) is to average scores of answered items and convert the average to the overall scale metric, as described earlier. Such a practice is viewed as acceptable as long as the scale has strong internal consistency reliability and participants have answered most items (Schafer & Graham 2002). For scoring purposes, at the scale level and for subscales with four or more items, participants were required to answer 75% of the scale's items for a valid score to be computed. For three-item subscales, participants were required to answer two of three items (67%). Thus, participants missing more than 25% of items (33% on three-item subscales) were assigned a missing value for the entire scale or subscale.

Data Analyses • To achieve our three main measurement goals—responsiveness, identification and refinement of tinnitus severity domains, and scaling of tinnitus severity for intake assessment—we used statistical analyses pertinent to each, employing SPSS versions 15 and 16. In Stage 2, we used results regarding item responsiveness and severity domains to reduce the 43-item Prototype 1 to 30 items for Prototype 2 and then evaluated the 30-item TFI using data from Prototype 1.

Evaluating Potential Responsiveness: Floor or Ceiling Effects, Missing Data, and Effect Sizes • For evaluating the TFI's potential responsiveness to treatment-related changes in tinnitus impact, we first examined frequency distributions of scores for TFI items and the overall TFI scale and subscales to identify the presence of floor or ceiling effects and missing data that could interfere with responsiveness. One purpose was to identify items having potential floor effects (i.e., scores clustered mostly at the 0 end of the 0 to 10 response scale) that would limit detection of improvement in a patient's tinnitus condition or ceiling effects (i.e., scores clustered mostly at the 10 end of the 0 to 10 response scale) that would limit detection of worsening tinnitus. Because the TFI was intended to detect treatment-related improvement as reflected by declines in the index score, floor effects were considered a greater limitation than ceiling effects. A second purpose was to identify items with relatively large amounts of missing data, which might signify items that were confusing or not applicable to participants.

Next, we examined effect sizes of the TFI items and overall TFI scale and subscales. As explained previously, we did not implement a specific treatment protocol to compute effect sizes comparing treatment and control groups. Instead, using an approach recommended by Lipsey (1990), we computed effect sizes for criterion groups that we expected to differ from one another to the extent that a treatment and control group would differ.

Our criterion groups were derived from participants' responses at 3 and 6 mo to the Global Perception of Change item, reported on a 7-point scale ranging from 1 (much approved) to 4 (no change) to 7 (much worse). For TFI Prototype 1, because of the small follow-up sample (described later), we collapsed response categories to create three criterion groups: Improved, Unchanged, and Worse. This allowed us to achieve minimally adequate sample sizes for estimating effect sizes.

For each of the criterion groups, we calculated effect sizes using Cohen's d (initial mean score minus follow-up mean score, divided by the pooled SD for the two scores). Cohen (1988) considered effect sizes ≥ 0.20 as small, ≥ 0.50 as moderate, and ≥ 0.80 as large. For Prototype 1, our prediction was that effect sizes for the Improved, Unchanged, and Worse groups would be positive, near zero, and negative, respectively.

Identifying Tinnitus-Severity Domains: Factor Analysis • To identify the key domains underlying the TFI, we used exploratory factor analysis both to provide validity evidence for internal structure of the TFI and to create TFI subscales. There is potential value in having subscales to measure important tinnitus severity factors, for example, to identify subgroups of tinnitus patients on the basis of differing profiles of subscale scores and to determine whether specific treatments result in differential effectiveness across subscales. Thus, factor analysis results added to the content and construct validity evidence for the TFI.

Exploratory factor analysis of the 43 items in TFI Prototype 1 was performed on the initial scores, beginning with principal components analysis (PCA) to determine the likely number of factors and followed by Principal Axis Factoring (PAF) to evaluate common factors. In all cases, we evaluated not only orthogonal (varimax) rotations but also oblique (Oblimin, delta = 0) rotations, because we anticipated substantial correlation between the various domains of tinnitus impact. Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy (SPSS versions 15 and 16) to evaluate factorability of the correlation matrix ranged from 0.94 to 0.97 for the various factor analyses in this study. We excluded from each analysis any participant missing data for 10% or more of the TFI items being analyzed (1.7–3.2% of participants were excluded, depending on the specific items being analyzed), and then conducted the analysis using mean substitution of the missing item responses.

Scaling of Tinnitus Severity: Descriptive Statistics, Reliability, and Validity • For scaling the severity of tinnitus for use in intake assessment, we examined descriptive statistics of the TFI scale and subscales, reliability (Cronbach's alpha for internal consistency, test–retest), and three aspects of construct validity (validity regarding internal structure of the TFI, convergent validity, and discriminant validity). Evidence for validity regarding internal structure was provided by factor analysis of TFI items as described earlier. Evidence for convergent validity was obtained by computing the Pearson *r* correlation between the TFI and the THI and VAS, two existing measures of tinnitus severity that were expected to correlate strongly with the TFI. Evidence for discriminant validity was obtained by correlating the TFI with the BDI-PC, a measure that was expected to be moderately associated with the TFI but with distinctly lower correlations than those for convergent validity. To obtain additional convergent validity evidence regarding clinically meaningful differences on the TFI, we examined the mean TFI values corresponding to each of five responses (not a problem, a small problem, a moderate problem, a big problem, a very big problem) that participants gave to the item, "How much of a problem is your tinnitus?" For scaling of tinnitus severity, it is desirable for distributions of TFI scores to be fine-grained and have adequately wide variability. Strong reliability of the TFI and subscales is important for its use in intake assessment.

Results

Prototype 1 Sample • We established the a priori criterion to accept as valid only those questionnaires in which at least 75% of the items were completed. Valid initial (intake) questionnaires were obtained from 326 of 327 participants who completed Prototype 1 questionnaires. Table 4 summarizes demographic characteristics as well as tinnitus and hearing characteristics of the initial sample. Combining data from all five sites was expected to produce ethnic and racial diversity, but only 6.7% of participants reported they were minority or mixed race (using the racial nomenclature recommended by current NIH guidelines), with 2.5% indicating they were Hispanic. The gender distribution was 81% male. Most participants were in their 50s, 60s, or 70s, with an average age of 62 yr. The response distributions to the five-level item "How much of a problem is your tinnitus?" demonstrated that the participant samples from the two tertiary-care centers (OHSU and CC) did, as expected, include higher percentages of cases with more severe tinnitus (described as "a big problem" or "a very big problem") than those from general audiology clinics (BPVA, JHVH, and HSI). See Supplemental Digital Content 8, Table B1, http://links.lww.com/EANDH/A60.

Retest data were provided by 30 participants. Follow-up data for Prototype 1 were provided by 65 participants at 3 mo and 42 at 6 mo (see Tables 3 and 4). Table 4 also summarizes tinnitus treatments tried by follow-up participants since their initial visit to the clinic as well as their Global Perception of Change regarding their tinnitus, ranging from "much improved" to "much worse."

Results for the 43-Item TFI Prototype 1

Responsiveness: Floor or Ceiling Effects and Missing Data for 43 Items in Prototype 1 • Item-level frequency distributions for initial data were examined for skewed distributions and missing data. Mean scores for 42 of the 43 items fell within the range 2.3 to 6.7, well above the desired lower limit of 2.0—the exception was one Somatic item (Feeling in ill health because of tinnitus), with a mean of 1.9—confirming that item responses generally were not affected by floor or ceiling effects. There was a small amount of missing data at the item level, suggesting that participants had few problems responding to the questionnaire. For each of 29 items, more than 99% of participants gave valid responses; fewer than 1% had missing responses. For another 11 items, 1.0 to 1.9% had missing responses. For the remaining three items, 2.0 to 3.1% had missing responses.

Responsiveness: Item-Level Effect Sizes for 43 Items in Prototype 1 • The main purpose of obtaining item-level effect sizes was to identify items with the strongest responsiveness for detecting treatment-related improvement. Item-level effect sizes were computed to estimate treatment-related change from initial data to 3 mo follow-up data for the three criterion groups: improved (n = 11), unchanged (n = 45), and worse (n = 9). Item-level effect sizes are included in Supplemental Digital Content 9, Table B2, http://links.lww.com/EANDH/A61. (Because follow-up responses at 6 mo numbered 42 cases and fell short of the numbers needed for valid statistical analysis when stratified into the three change categories, 6 mo effect sizes were not computed for TFI Prototype 1 items, overall scale, or subscales.)

For 41 of the 43 items in TFI Prototype 1, effect sizes for the Improved group ranged from 0.24 to 1.50 but mainly fell within the range 0.60 to 1.00. Using the size categories suggested by Cohen (1988), 14 of the items had large effect sizes (≥0.80), 20 items had moderate effect sizes (in the range 0.50–0.79), and 7 had "small" effect sizes (in the range 0.20–0.49). For the

TABLE 4. Sample characteristics for Tinnitus Functional Index Prototypes 1 and 2 at initial and follow-up data collection

Initial Data Collection	Prototype 1, n (%) n = 326	Prototype 2, n (%) n = 347
Demographic characteristics		
Gender		
Male	265 (81.3)	285 (82.1)
Female	61 (18.7)	60 (17.3)
Unreported	—	2 (0.6)
Age (yr)		
Mean	62.2	60.2
SD	12.7	11.8
Range	17–87	22–90
Ethnicity		
Hispanic	8 (2.5)	12 (3.5)
Non-Hispanic	312 (95.7)	325 (93.7)
Missing or prefer not to answer	6 (1.8)	10 (2.9)
Race		
Minority or mixed race*	22 (6.7)	31 (8.9)
White only	299 (91.7)	307 (88.5)
Missing or prefer not to answer	5 (1.5)	9 (2.6)
Tinnitus and hearing characteristics		
About how long have you been aware of hearing tinnitus?		
Less than 1 yr	50 (15.3)	58 (16.7)
1–2 yr	31 (9.5)	42 (12.1)
3–5 yr	50 (15.3)	38 (11.0)
6–10 yr	43 (13.2)	43 (12.4)
11–20 yr	41 (12.6)	38 (11.0)
20+ yr	108 (33.1)	128 (36.9)
About how often does your tinnitus seem to be present?		
Present occasionally	20 (6.1)	10 (2.9)
Present some of the time	30 (9.2)	28 (8.1)
Present most of the time	76 (23.3)	73 (21.0)
Present always	197 (60.4)	236 (68.0)
How much of a problem is your tinnitus?		
Not a problem	29 (8.9)	8 (2.3)
A small problem	57 (17.5)	45 (13.0)
A moderate problem	100 (30.7)	100 (28.8)
A big problem	82 (25.2)	113 (32.6)
A very big problem	55 (16.9)	81 (23.3)
Are you having any problems hearing speech or other sounds?		
No problem	44 (13.5)	43 (12.4)
A small problem	64 (19.6)	63 (18.2)
A moderate problem	98 (30.1)	100 (28.8)
A big problem	79 (24.2)	89 (25.6)
A very big problem	31 (9.5)	49 (14.1)
Participation in 3 mo and 6 mo follow-up		
3 mo and 6 mo follow-up	33 (10.1)	76 (21.9)
Only 3 mo follow-up	32 (9.8)	79 (22.8)
Only 6 mo follow-up	9 (2.8)	10 (2.9)
Neither follow-up	252 (77.3)	182 (52.4)
3 mo Follow-Up Data Collection	**n = 65**	**n = 155**
No. of tinnitus treatments tried since visit to clinic		
0	9 (13.8)	46 (29.7)
1	31 (47.7)	25 (16.1)
2	17 (26.2)	32 (20.6)
3–5	8 (12.3)	39 (25.2)
6–9	0 (0.0)	13 (8.4)
Specific tinnitus treatments tried		
Hearing aid for one or both ears	29 (44.6)	61 (39.4)
Tinnitus masker/sound generator	9 (13.8)	28 (18.1)
Portable sound-generating device	22 (33.8)	43 (27.7)
Medical treatment (medications, sleep therapy)	13 (20.0)	40 (25.8)
Psychological counseling (in a group or alone)	5 (7.7)	16 (10.3)

(*Continued*)

TABLE 4. Continued.

3 mo Follow-Up Data Collection	n = 65	n = 155
Relaxation training	6 (9.2)	21 (13.5)
Cognitive behavioral training	3 (4.6)	12 (7.7)
Explanation and information about tinnitus	†	76 (49.0)
Other treatment	9 (13.8)	18 (11.6)
How is your overall tinnitus condition now, compared to your first visit to this clinic?		
Much improved	2 (3.1)	8 (5.2)
Moderately improved	3 (4.6)	14 (9.0)
Slightly improved	6 (9.2)	30 (19.4)
No change	45 (69.2)	68 (43.9)
Slightly worse	7 (10.8)	13 (8.4)
Moderately worse	1 (1.5)	12 (7.7)
Much worse	1 (1.5)	5 (3.2)
Missing	0 (0.0)	5 (3.2)
6 mo Follow-Up Data Collection	**n = 42**	**n = 86**
No. of tinnitus treatments tried since visit to clinic		
0	20 (47.6)	29 (33.7)
1	13 (31.0)	7 (8.1)
2	4 (9.5)	17 (19.8)
3–5	5 (11.9)	24 (27.9)
6–9	0 (0.0)	9 (10.5)
Specific tinnitus treatments tried		
Hearing aid for one or both ears	11 (26.2)	29 (33.7)
Tinnitus masker/sound generator	5 (11.9)	16 (18.6)
Portable sound-generating device	9 (21.4)	31 (36.0)
Medical treatment (medications, sleep therapy)	7 (16.7)	22 (25.6)
Psychological counseling (in a group or alone)	0 (0.0)	12 (14.0)
Relaxation training	1 (2.4)	16 (18.6)
Cognitive behavioral training	1 (2.4)	10 (11.6)
Explanation and information about tinnitus	†	38 (44.2)
Other treatment	4 (9.5)	15 (17.4)
How is your overall tinnitus condition now, compared to your first visit to this clinic?		
Much improved	0 (0.0)	3 (3.5)
Moderately improved	2 (4.8)	13 (15.1)
Slightly improved	4 (9.5)	13 (15.1)
No change	25 (59.5)	34 (39.5)
Slightly worse	7 (16.7)	10 (11.6)
Moderately worse	2 (4.8)	11 (12.8)
Much worse	2 (4.8)	1 (1.2)
Missing	0 (0.0)	1 (1.2)

For any characteristic, when numbers for the initial sample do not total to 326 (Prototype 1) or 347 (Prototype 2) or percentages do not total 100%, it is because of missing data (e.g., blank, refused, don't know).

**Minority or mixed race for Prototype 1 included 22 individuals, with specific race categories indicated (participants could check more than one): African American or Black (13), American Indian/Alaska Native (9), Asian (2), Hawaiian or Pacific Islander (1), White (8), Other (1). Minority or mixed race for Prototype 2 included 31 individuals, with specific race categories indicated (participants could check more than one): African American or Black (16), American Indian/Alaska Native (5), Asian (2), Hawaiian or Pacific Islander (1), White (4), Other (7).*

†For Prototype 1, "Explanation and information about tinnitus" was not specified as a treatment option in the questionnaire but was added for Prototype 2.

remaining two items, TFI scores for the Improved group were worse at 3 mo follow-up: Question 31, "How much did your tinnitus cause you to have headaches?" (ES = −0.49) and Question 35, "How much of the time did your tinnitus cause you to feel less interested in going out?" (ES = −0.42).

As expected, most effect sizes for the Unchanged group were relatively close to zero, ranging from −0.17 to 0.32, with more than half the items (22 of 43) having effect sizes very near zero (±0.09). Contrary to expectations, only 18 of 43 items for the Worse group were negative, whereas 13 items had positive effect sizes of 0.20 or larger, with four of the effect sizes exceeding 0.50. With only 11 participants in the Improved group and 9 in the Worse group, Prototype 1 effect sizes should be interpreted with caution. Effect sizes needed to be estimated using larger criterion group sizes, which were obtained later in Prototype 2.

Tinnitus-Severity Domains: Factor Analysis of 43 Items in Prototype 1 • Starting by selecting factors with eigenvalues ≥1.0, the number of factors extracted was varied in successive analyses by specifying extraction of five, six, seven, and eight factors. The clearest, most easily interpreted factor structure was obtained using oblique PAF. Factors five to eight had initial eigenvalues <1.0 but were nevertheless retained as meaningful factors on the basis of guidelines by a number of researchers who generally recommended erring on the side of overextraction rather

than underextraction of factors (e.g., Wood et al. 1996; Reise et al. 2000; Tabachnick & Fidell 2001).

We also experimented with varying the analysis sample by selecting participants on the basis of their responses to the question, "How much of a problem is your tinnitus?" and omitting 29 participants who responded with "not a problem." We then sequentially omitted weak items from the factor analysis on the basis of small factor loadings and poor item-level effect sizes found earlier.

The final analysis (oblique PAF of 30 items, with 8 factors specified, from 285 participants whose tinnitus was equal to or greater than A small problem) explained 82.4% of the variance. Of the eight factors shown in Supplemental Digital Content 10, Table B3, http://links.lww.com/EANDH/A62, seven exhibited clear patterns of factor loadings whereas an eighth factor (Quality of Life) had a mixed pattern of loadings.

Identification of this eighth factor, representing a cross-factor domain, was not completed until after detailed review and discussion of the 30-item PAF, focusing on seven items that did not load strongly on any single factor (most of them displaying moderate factor loadings on two or more factors and having absolute values in the range 0.22–0.64). It became evident that these items could appropriately be combined into a Quality of Life factor or subscale.

In summary, 8 of the 13 domains that had been identified earlier by the Item Selection Panel were corroborated as separate factors by the factor analyses of Prototype 1 data. Another 4 domains of the original 13 domains were important contributors to the eight factor-based subscales shown in Supplemental Digital Content 9, Table B2, http://links.lww.com/EANDH/A61. For example, items representing the original Intrusiveness and Persistence domains were merged to generate a single Intrusive subscale. As described earlier, the Social Distress, Leisure, and Work domains were successfully merged into the Quality of Life subscale. Of the original list of 13 domains, the only domain that was not represented for Prototype 2 was the Somatic domain. All three items representing the Somatic domain exhibited low factor loadings on all factors except the Emotional factor, suggesting that the Somatic domain was not useful as a separate subscale for measuring the severity of tinnitus. Only one of the Somatic items ("How fatigued has tinnitus caused you to feel?") was retained for Prototype 2, because it was the only Somatic item with a mean score above 3.5 (on the 0 to 10 scale) and was therefore retained as a potentially helpful item contributing to the Emotional subscale.

Selection of 30 Prototype 1 Items for TFI Prototype 2 • Using results regarding responsiveness and the 8 identified severity domains, we created TFI Prototype 2 by selecting 30 of the 43 Prototype 1 items with the strongest effect sizes, clearest factor loadings, fewest missing data, and best distributions to minimize floor and ceiling effects. We also selected items for the eight Prototype 2 subscales using these same criteria plus the requirement that each subscale be composed of at least three items. We removed 13 items as described in Supplemental Digital Content 9, Table B2, http://links.lww.com/EANDH/A61.

Results for the 30 Prototype 1 Items Retained for TFI Prototype 2 • Before testing Prototype 2 with a new sample in Stage 3, we used the Prototype 1 sample of 326 to examine the 30-item TFI with respect to both scaling of tinnitus severity for intake assessment and responsiveness.

Validity for Scaling of Tinnitus Severity

Descriptive Statistics • We computed an overall score for the 30-item TFI and scores for its eight subscales and examined the distributions of overall and subscale scores. The initial scores for the 30-item TFI exhibited a smoothly rising cumulative distribution over the total 0 to 100 range, with no flat spots or sudden "jumps." The mean TFI (45.4 ± 27.6, N = 326) fell between the mean values of the THI (37.7 ± 27.9, N = 310) and severity VAS (57.3 ± 28.3, N = 309). Median values for the three overall scores were close to the means: 45.2 (TFI), 30.0 (THI), and 61.3 (severity VAS).

For the 30 Prototype 1 items retained for Prototype 2, none of the 326 Prototype 1 participants had a missing score on the overall 30-item TFI scale (using the 75% cutoff). For seven of the Prototype 1 subscales, 0.3 to 0.9% of those participants had missing scores (using the 67% cutoff described earlier). For the one four-item subscale, 1.8% of the participants had missing scores.

Reliability • Internal consistency reliability of the TFI was excellent (Cronbach's alpha = 0.98), and test–retest reliability was high ($r = 0.91$). For the eight subscales, Cronbach's alpha values ranged from 0.87 to 0.97, and test–retest reliability ranged from 0.71 to 0.92.

Convergent and Discriminant Validity • High Pearson correlations were found between TFI initial scores and those for the THI ($r = 0.89$) and severity VAS ($r = 0.74$, $p < 0.001$ for both). In addition, the initial TFI scores exhibited a moderate correlation ($r = 0.57$) with the BDI-PC.

The convergent validity of Prototype 1 was further supported when TFI mean scores were compared across the five response levels of the global severity item, "How much of a problem is your tinnitus?" as displayed in Table 5. Rounded to whole numbers, the TFI means were 7 (not a problem), 20 (a small problem), 38 (a moderate problem), 65 (a big problem), and 77 (a very big problem). There was a strong association between TFI means and the five levels of the tinnitus "problem" item ($F[4,318] = 187.0$, $p < 0.001$).

Validity Regarding Responsiveness • We compared effect sizes for the TFI with effect sizes for the VAS and THI. Table 6 shows that, for the Improved group, the effect size of 0.90 for the overall Prototype 1 was larger than the Improved group's effect size of 0.57 for the THI and of 0.45 for the severity VAS. For the Unchanged and Worse groups, Prototype 1 effect sizes were fairly close to zero, similar to the THI and VAS.

We also examined the effect sizes of the Improved group on the eight factor-based subscales at 3 mo follow-up (Supplemental Digital Content 9, Table B2, http://links.lww.com/EANDH/A61). Six of the subscales exhibited large effect sizes, and two had moderate effect sizes.

Conclusions

TFI Prototype 1 displayed the following desirable measurement properties:

1. Participants had few problems responding to the items. Missing data were minimal and item distributions did not exhibit significant floor or ceiling effects.
2. Excellent convergent validity was found when Prototype 1 (on the basis of 30 of 43 items) was compared with other published scales for tinnitus severity (the THI and

TABLE 5. Comparison of performance of three versions of the TFI

Questionnaire Attributes	TFI Version: TFI Prototype 1 (30 Items) n = 326	TFI Prototype 2 (30 Items) n = 347	Final TFI (25 Items) n = 347
Validity regarding internal structure			
Number of factors identified	8	8	8
Variance accounted for by principal axis factoring (oblique rotation with eight factors extracted)*	82.4%	78.5%	79.5%
Validity regarding responsiveness			
3 mo effect sizes			
Improved group	0.90	0.83	0.84
Unchanged group	0.08	0.29	0.30
Worse group	0.03	−0.14	−0.19
6 mo effect sizes			
Improved group	†	1.46	1.47
Unchanged group	†	0.14	0.15
Worse group	†	0.12	0.09
Validity regarding scaling of tinnitus severity			
Convergent validity (Pearson correlations with other severity scales)			
r with THI	0.89	0.87	0.86
r with VAS	0.74	0.75	0.75
Discriminant validity			
r with BDI-PC	0.57	0.57	0.56
Convergent validation by analysis of variance comparing mean TFI across five levels of "How much of a problem is your tinnitus?"	$F[4,318] = 187.0$, $p < 0.001$	$F[4,342] = 152.2$, $p < 0.001$	$F[4, 342] = 149.9$, $p < 0.001$
	n / Mean (SD)	n / Mean (SD)	n / Mean (SD)
1 = Not a problem	29 / 7 (6)	8 / 13 (12)	8 / 14 (12)
2 = A small problem	57 / 20 (13)	45 / 20 (10)	45 / 21 (10)
3 = A moderate problem	100 / 38 (16)	100 / 42 (16)	100 / 42 (16)
4 = A big problem	82 / 65 (17)	113 / 64 (15)	113 / 65 (15)
5 = A very big problem	55 / 77 (16)	81 / 77 (16)	81 / 78 (16)
Reliability			
Cronbach's alpha	0.98	0.98	0.97
Test–retest	0.91	0.76	0.78
Descriptive statistics			
Mean ± SD (initial data)	45.4 ± 27.6	53.8 ± 25.0	54.4 ± 24.7
Median	45.2	56.0	57.5

Mean (SD) values are rounded to whole numbers.

BDI-PC, Beck Depression Inventory-Primary Care; TFI, Tinnitus Functional Index; THI, Tinnitus Handicap Inventory; VAS, Visual Analog Scale.

**For Prototype 1, factor analysis excluded participants who reported "not at all" to how much of a problem their tinnitus is. For Prototype 2 and final TFI, factor analysis excluded participants who reported "not at all" or "small" to the same question.*

†For Prototype 1, 6 mo effect sizes were not computed because of small group sizes.

severity VAS) and with the global severity item, "How much of a problem is your tinnitus?"

3. Prototype 1 exhibited a clear eight-factor structure incorporating eight widely acknowledged components of negative tinnitus impact, thus refining the definition of the construct being measured and providing an expanded set of measures for evaluating treatment effects.
4. The effect size for the group of participants with Improved tinnitus was 0.90 (compared with small or near-zero values for those whose tinnitus was Unchanged or Worse), indicating that the overall questionnaire possessed good responsiveness for use as an outcome measure.
5. Of the 43 individual items in Prototype 1, 34 (79%) displayed effect sizes in the moderate to high range for the Improved group, indicating that the Item Selection Panel had succeeded in identifying items that were responsive to treatment-related change.
6. By using predetermined criteria for selecting the best-functioning items, we were able to reduce the size of the questionnaire to a total of 30 items, thus generating TFI Prototype 2.

STAGE 3: CLINICAL EVALUATION OF TFI PROTOTYPE 2

Methods

The goal of Stage 3 was to use a new sample to evaluate the 30-item Prototype 2 in terms of responsiveness, key domains (internal structure), and scaling of tinnitus severity and to retain

TABLE 6. Comparison of mean change at 3 mo follow-up for improved, unchanged, or worse groups on overall scores for TFI Prototype 1 (30 of 43 items retained for Prototype 2), Tinnitus Handicap Inventory, and severity VAS

Group	Number in Group*	Mean Change	Pooled SD	Effect Size
Improved				
TFI Prototype 1	11	22.6	25.0	0.90
THI	11	15.4	27.2	0.57
Severity VAS	11	12.5	27.6	0.45
Unchanged				
TFI Prototype 1	44	2.2	26.7	0.08
THI	44	2.4	26.2	0.09
Severity VAS	38	4.2	25.8	0.16
Worse				
TFI Prototype 1	9	0.6	22.2	0.03
THI	9	5.3	28.6	0.19
Severity VAS	8	0.7	20.4	0.03

Pooled SD combines SDs for initial and 3 mo scores. In preliminary presentations of this work (Meikle, Henry et al. 2008; Meikle, Stewart et al. 2008), computations of effect sizes took advantage of the repeated-measures design by computing effect sizes using initial minus follow-up mean differences, divided by the SD of the difference scores. Later, we adopted a different method for calculating effect sizes, changing the denominator to the pooled SD of the initial and follow-up scores. The latter method makes the effect sizes of repeated-measures designs comparable to the effect sizes obtained in studies comparing different groups of subjects (e.g., comparing treated with untreated subjects), thus facilitating meta-analyses that may be conducted in the future.

TFI, Tinnitus Functional Index; THI, Tinnitus Handicap Inventory; VAS, Visual Analog Scale.

**Not all of the 65 follow-up participants completed all three of the questionnaires: 64 participants completed TFI Prototype 1, 64 completed the THI, and 57 completed the VAS.*

the best Prototype 2 items for the final TFI. In general, Stage 3 methods and data analysis were the same as for Stage 2; any differences are noted in later sections.

One of the five sites, HSI, discontinued participation in the study. Retest data were collected at only one site (OHSU). Payment to retest participants was increased to $20 in the effort to increase the number of retest responses completed.

To increase the percentage of participants likely to provide follow-up responses, recruitment efforts at the two VA medical centers were changed slightly to reduce the number of participants reporting the magnitude of their tinnitus problem as not a problem or a small problem. Although this introduced a systematic bias in favor of patients with more severe tinnitus, we believe that this improved our ability to evaluate responsiveness of Prototype 2, because individuals with tinnitus that is not problematic are unlikely to be compliant with treatment and are therefore inappropriate for evaluating treatment-related improvement in tinnitus.

To increase the numbers of follow-up participants, OHSU joined with the other sites in acquiring follow-up data. Also, follow-up payments were increased to $20 to increase follow-up participation and were provided for CC, BPVA, and OHSU participants (again, without follow-up payments at JHVH).

Results

Prototype 2 Sample • Valid initial (intake) questionnaires were obtained from 347 of 350 participants who completed Prototype 2 questionnaires (see Table 4). Of Prototype 2 participants, 8.9% reported they were minority or mixed race, with 3.5% indicating they were Hispanic. The gender distribution was 82% male, and the average age was 60 yr.

Retest data were provided by 37 participants. Follow-up data for Prototype 2 were provided by 155 participants at 3 mo and 85 at 6 mo (see Tables 3 and 4). We found higher tinnitus severity levels for the Prototype 2 sample than for Prototype 1, as intended, as a result of recruitment efforts to limit inclusion of patients with mild tinnitus (see Supplemental Digital Content 8, Table B1, http://links.lww.com/EANDH/A60).

Results for the 30-Item TFI Prototype 2

Prototype 2 Descriptive Statistics • Using a new sample of 347 tinnitus patients, item-level frequency distributions for initial data from the 30-item Prototype 2 were examined for skewed distributions to identify potential floor or ceiling effects and missing data. Initial mean scores for all 30 items fell within the range 3.8 to 7.2 on the 0- to 10-point scale. As before, there were few missing values.

We computed an overall score for Prototype 2 and estimated reliability and validity statistics that could be used for comparison with reliability and validity results of the 30-item TFI obtained with the Prototype 1 sample. Because of decreasing the proportion of Prototype 2 participants reporting their tinnitus was not a problem, the overall mean scores for all three outcome variables were 4 to 8 points higher than in Stage 2 results (see Table 5). Again, the mean TFI (53.8 $\pm$ 25.0, N = 347) fell between the mean values of the THI (46.0 $\pm$ 26.2, N = 346) and the severity VAS (61.7 $\pm$ 23.8, N = 335). Median values for the three scales were close to the means: 56.0 (TFI Prototype 2), 42.9 (THI), and 65.0 (severity VAS).

Prototype 2 Validity Regarding Internal Structure • For Prototype 2, we considered doing a confirmatory factor analysis using a higher order model but decided against it because creating Prototype 2 entailed making important changes to the overall questionnaire. We deleted nearly one-third of the Prototype 1 items, changed the ordering of items in the questionnaire, and changed the wording of a response option for one item. Instead, we again chose to use exploratory factor analysis to test whether the same factor structure was observed in the new sample using the new questionnaire.

Factor analysis of Prototype 2 initial responses employed techniques identical to those used with Prototype 1 (Principal Components Analysis followed by Principal Axis Factoring, both models using orthogonal followed by oblique rotation). The same eight factors were again found for Prototype 2 and confirmed that despite its reduced number of items, Prototype 2 retained the comprehensive coverage of tinnitus impact found for Prototype 1. Once again, the factor structure was clearest when the factor analysis omitted individuals with mild tinnitus, retaining only those participants whose tinnitus was definitely problematic (tinnitus described as a moderate problem, a big problem, or a very big problem), yielding a total of 288 participants for the analysis. (See Supplemental Digital Content 11, Table C1, http://links.lww.com/EANDH/A63.)

Prototype 2 Validity Regarding Responsiveness • The overall TFI effect size for the Improved group was 0.83 at 3 mo and 1.46 at 6 mo. These Improved effect sizes for the overall TFI were somewhat larger than the effect sizes for the overall THI, which were 0.56 at 3 mo and 1.22 at 6 mo. For the severity VAS, the Improved effect size of 0.83 was identical to that for Prototype 2 at 3 mo but substantially smaller at 6 mo (0.80 for VAS compared with 1.46 for Prototype 2). Effect sizes for the eight factor-based subscales of Prototype 2 did not differ substantially from those obtained with the 30-item preliminary results obtained using Prototype 1 data.

TABLE 7. Results of an eight-factor solution using principal-axis factor analysis with oblique rotation of 25 items from Prototype 2 that were retained for the final TFI

TFI Item Content (Prototype 2 Item Number)	Mean	SD	Factor Loadings From the Pattern Matrix: 1 Cognitive	2 Auditory	3 Intrusive	4 Sleep	5 Relaxation	6 Quality of Life	7 Emotional	8 Sense of Control	h^2
TFI 9: Think clearly	5.31	2.84	**0.94**	0.01	−0.01	−0.01	−0.03	−0.01	0.01	0.01	0.88
TFI 8: Concentrate	5.85	2.59	**0.85**	0.06	−0.01	−0.01	0.12	0.05	−0.03	−0.02	0.90
TFI 10: Focus attention	5.33	2.71	**0.78**	0.00	0.05	−0.05	0.01	0.01	0.10	0.01	0.84
TFI 12: Understand people	5.70	3.01	0.00	**1.01**	−0.04	−0.02	−0.04	−0.04	0.02	0.02	0.97
TFI 11: Hear clearly	5.83	2.98	0.01	**0.92**	0.08	0.03	−0.01	−0.05	0.06	0.01	0.89
TFI 13: Follow conversations	5.99	3.11	0.01	**0.91**	−0.03	0.00	0.04	0.10	−0.07	−0.03	0.86
TFI 1R: Recoded % aware 0–10	7.74	2.50	0.01	0.03	**0.81**	−0.05	0.02	0.04	0.01	−0.03	0.73
TFI 25R: Recoded % annoyed 0–10	6.92	2.80	−0.01	0.03	**0.51**	−0.18	0.03	0.18	0.09	0.17	0.75
TFI 2: Strong or loud	7.22	2.03	0.13	0.12	**0.35**	−0.04	0.18	−0.15	0.03	0.25	0.59
TFI 27: As much sleep	5.84	3.45	−0.03	0.02	0.02	**−0.99**	−0.05	0.00	0.05	−0.05	0.95
TFI 28: Keep from sleeping as deep-peaceful	5.74	3.46	−0.01	0.00	0.01	**−0.95**	0.01	0.00	−0.04	0.01	0.86
TFI 26: Difficult fall asleep-stay asleep	6.26	3.37	0.06	−0.03	0.01	**−0.81**	0.10	−0.02	0.01	0.02	0.82
TFI 15: Quiet resting activities	6.67	2.66	0.08	0.05	0.01	−0.07	**0.71**	0.05	0.12	0.02	0.86
TFI 17: Enjoy peace and quiet	7.26	2.71	−0.02	−0.01	0.16	−0.04	**0.68**	0.08	0.08	0.11	0.84
TFI 16: Ability to relax	6.37	2.77	0.15	0.00	−0.11	−0.19	**0.61**	0.06	0.10	0.11	0.88
TFI 19: Enjoy social activities	5.61	3.14	0.01	0.17	0.04	−0.11	0.11	**0.64**	0.09	−0.02	0.81
TFI 20: Enjoyment of life	5.83	2.89	0.16	−0.02	0.11	0.02	0.21	**0.59**	0.07	0.06	0.80
TFI 7: Relationships family-friend	5.24	2.98	**0.31**	0.09	0.10	−0.03	−0.17	**0.36**	0.16	0.15	0.69
TFI 29: Difficulty perform work or tasks	4.78	3.18	0.29	0.05	−0.01	−0.13	−0.06	**0.35**	0.13	0.12	0.65
TFI 21: Anxious or worried	5.66	3.09	0.05	0.00	0.07	0.04	0.09	−0.04	**0.92**	−0.10	0.87
TFI 23: Bothered or upset	6.18	3.11	−0.01	0.03	0.04	−0.03	0.04	0.03	**0.77**	0.11	0.82
TFI 24: Depressed	4.52	3.30	0.02	0.01	−0.12	−0.12	−0.04	0.10	**0.73**	0.10	0.77
TFI 3: Feel in control	7.40	2.85	−0.02	0.02	−0.04	0.01	0.05	0.04	0.04	**0.65**	0.47
TFI 6: Easy to ignore	7.55	2.42	0.09	0.03	0.27	−0.05	0.07	−0.06	0.00	**0.51**	0.63
TFI 5: Easy to cope	6.15	2.36	0.26	−0.05	0.07	−0.08	0.05	0.03	0.13	**0.47**	0.74

The analysis was computed using 283 Prototype 2 participants who reported that their tinnitus problem was a moderate problem, a big problem, or a very big problem.
There were 11 additional participants who reported that their tinnitus problem was a moderate, big, or very big problem, but were excluded from the analysis because they were missing responses to more than 10% (three or more items) of the final 25-item TFI. The factor analysis was then computed using mean substitution of missing data for the 283 participants.
Bold values indicate factor loadings whose absolute values are 0.30 or greater. Table entries are factor loadings from the pattern matrix and are standardized regression coefficients (beta weights) for predicting each item from the eight factors. The percentage of variance explained by the eight factors was 79.5%.
TFI, Tinnitus Functional Index; h^2, communality (percentage of variance in each item explained by the eight factors).

Prototype 2 Reliability and Validity for Scaling of Tinnitus Severity • Internal consistency reliability for Prototype 2 remained very high (coefficient alpha = 0.98). Test–retest reliability (r = 0.76) was good, although somewhat lower than 0.91 found for Prototype 1. Possible reasons for the lower test–retest correlation are (a) the reduced number of items in Prototype 2, (b) the elimination from Prototype 2 of Prototype 1 items that were found relatively insensitive to change (i.e., were more stable over time) in Stage 2 results, and (c) the restriction of range resulting from a sample having fewer participants with mild tinnitus.

Excellent convergent validity was again found, with high correlations between participants' Prototype 2 initial scores and the corresponding scores for the THI (r = 0.87) and severity VAS (r = 0.75). Discriminant validity was good, with a moderate correlation of 0.57 with the BDI-PC. Overall Prototype 2 initial scores were again significantly associated with participants' responses to the global severity item concerning the extent of their "problem" tinnitus (Table 5).

Conclusions Regarding the 30-Item TFI Prototype 2 • Despite reducing the length of the TFI to 30 items, Prototype 2 performed well in a new sample in regard to measurement properties, including retaining a consistent and clear factor structure, displaying high internal consistency reliability and good test–retest reliability, and showing strong construct validity for scaling of tinnitus severity. For evaluating treatment outcomes, Prototype 2 exhibited moderately high responsiveness at 3 mo and high responsiveness at 6 mo. We were therefore encouraged to proceed with reducing the TFI length while retaining at least three items per subscale.

Selection of 25 Items for the Final TFI • As before, the goal was to select the best-functioning items for inclusion in the final TFI. After careful evaluation and discussion of the Prototype 2 item set, five items were removed: discomfort caused by tinnitus, interference of tinnitus with participation in social events, interference of tinnitus with leisure activities, fatigue caused by tinnitus, and amount of time that overall quality of life was reduced by tinnitus.

The final TFI has eight subscales (Intrusive, Sense of Control, Cognitive, Sleep, Auditory, Relaxation, Quality of Life, and Emotional), with three items each for seven of the subscales and four items for the Quality of Life subscale. All analyses used for evaluating Prototype 2 were repeated for the 25-item Final TFI using the data obtained with the Prototype 2 sample.

Results for the Final 25-Item TFI • On the basis of the responses to the 25 items (taken from participants' responses to the 30-item Prototype 2), the 25-item TFI questionnaire was found to retain satisfactory reliability and validity. Table 5 summarizes its properties in comparison with those of the 30-item Prototype 2.

Final TFI Validity Regarding Internal Structure • Table 7 shows the pattern matrix for the oblique Principal Axis Factoring model when applied to the 25-item TFI. It can be seen that the shorter version performed well, generally duplicating the results obtained with Prototype 2. So that readers can see the difference between doing the analyses with all tinnitus problem levels versus moderate or greater tinnitus problems, we have included the factor pattern matrix for the final 25-item TFI for both samples (n = 336 for all tinnitus problem levels and N = 283 for moderate, big, and very big tinnitus problem levels). For the former group (n = 336) results, see Supplemental Digital Content 12, Table C2, http://links.lww.com/EANDH/A64. For the latter group (n = 283) results, see Table 7.

Including all participants in the factor analysis provided a larger sample size and resulted in more explained variance. This larger variance, however, occurred in large part because participants whose tinnitus was either not a problem or a small problem tended to answer most TFI items with a 0, 1, or 2 on the 0 to 10 scale; this bolus of consistently low item scores resulted in increased correlations among items and a less clear pattern of factor loadings. In contrast, computing the factor analyses with participants reporting that their tinnitus was A moderate problem or worse resulted in clearer patterns of factor loadings, more distinct factors, and smaller correlations among factors.

Final TFI Validity Regarding Responsiveness • Good response resolution in the final TFI was shown by the lack of extreme item means but relatively large item SD values. The 25 item means ranged from 3.8 to 7.2 (on the 0 to 10 scale), with item SD values ranging from 2.3 to 3.7. Four of the 25 items did display a mild ceiling effect, in that 25 to 34% of patients endorsed the most severe response of 10. Because our aim was to measure improvement of tinnitus resulting from treatment, and not to measure the worsening of tinnitus, we did not have a high level of concern about ceiling effects. In fact, we viewed such items as desirable for detecting improvement because it is tinnitus patients with moderate to high severity who are most likely to need and seek treatment. Supplemental Digital Content 13, Table C3, http://links.lww.com/EANDH/A65, summarizes each of the 25 items in the final TFI the percentage of patients giving each response (0 to 10) and the percentage missing. Missing data fell below 3% for all items.

In estimating effect sizes for the TFI and its subscales, we decided that we had adequate group sizes (see Table 4) to collapse the seven-level perceived change item (Much improved to Much worse) into a five-level perceived change variable. The Much improved and Moderately improved levels were combined, as were the Much worse and Moderately worse levels. Table 8 lists the 3 and 6 mo effect sizes for the overall TFI and its subscales and for the VAS, THI, and THI subscales. Effect sizes were moderate to high at 3 mo and even higher at 6 mo. At 3 mo, the TFI and VAS exhibited similar effect sizes, but the TFI had larger negative effect sizes at 6 mo. Overall, effect sizes for the Improved groups were somewhat larger for the TFI than for the THI, but the Worse groups had slightly larger negative effect sizes for the THI compared with the TFI. Figure 2 displays the effect size estimates for the overall scales (final TFI, VAS, THI) at 3 and 6 mo.

Final TFI Validity Regarding Scaling of Tinnitus Severity • As with Prototypes 1 and 2, the final TFI exhibited strong evidence for scaling of tinnitus severity. Supplemental Digital Content 14, Table C4, http://links.lww.com/EANDH/A66, presents reliability, descriptive statistics, convergent validity, and discriminant validity for the final TFI and its subscales. Missing data percentages were 0% for the overall TFI and 0.3 to 1.7% for the subscales.

Using the New TFI

Clinical Significance of TFI Scores for Determining Severity of Tinnitus • For evaluating the severity of tinnitus in individuals at intake (e.g., for staging or for screening purposes), we stratified TFI scores using participants' responses to the global severity item, "How much of a problem is your tinnitus?" We intentionally used this item to stratify our sample for the purpose of comparing mean TFI scores across the five levels: not a problem, a small problem, a moderate problem, a big problem, and a very big problem. Comparing the mean TFI scores (and TFI score distributions) of patients across these five problem levels allows clinicians to have a practical sense of how to interpret TFI scores. For example, on the 25-item TFI, the mean scores were 14, 21, 42, 65, and 78 for the five successive levels of the "problem" responses.

Figure 3 shows the frequency distributions of the 25-item TFI scores for each of the five response levels, as well as the mean and SD for each level. The modal TFI values were in the range of 10 to 20 for the Small problem group, 30 to 40 for the Moderate problem group, 40 to 60 for the big problem group, and 60 to 90 for the very big problem group. These results provide preliminary support for classifying TFI scores below a value of about 25 as indicating relatively mild tinnitus, typically with little or no need for intervention. TFI scores from about 25 to about 50 would suggest more significant problems with tinnitus, indicating a possible or borderline need for professional attention (including appropriate advice such as avoiding worsening of the tinnitus by exposure to loud noise). Last, TFI scores above about 50 are likely to indicate tinnitus severe enough to qualify for more aggressive efforts to provide relief, possibly involving referral to specialty tinnitus care.

Minimum Clinically Important Change in TFI Scores • This is an important topic and one that has generated substantial debate among measurement experts (Norman et al. 2003; Terwee et al. 2003). One major issue is that there are considerable individual differences between patients in regard to what they consider a meaningful change in their tinnitus. Another issue is that statistical demonstrations of differences between treatment groups are not necessarily indicative of changes that patients consider important or meaningful, particularly when

TABLE 8. Effect size estimates for the overall scales (final 25-item TFI, VAS, THI) and subscales (TFI, THI) at 3 and 6 mo follow-up

Scales and Subscales	Perceived Change in Overall Tinnitus Condition From Initial Visit to 3 mo Follow-Up				
	Much or Moderately Improved (n = 21–22)	Slightly Improved (n = 26–30)	Unchanged (n = 55–68)	Slightly Worse (n = 12–13)	Moderately or Much Worse (n = 13–17)
	3 mo effect size estimates				
TFI	1.01	0.74	0.29	0.14	−0.36
VAS	0.98	0.74	0.24	0.08	−0.93
THI	0.79	0.40	0.22	−0.04	−0.48
TFI subscales					
Intrusive	1.05	0.80	0.24	0.21	−0.51
Sense of Control	0.82	0.80	0.38	0.25	−0.88
Cognitive	0.57	0.55	0.24	0.13	−0.41
Sleep	0.50	0.49	0.17	0.14	−0.12
Auditory	0.71	0.49	0.10	−0.25	−0.01
Relaxation	1.19	0.72	0.28	0.07	−0.42
Quality of Life	0.70	0.66	0.26	0.09	−0.16
Emotional	1.14	0.62	0.33	0.20	−0.33
THI subscales					
Functional	0.56	0.27	0.15	−0.17	−0.48
Emotional	0.72	0.36	0.17	−0.04	−0.37
Catastrophic	1.18	0.69	0.41	0.28	−0.52

Scales and Subscales	Perceived Change in Overall Tinnitus Condition From Initial Visit to 6 mo Follow-Up				
	Much or Moderately Improved (n = 14–16)	Slightly Improved (n = 12–13)	Unchanged (n = 32–34)	Slightly Worse (n = 8–10)	Moderately or Much Worse (n = 11–12)
	6 mo effect size estimates				
TFI	1.88	1.19	0.14	0.14	−0.05
VAS	0.96	0.72	0.10	0.13	−0.05
THI	1.33	1.12	−0.02	0.27	−0.24
TFI subscales					
Intrusive	1.88	0.58	0.31	0.31	0.29
Sense of Control	1.79	0.66	0.30	0.29	0.06
Cognitive	1.29	1.27	0.11	−0.15	−0.13
Sleep	0.63	1.32	0.01	0.03	0.10
Auditory	0.89	−0.04	−0.07	0.04	0.01
Relaxation	2.09	1.26	0.06	0.16	−0.03
Quality of Life	1.37	1.20	0.16	0.15	−0.14
Emotional	1.81	1.11	0.12	0.17	−0.20
THI subscales					
Functional	0.89	0.85	−0.16	0.27	−0.30
Emotional	1.36	1.07	0.07	0.22	−0.14
Catastrophic	1.91	1.26	0.15	0.32	−0.25

Regarding missing data at 3 mo, 5 of the 155 participants providing 3 mo follow-up data did not answer the perceived-change item, so the total sample size used for this table was 150. In estimating effect sizes for each of the five perceived-change levels, no scale or subscale (except for the VAS) had more than one participant missing a score. For the VAS, 22 participants had a missing score, with the number missing in each perceived-change level as follows: 1 in much or moderately improved, 4 in slightly improved, 13 in unchanged, 0 in slightly worse, and 4 in moderately or much worse.

Regarding missing data at 6 mo, 1 of the 86 participants providing 6-mo follow-up data did not answer the perceived-change item, so the total sample size used for this table was 85. As with 3 mo, in estimating effect sizes for each of the five perceived change levels, no scale or subscale (except for the VAS) had more than one participant missing a score.

TFI, Tinnitus Functional Index; THI, Tinnitus Handicap Inventory; VAS, Visual Analog Scale.

the treatment groups are large enough to permit reliable detection of small differences in treatment outcomes.

What changes in tinnitus might participants in our study consider meaningful? We used Lipsey's criterion groups approach again, stratifying the follow-up results at 3 and 6 mo by the Global Perception of Change variable. The mean TFI change scores for the five change groups at 3 and 6 mo are shown in Figure 4, where it can be seen that the mean change scores exhibit an orderly progression from Much or Moderately improved through Unchanged to Moderately or Much worse.

Acknowledging the preliminary nature of these results from an observational study, we interpret the data shown in Figure 4 as suggesting that a reduction in TFI scores of around 13 points should be meaningful to patients—that is, at 3 mo, the score reduction for the "much to moderately improved" group (−21.1) compared with the reduction for the "unchanged" group (−7.2) is about 14 points, slightly larger than one-half of the SD observed for the initial scores of the overall group of 347 participants (SD = 24.7, see Table 5). Likewise, for the 6 mo follow-up data, the TFI reduction for the Slightly improved group (−20.9) compared with that for the Unchanged group (−3.4) is about 17 points, again somewhat larger than one-half the initial SD. In terms of effect sizes, these observed differences between Improved and Unchanged would represent values greater than 0.5, constituting moderate effect sizes using Cohen's terminology (Cohen

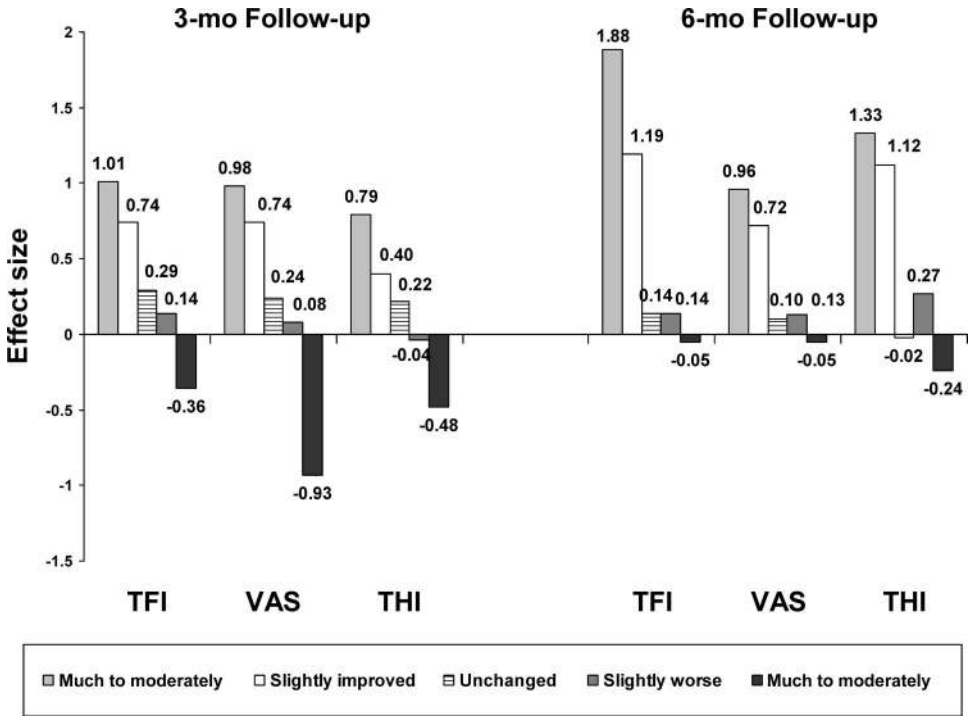

Fig. 2. Effect sizes for the final 25-item TFI, VAS, and THI at 3 and 6 mo follow-up. TFI, Tinnitus Functional Index; THI, Tinnitus Handicap Inventory; VAS, Visual Analog Scale.

1988). These estimates agree well with a recent meta-analytic review by Norman et al., in which they concluded "under many circumstances, when patients with a chronic disease are asked to identify minimal change, the estimates fall very close to half a SD" (Norman et al. 2003, p. 590). Pending further clinical observations using the TFI, which may help to refine our estimate, we suggest that a reduction of 13 points be used as the criterion for meaningful reduction in TFI outcome scores.

Use of TFI and THI Subscales • The effect-size estimates in Table 8 show that subscales differed considerably from one another, with some subscales exhibiting effect sizes that were two to three times larger than those of other subscales. An unexpected finding regarding subscale effect sizes was the large effect sizes for the THI Catastrophic subscale. An important difference between the TFI and THI is that the THI's domain of catastrophic responses was not included in the TFI. The Tinnitus Research Consortium, which funded this study, had requested that the TFI minimize negative ideation and exclude catastrophic items. One reason for omitting such items is that persons with mild tinnitus sometimes become concerned that such catastrophic feelings will be their fate eventually, creating a negative starting point for intervention. The five items on the THI Catastrophic subscale focus on patients' feelings of desperation, inability to escape from tinnitus, fear of having a terrible disease, loss of control, and inability to cope. Two of the three items on the new TFI Sense of Control subscale have

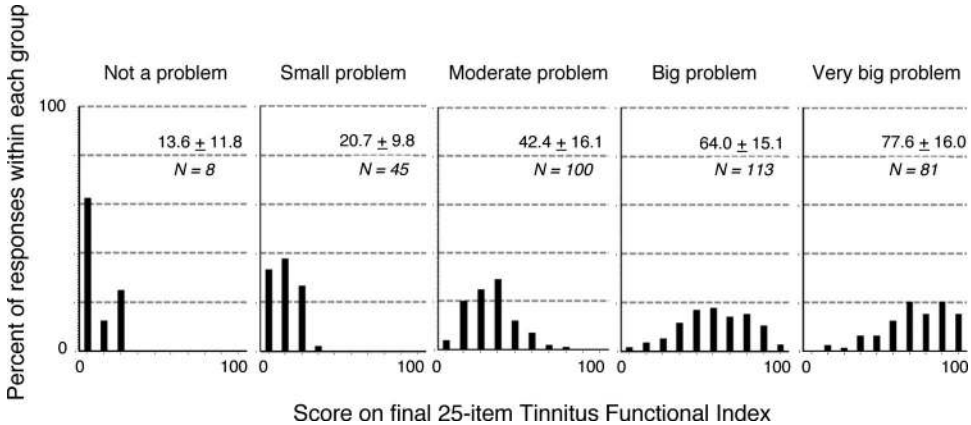

Fig. 3. Frequency distributions of overall TFI scores after stratification by responses to the item "How much of a problem is your tinnitus?" Horizontal axis shows overall TFI scores (in 10 bins ranging from 0 to 100 for each of the five small graphs) for each of five groups of participants identified by their responses to the tinnitus problem question, ranging from those responding "not a problem" to those responding "a very big problem." Vertical axis shows the percentage of participants' responses occurring in each bin of the five successive frequency distributions. Numbers at upper right corner in each graph display the group sizes together with the group mean TFI scores and SDs. TFI, Tinnitus Functional Index.

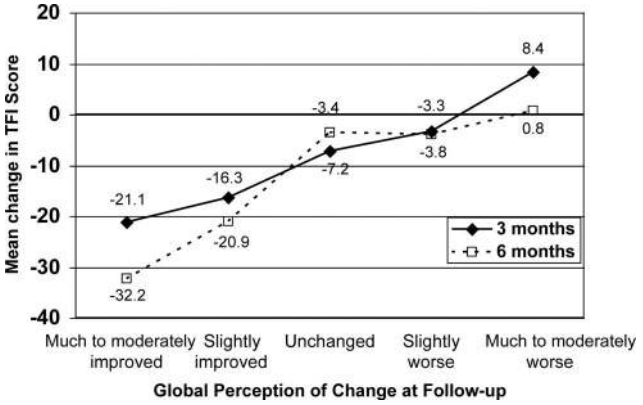

Fig. 4. Overall mean TFI change scores at 3 and 6 mo follow-up after stratification by participants' Global Perception of Change. Horizontal axis shows five levels of participants' Global Perception of Change. Vertical axis shows mean change in scores for overall TFI (follow-up score minus initial score) for each of the two follow-up intervals (3 and 6 mo). Using an analysis of variance to compare group means, both the 3 and 6 mo means displayed significant overall differences, resulting in F ratios of 10.23 at 3 mo and 12.96 at 6 mo ($p < 0.001$ for both). TFI, Tinnitus Functional Index.

some similarity to the two THI items covering control and coping. The two TFI items are not as extreme as the THI counterparts because their item stems are worded more neutrally and the 0 to 10 response options include a positive label at one end. The two TFI items are as follows: "Over the PAST WEEK, (1) Did you feel IN CONTROL in regard to your tinnitus?" 0 = very much in control, 10 = never in control and (2) "How easy was it for you to COPE with your tinnitus?" 0 = very easy to cope, 10 = impossible to cope.

To determine whether excluding catastrophic reactions from the TFI content resulted in a serious gap, we first correlated the THI Catastrophic subscale with the TFI and its subscales. The correlations were moderate, ranging from 0.40 to 0.73, suggesting that the Catastrophic subscale may measure a somewhat different severity domain than the TFI and its subscales. Like several of the TFI subscales with large effect sizes, the THI Catastrophic subscale may work particularly well as an outcome measure. Because the Catastrophic subscale had the largest effect sizes of the three THI subscales, it disproportionately affected the overall THI effect size (see Table 8). These additional results and comparisons point to the potential utility of examining both THI and TFI subscales in intervention trials.

The Final 25-Item TFI • To facilitate computation of participants' scores on the eight subscales (corresponding to the eight factors consistently identified by repeated factor analyses), the ordering of items within the questionnaire was revised slightly to place all items corresponding to a given subscale together. The 25 items of the final version of the TFI questionnaire were reformatted to fit two standard pages (see Fig. 5). The final questionnaire is also presented online, together with brief scoring instructions, all of which can be downloaded and printed (see Supplemental Digital Content 15, Figure C1, http://links.lww.com/EANDH/A67).

TINNITUS FUNCTIONAL INDEX

Today's Date _____ (Month / Day / Year) Your Name _____ (Please Print)

Please read each question below carefully. To answer a question, select *ONE* of the numbers that is listed for that question, and draw a *CIRCLE* around it like this: (10%) or (1).

I Over the PAST WEEK...

1. What percentage of your time awake were you consciously **AWARE OF** your tinnitus?
Never aware ► 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100% ◄ *Always aware*

2. How **STRONG** or **LOUD** was your tinnitus?
Not at all strong or loud ►0 1 2 3 4 5 6 7 8 9 10 ◄ *Extremely strong or loud*

3. What percentage of your time awake were you **ANNOYED** by your tinnitus?
None of the time ► 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100% ◄ *All of the time*

SC Over the PAST WEEK...

4. Did you feel **IN CONTROL** in regard to your tinnitus?
Very much in control ►0 1 2 3 4 5 6 7 8 9 10 ◄ *Never in control*

5. How easy was it for you to **COPE** with your tinnitus?
Very easy to cope ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Impossible to cope*

6. How easy was it for you to **IGNORE** your tinnitus?
Very easy to ignore ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Impossible to ignore*

C Over the PAST WEEK...

7. Your ability to **CONCENTRATE**?
Did not interfere ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Completely interfered*

8. Your ability to **THINK CLEARLY?**
Did not interfere ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Completely interfered*

9. Your ability to **FOCUS ATTENTION** on other things besides your tinnitus?
Did not interfere ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Completely interfered*

SL Over the PAST WEEK...

10. How often did your tinnitus make it difficult to **FALL ASLEEP or STAY ASLEEP**?
Never had difficulty ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Always had difficulty*

11. How often did your tinnitus cause you difficulty in getting **AS MUCH SLEEP** as you needed?
Never had difficulty ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Always had difficulty*

12. How much of the time did your tinnitus keep you from **SLEEPING** as **DEEPLY** or as **PEACEFULLY** as you would have liked?
None of the time ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *All of the time*

 08.15.08

TINNITUS FUNCTIONAL INDEX PAGE 2

Please read each question below carefully. To answer a question, select *ONE* of the numbers that is listed for that question, and draw a *CIRCLE* around it like this: (10%) or (1).

A	**Over the PAST WEEK, how much has your tinnitus interfered with...**	*Did not interfere*									*Completely interfered*
13.	Your ability to **HEAR CLEARLY**?	0	1	2	3	4	5	6	7	8	9	10
14.	Your ability to **UNDERSTAND PEOPLE** who are talking?	0	1	2	3	4	5	6	7	8	9	10
15.	Your ability to **FOLLOW CONVERSATIONS** in a group or at meetings?	0	1	2	3	4	5	6	7	8	9	10
R	**Over the PAST WEEK, how much has your tinnitus interfered with...**	*Did not interfere*									*Completely interfered*
16.	Your **QUIET RESTING ACTIVITIES**?	0	1	2	3	4	5	6	7	8	9	10
17.	Your ability to **RELAX**?	0	1	2	3	4	5	6	7	8	9	10
18.	Your ability to enjoy **PEACE AND QUIET** ?	0	1	2	3	4	5	6	7	8	9	10
Q	**Over the PAST WEEK, how much has your tinnitus interfered with...**	*Did not interfere*									*Completely interfered*
19.	Your enjoyment of **SOCIAL ACTIVITIES**?	0	1	2	3	4	5	6	7	8	9	10
20.	Your **ENJOYMENT OF LIFE**?	0	1	2	3	4	5	6	7	8	9	10
21.	Your **RELATIONSHIPS** with family, friends and other people?	0	1	2	3	4	5	6	7	8	9	10

22. How often did your tinnitus cause you to have difficulty performing your **WORK OR OTHER TASKS**, such as home maintenance, school work, or caring for children or others?
Never had difficulty ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Always had difficulty*

E Over the PAST WEEK...

23. How **ANXIOUS** or **WORRIED** has your tinnitus made you feel?
Not at all anxious or worried ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Extremely anxious or worried*

24. How **BOTHERED** or **UPSET** have you been because of your tinnitus?
Not at all bothered or upset ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Extremely bothered or upset*

25. How **DEPRESSED** were you because of your tinnitus?
Not at all depressed ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Extremely depressed*

 08.15.08

Fig. 5. Final version of the TFI questionnaire.

DISCUSSION

The results of this study provide evidence that the 25-item TFI achieved the three original design goals: comprehensive coverage of a broad spectrum of tinnitus-related problems, high responsiveness to treatment-related changes in tinnitus, and excellent properties for scaling the severity of tinnitus for use in intake assessment. In addition, this work refined the definition of the construct being measured (the severity and negative impact of tinnitus) by elucidating eight factor-based subscales that provide an expanded set of measures for evaluating the severity of tinnitus at intake and for evaluating treatment effects. Moreover, participants completing the TFI seemed to do so with little if any difficulty.

Relation to Previous Work

It is important to acknowledge the dependence of the TFI on nine previously existing tinnitus questionnaires (cited in Table 1), all of which were published within the span of 12 yr (1988–1999). Because these questionnaires were used as the primary starting point for our potential pool of items, there is content overlap between the TFI and other measures. Like these measures, the TFI has been documented for internal consistency, test-retest reliability, and convergent and discriminant validity.

There are also notable differences between the TFI and other measures. The primary difference is that the TFI was developed with a systematic focus on responsiveness, which involved the following: (a) content validity evaluation by an expert panel of 17 judges to identify potentially responsive items; (b) use of a 0 to 10 response option to achieve fine discrimination for change; (c) quantitative testing of TFI Prototypes 1 and 2 using large samples of tinnitus patients from multiple geographic locations; and (d) selection of items shown to have strong effect sizes for detecting change.

The THI is one of the best known and most widely used tinnitus instruments, with well-established reliability and construct validity (Newman et al. 1996, 1998). For these reasons, we chose it as a criterion measure for establishing convergent validity for the TFI. The key advantages of the TFI compared with the THI are its extensive emphasis on content validity in item selection, its 0 to 10 response format (versus the THI's three-level format), greater responsiveness, and eight subscales (versus three for the THI). If the TFI's greater responsiveness is upheld in clinical trials, it could result in future trials needing fewer patients to achieve statistical significance.

The nine questionnaires that provided content for the TFI constitute a valuable body of expert opinion on the basis of substantial clinical experience working with tinnitus patients. The use of different questionnaires across studies, however, has made it difficult to compare results and to conduct meta-analyses—ultimately impeding progress in tinnitus outcomes research. The measurement approach used in developing the TFI and the encouraging results obtained suggest that the TFI can serve as a valuable tool for facilitating understanding of tinnitus and for refining treatments. It is our hope that the TFI will contribute toward widespread use of quantitative tinnitus assessment techniques that achieve a level of standardization comparable to that already achieved for audiometric measures. As with any new questionnaire, however, researchers conducting clinical trials should consider using two or more tinnitus outcome measures until there is more definitive evidence about the relative responsiveness and utility of existing scales, including the new TFI.

Development of Eight Subscales

In its Request for Applications, the funding agency (Tinnitus Research Consortium) emphasized that the new questionnaire should cover the important domains of negative tinnitus impact. The agency specified 10 domains that were required to be included in the yet-to-be-developed TFI. The new questionnaire was intended to be comprehensive and not omit any important domains of negative functional impact as a result of tinnitus. According to recent advances in measurement science regarding the development of responsive outcome measures for intervention trials, it is important to measure each of the key domains with at least three items, if possible. Moreover, the availability of reliable but brief subscales will allow clinicians and researchers to obtain a preliminary view of how patients are faring in those separate domains.

In addition to the 10 domains specified by the granting agency, we identified three more domains during the content validity evaluation, resulting in a total of 13 domains reflected in the 43-item Prototype 1. In the factor analysis of Prototype 1 data, eight of the original 13 domains were corroborated as separate factors, with items from four of the remaining five domains loading on these factors. One of those 13 domains (Somatic) had weak factor analysis results and low mean values and eventually was excluded from the final 25-item TFI.

For Prototype 2, we had considered doing a confirmatory factor analysis using a higher order model. We refrained, as described earlier, because creating Prototype 2 entailed making important changes to the overall questionnaire. Because of those changes, we did not feel it was appropriate to use the confirmatory approach. To address the concepts relevant to a future confirmatory analysis, we have conducted exploratory analyses including TFI subscale intercorrelations and factor analysis of the eight TFI subscales. We used results from these analyses to generate hypotheses about the higher-order internal structure of the TFI. In particular, we hypothesized a model of internal structure for the TFI that is composed of eight first-order domain-specific factors and a general second-order tinnitus severity factor.

It is noteworthy that, using the eight TFI subscales as variables and checking intercorrelations between them, a fairly clear pattern of low to moderate correlations of the Auditory subscale was observed with the other seven subscales (r = 0.24–0.48, median r = 0.32). In contrast, the other seven subscales exhibited moderate to moderately high intercorrelations (r = 0.51–0.80, median r = 0.65). (See Supplemental Digital Content 16, Table C5, http://links.lww.com/EANDH/A68, for subscale intercorrelations.) We then did a principal components factor analysis of all eight subscales, with oblique rotation, which provided evidence for two possible structures. In the first, there is a general tinnitus severity factor underlying all eight subscales. In the second, there is a general tinnitus severity factor underlying seven of the eight subscales, with the Auditory subscale representing an underlying specific factor. From our exploratory analyses, we think that four other subscales (Cognitive, Sleep, Relaxation, Sense of Control) may also represent specific factors. However, we are not sure

whether the remaining three subscales (Intrusive, Quality of Life, Emotional) reflect only a general tinnitus severity factor or both general and specific factors.

We continue to think that it is important to retain the TFI subscale, Auditory, that measures hearing difficulty attributed to tinnitus, but we also think that its underlying dimension may be of a different flavor compared with the other seven subscales. Tinnitus and hearing difficulty may interact in two key ways. First, for some patients who feel that their tinnitus interferes with their hearing, it is the impaired hearing, not the tinnitus, that is the primary problem needing intervention (Coles 1995; Zaugg et al. 2002; Dobie 2004b). Second, for a smaller subset of patients, tinnitus may actually interfere with hearing sensitivity (e.g., a car mechanic listening for engine sounds, a doctor listening for heart sounds). The Auditory subscale enables this important aspect of tinnitus disturbance to be assessed.

During our factor analysis activities, another topic receiving considerable attention was the issue of underextraction versus overextraction, with the former usually viewed as leading to more misinterpretation of factor structure than the latter. We erred on the side of extracting more factors, especially because we began with some relatively well-specified domains of tinnitus severity. At this point, we look forward to doing a confirmatory factor analysis as well as a bifactor analysis (Reise et al. 2007), which may be even better at allowing us to determine whether one general factor underlies the 25-item TFI, with some subscales also having a specific factor and some subscales not.

The eight-factor solution obtained in the exploratory factor analyses for the final 25-item TFI (as well as Prototypes 1 and 2) accords quite well with content included in the 10 tinnitus impact domains listed by the funding agency in its Request for Applications and with the 13 domains identified in the content validity results from Stage 1. That earlier tinnitus questionnaires had not previously identified these eight factors as such may be because no previous questionnaire appears to have involved factor analytic techniques similar to those used here. In addition, no previous questionnaire has systematically included the recommended minimum of three to four items for each of the eight domains, as was done for the TFI.

It is possible that the eight factors of the final 25-item TFI can provide useful diagnostic information by serving as subscales for evaluating the specific profile of tinnitus-related problems characterizing any given patient. If so, the eight subscales may contribute to further standardization of tinnitus measures both for diagnostic assessment and outcomes evaluation. The fact that most of these eight factors or subscales displayed moderate to large effect sizes at 3 months and even larger effect sizes at 6 months suggests that some or all of the subscales may prove useful as subsidiary outcome measures in studies of treatment efficacy.

Limitations of the Project

Given the resources available, use of a controlled clinical trial to develop and test the TFI was not feasible. Instead, we used an approach recommended by Lipsey (1983, 1990), in which nonexperimental data are used to evaluate the potential effect sizes that candidate outcome measures (i.e., the TFI, THI, and VAS in our study) might exhibit in a future intervention trial. We used the approach from a published pain study as a model for testing our instrument (Clark et al. 2003). In particular, we used an observational approach in which patients who were measured at 3 and 6 mo follow-up provided self-reports of treatments received and estimated how their tinnitus had changed (Improved, Unchanged, Worsened) since their initial questionnaire.

Because this research was observational rather than experimental, no attempt was made to control for individual differences between participants in regard to age, hearing status, or tinnitus duration and etiology. Furthermore, despite the use of a multisite design incorporating widely separated patient groups in Oregon, Ohio, and Florida, the participants in the present study did not represent much ethnic or racial diversity. Further work is therefore needed to evaluate demographic or other variables that may affect the sensitivity of the TFI to treatment-related change. For example, it is possible that use of the TFI in other populations or with other treatments not evaluated in the present study may reveal additional content domains or items that could or should be investigated.

The present study evaluated TFI changes using a mixed treatments design—that is, every study site applied its own standard of care, which varied from site to site and from patient to patient. Differences between the various treatments may have led to increased variance in regard to treatment-related improvements in tinnitus, possibly restricting the range of effect sizes observed in the Improved group. Despite that possibility, this study demonstrated that the 25-item TFI exhibited clear, significant changes in participants who reported their tinnitus was improved at 3 and 6 mo follow-up (contrasting with small or negative TFI effect sizes in participants whose tinnitus was unchanged or worse at follow-up). Further work is now needed to evaluate the use of the TFI in controlled treatment trials. Assuming that effective treatment is employed, such studies may demonstrate even larger improvement-related effect sizes than were found in the present mixed treatment design.

Gender Representation • While one of the study's strengths is its multisite design, 53% of the Prototype 1 sample and 62% of the Prototype 2 sample came from VA sites, who are predominantly male. VA participants included 98.3% males in Prototype 1 and 95.8% males in Prototype 2. Non-VA participants included 62.3% males in Prototype 1 and 61.4% males in Prototype 2. Although veterans may have a significantly greater risk of chronic tinnitus than nonveterans (Hoffman & Reed 2004; Folmer et al. 2011), the disproportionate inclusion of veterans in our study resulted in the overall sample not being representative of the gender distribution of patients with tinnitus. We think that the gender distribution in the non-VA sites is likely closer to the gender distribution of the tinnitus patient populations in the United States and Norway, as described by Hoffman and Reed (2004, p. 27): "males are nearly 50% more likely to have reported chronic tinnitus or bothersome tinnitus" compared with females. Furthermore, analysis of variance confirmed that men and women in our study did not differ significantly in their average scores on the TFI, THI, and severity VAS (p = 0.38, 0.08, and 0.41, respectively).

Validity of the Final 25-Item TFI • Data from the sample of 347 patients who completed the 30-item Prototype 2 were used to compute reliability and validity statistics on the 25-item "final" version of the TFI. (The selection of 30 items for Prototype 2

provided a cushion in case some items did not continue to be as strong as in Prototype 1. When reducing Prototype 2 from 30 items to 25 for the final TFI, we were balancing a recommendation from the funding agency to produce a 12- to 22-item measure with our own preference for a somewhat longer scale. It is possible that these reliability and validity statistics could have differences compared with any future results from a new sample of patients completing the 25-item TFI. Large differences would be unlikely, however, because the general format of the 25-item TFI is nearly identical to TFI Prototypes 1 and 2. The order of items within subscales is very similar across the 30-item TFI Prototype 2 and the final 25-item TFI, although the order of subscales has been changed somewhat. Items were worded identically across prototypes, with the exception of three minor changes from Prototype 1 to Prototype 2 (response anchors for one item were changed, a four-word phrase in lowercase was capitalized, and the word "as" was added to one item for clarity). Because the samples used for Prototypes 1 and 2 were relatively large, we expect that results for the 25-item TFI with a new sample will be similar in terms of response distributions, missing data, reliability, validity, and responsiveness.

CONCLUSIONS

The potential benefits to be derived from developing a "core set" of measures have been emphasized for tinnitus research and clinical practice (Axelsson 1992; Meikle & Griest 2002) as well as in many other health contexts (Tugwell & Boers 1993; van der Heijde et al. 1997; Cox et al. 2000; Turk et al. 2003). The major components of all core measures include (a) identification of widely accepted measures (e.g., clinical tests or observations, patients' self-reported problems); (b) development of expert consensus regarding the most beneficial and cost-effective of the various measures available; and (c) implementation of a general agreement among clinicians and researchers to use the core set of measures routinely, to facilitate meta-analyses and other comparisons between different clinical trials or research studies.

In the present research, component (a) was addressed by enlisting collaborative efforts from a diverse group of clinicians and researchers in health-related disciplines with collective expertise in tinnitus measurement, diagnosis, and treatment (the authors of this article) and using their combined judgments to select the domains and items to be included in the TFI. To address component (b), this research provided evidence concerning the effect sizes obtained in clinical use of the TFI and its eight subscales. This evidence suggests that the TFI may contribute to cost-effectiveness of research by providing relatively large effect sizes when used in clinical trials. These are clearly only the first steps in developing more widespread consensus regarding use of the TFI as one part of a core set of measures. To address component (c), we hope that other investigators will join with us in testing the use of the TFI more widely in a variety of clinical and research settings.

Standardizing outcome measures would facilitate comparisons across treatments and allow for meta-analyses summarizing results across treatment. We hope that the present results will facilitate the provision of evidence-based treatment decisions for people with tinnitus.

ACKNOWLEDGMENTS

The authors thank the Tinnitus Research Consortium for providing the principal funding support for this research. They also acknowledge the following institutions for providing substantial support in the form of clinical and administrative facilities: The Oregon Health & Science University, Portland, Oregon; the Cleveland Clinic, Cleveland, Ohio; the Bay Pines VA Healthcare System, Bay Pines, Florida; the James A. Haley Veterans' Hospital in Tampa, Florida; the VA National Center for Rehabilitative Auditory Research, Portland VA Medical Center, Portland, Oregon; and the Hearing & Speech Institute, Portland, Oregon.

In the Request for Applications for the grant supporting this study, helpful suggestions were provided by the Tinnitus Research Consortium concerning domains of tinnitus impact associated with chronic tinnitus. The authors also received helpful comments during preparation of this report from Robert Dobie, MD, Department of Otolaryngology, University of California at Davis; James B. Snow, Jr., MD, Professor Emeritus of Otorhinolaryngology, University of Pennsylvania; and P. Ashley Wackym, MD, Balance and Hearing Center Northwest, Portland, OR (formerly at Department of Otolaryngology and Communication Sciences, Medical College of Wisconsin). Susan Meikle Mandell, BA, is specially appreciated for editing the final manuscript.

The authors thank the following individuals who provided timely, accurate, and complete acquisition of participants' questionnaires and responsible handling of patient confidentiality procedures: MaryCatherine Corrigan and Lisa Oswald, AuD (Cleveland Clinic); Monica Mejia, MS, and Danielle DeBiase (Bay Pines VA Medical Center); Deirdre Hamaker and Beth Priest (James A. Haley Veterans' Hospital); and Gail Oakes and Ann Tucker, MS (Oregon Health & Science University).

The authors thank Thomas Creedon for designing and maintaining the interactive website that provided essential support for all phases of this research, including the multiple web-based activities of the Item Selection Panel; for analyzing the data resulting from the panel's work; and for creating web tools that facilitated error-free tracking and logging-in of all questionnaires mailed from participating sites to the core facility at the Oregon Health & Science University.

This work was supported by a research grant from the Tinnitus Research Consortium. Administrative support and facilities were provided by Oregon Health & Science University, Portland, Oregon; Cleveland Clinic, Cleveland, Ohio; Bay Pines VA Healthcare System, Bay Pines, Florida; James A. Haley Veterans' Hospital, Tampa, Florida; VA National Center for Rehabilitative Auditory Research, Portland, Oregon; and Hearing and Speech Institute, Portland, Oregon.

The contents of this report do not represent the views of the Department of Veterans Affairs or the U.S. Government.

Address for correspondence: James A. Henry, PhD, VA Medical Center, National Center for Rehabilitative Auditory Research, PO Box 1034, Portland, OR 97207, USA. E-mail: james.henry@va.gov.

Received February 5, 2010; accepted July 19, 2011.

REFERENCES

Axelsson, A. (1992). Conclusion to Panel Discussion on Evaluation of Tinnitus Treatments. In J.-M. Aran & R. Dauman (Eds). *Tinnitus 91. Proceedings of the Fourth International Tinnitus Seminar* (pp. 453–455). New York, NY: Kugler Publications.

Beck, A. T., Guth, D., Steer, R. A., et al. (1997). Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behav Res Ther*, *35*, 785–791.

Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. *Am Psychol*, *29*, 512–518.

Castle, N. G., & Engberg, J. (2004). Response formats and satisfaction surveys for elders. *Gerontologist*, *44*, 358–367.

Clark, M. E., Gironda, R. J., Young, R. W. (2003). Development and validation of the Pain Outcomes Questionnaire-VA. *J Rehabil Res Dev*, *40*, 381–395.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Coles, R. R. A. (1995). Classification of Causes, Mechanisms of Patient Disturbance, and Associated Counseling. In J. A. Vernon & A. R. Moller (Eds). *Mechanisms of Tinnitus* (pp. 11–19). Needham Heights, MA: Allyn & Bacon.

Coles, R. R. A., Lutman, M. E., Axelsson, A., et al. (1992). Tinnitus Severity Gradings: Cross-Sectional Studies. In J.-M. Aran & R. Dauman (Eds). *Tinnitus 91. Proceedings of the Fourth International Tinnitus Seminar* (pp. 453–455). New York, NY: Kugler Publications.

Cox, R., Hyde, M., Gatehouse, S., et al. (2000). Optimal outcome measures, research priorities, and international cooperation. *Ear Hear*, *21(4 Suppl)*, 106S–115S.

Dobie, R. A. (2004a). Clinical Trials and Drug Therapy for Tinnitus. In J. B. Snow (Ed). *Tinnitus: Theory and Management* (pp. 266–277). Lewiston, NY: BC Decker Inc.

Dobie, R. A. (2004b). Overview: Suffering From Tinnitus. In J. B. Snow (Ed). *Tinnitus: Theory and Management* (pp. 1–7). Lewiston, NY: BC Decker Inc.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., et al. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods*, *4*, 272–299.

Folmer, R. L., McMillan, G. P., Austin, D. F., et al. (2011). Audiometric thresholds and prevalence of tinnitus among male veterans in the United States: Data from the National Health and Nutrition Examination Survey, 1999–2006. *J Rehabil Res Dev*, *48*, 503–516.

Guyatt, G. (1988). Measuring health status in chronic airflow limitation. *Eur Respir J*, *1*, 560–564.

Guyatt, G. H., Berman, L. B., Townsend, M., et al. (1987). A measure of quality of life for clinical trials in chronic lung disease. *Thorax*, *42*, 773–778.

Halford, J. B. S., & Anderson, S. D. (1991). Tinnitus severity measured by a subjective scale, audiometry and clinical judgement. *J Laryngol Otol*, *105*, 89–93.

Hallam, R. S., Jakes, J. C., Hinchcliffe, R. (1988). Cognitive variables in tinnitus annoyance. *Br J Clin Psychol*, *27*, 213–222.

Haynes, S. N., Richard, D. C., Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychol Assess*, *7*, 238–247.

Henry, J. A., Schechter, M. A., Zaugg, et al. (2006). Clinical trial to compare Tinnitus Masking and Tinnitus Retraining Therapy. *Acta Otolaryngol Suppl*, *126*, 64–69.

Hoffman, H. J., & Reed, G. W. (2004). Epidemiology of Tinnitus. In J. B. Snow (Ed). *Tinnitus: Theory and Management* (pp. 16–41). Lewiston, NY: BC Decker Inc.

Jastreboff, M. M., & Jastreboff, P. J. (1999). Questionnaires for Assessment of the Patients and Their Outcomes. In J. W. P. Hazell (Ed). *Proceedings of the Sixth International Tinnitus Seminar* (pp. 487–491). London: The Tinnitus and Hyperacusis Centre.

Kamalski, D. M., Hoekstra, C. E., van Zanten, B. G., et al. (2010). Measuring disease-specific health-related quality of life to evaluate treatment outcomes in tinnitus patients: A systematic review. *Otolaryngol Head Neck Surg*, *143*, 181–185.

Kirshner, B., & Guyatt, G. (1985). A methodological framework for assessing health indices. *J Chronic Dis*, *38*, 27–36.

Kuk, F. K., Tyler, R. S., Russell, D., et al. (1990). The psychometric properties of a tinnitus handicap questionnaire. *Ear Hear*, *11*, 434–445.

Lipsey, M. W. (1983). A scheme for assessing measurement sensitivity in program evaluation and other applied research. *Psychol Bull*, *94*, 152–165.

Lipsey, M. W. (1990). Design Sensitivity: Statistical Power for Experimental Research. Newbury Park, CA: Sage.

Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annu Rev Psychol*, *51*, 345–375.

Meikle, M. B. (1992). Methods for Evaluation of Tinnitus Relief Procedures. In J.-M. Aran & R. Dauman (Eds). *Tinnitus 91. Proceedings of the Fourth International Tinnitus Seminar* (pp. 555–562). New York, NY: Kugler Publications.

Meikle, M. B., Creedon, T. A., Griest, S. E. (2004). *Tinnitus Archive* (2nd ed.). Retrieved February 4, 2011, from http://www.tinnitusarchive.org/.

Meikle, M. B., & Griest, S. E. (2002). Tinnitus Severity and Disability: Prospective Efforts to Develop a Core Set of Measures. In R. Patuzzi (Ed). *VIIth International Tinnitus Seminar* (pp. 157–161). Perth: Physiology Department, University of Western Australia.

Meikle, M. B., Griest, S. E., Stewart, B. J., et al. (1995). Measuring the Negative Impact of Tinnitus: A Brief Severity Index. In A. Ryan (Ed). *Abstracts of the Eighteenth Midwinter Research Meeting, Association for Research in Otolaryngology* (p. 167). Des Moines, IA: Association for Research in Otolaryngology.

Meikle, M. B., Stewart, B. J., Griest, S. E., et al. (2008). Tinnitus outcomes assessment. *Trends Amplif*, *12*, 223–235.

Meikle, M. B., Stewart, B. J., Griest, S. E., et al. (2007). Assessment of tinnitus: Measurement of treatment outcomes. *Prog Brain Res*, *166*, 511–521.

Moller, A. R. (2007). Tinnitus and pain. *Prog Brain Res*, *166*, 47–53.

Moran, L. A., Guyatt, G. H., Norman, G. R. (2001). Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J Clin Epidemiol*, *54*, 571–579.

Newman, C. W., Jacobson, G. P., Spitzer, J. B. (1996). Development of the Tinnitus Handicap Inventory. *Arch Otolaryngol Head Neck Surg*, *122*, 143–148.

Newman, C. W., & Sandridge, S. A. (2004). Tinnitus Questionnaires. In J. B. Snow (Ed). *Tinnitus: Theory and Management* (pp. 237–254). Lewiston, NY: BC Decker Inc.

Newman, C. W., Sandridge, S. A., Jacobson, G. P. (1998). Psychometric adequacy of the Tinnitus Handicap Inventory (THI) for evaluating treatment outcome. *J Am Acad Audiol*, *9*, 153–160.

Noble, W., & Tyler, R. (2007). Physiology and phenomenology of tinnitus: Implications for treatment. *Int J Audiol*, *46*, 569–574.

Norman, G. R., Sloan, J. A., Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Med Care*, *41*, 582–592.

Nunnally, J. C. (1978). Validity. In J. C. Nunnally (Ed.), *Psychometric Theory* (pp. 86–113). New York, NY: McGraw-Hill.

Puhan, M. A., Bryant, D., Guyatt, G. H., et al. (2005). Internal consistency reliability is a poor predictor of responsiveness. *Health Qual Life Outcomes*, *3*, 33.

Reise, S. P., Morizot, J., Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res*, *16(Suppl 1)*, 19–31.

Reise, S. P., Waller, N. G., Comrey, A. L. (2000). Factor analysis and scale revision. *Psychol Assess*, *12*, 287–297.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychol Methods*, *7*, 147–177.

Stewart, B. J., & Archbold, P. G. (1992). Nursing intervention studies require outcome measures that are sensitive to change: Part One. *Res Nurs Health*, *15*, 477–481.

Stewart, B. J., & Archbold, P. G. (1993). Nursing intervention studies require outcome measures that are sensitive to change: Part Two. *Res Nurs Health*, *16*, 77–81.

Stouffer, J. L., & Tyler, R. S. (1990). Characterization of tinnitus by tinnitus patients. *J Speech Hear Disord*, *55*, 439–453.

Sweetow, R. W., & Levy, M. C. (1990). Tinnitus severity scaling for diagnostic/therapeutic usage. *Hear Instrum*, *41*, 44–46.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th Ed.). Boston, MA: Allyn & Bacon.

Terwee, C. B., Dekker, F. W., Wiersinga, W. M., et al. (2003). On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Qual Life Res*, *12*, 349–362.

Tugwell, P., & Boers, M. (1993). OMERACT conference on outcome measures in rheumatoid arthritis clinical trials: Introduction. *J Rheumatol*, *20*, 528–530.

Turk, D. C. (2002). The Three Most Important Words in Hearing Health Care: Outcomes, Outcomes, Outcomes! In R. Patuzzi (Ed). *Proceedings of the Seventh International Tinnitus Seminar (Suppl*). Perth: Physiology Department, University of Western Australia.

Turk, D. C., & Burwinkle, T. M. (2005). Assessment of chronic pain in rehabilitation: Outcomes measures in clinical trials and clinical practice. *Rehabil Psychol*, *50*, 56–64.

Turk, D. C., Dworkin, R. H., Allen, R. R., et al. (2003). Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain*, *106*, 337–345.

Tyler, R. S. (1993). Tinnitus disability and handicap questionnaires. *Semin Hear*, *14*, 377–383.

Tyler, R. S., & Baker, L. J. (1983). Difficulties experienced by tinnitus sufferers. *J Speech Hear Disord*, *48*, 150–154.

U.S. Department of Health and Human Services. (2006). Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville, MD: Food and Drug Administration.

van der Heijde, D., Bellamy, N., Calin, A., et al. (1997). Preliminary core sets for endpoints in ankylosing spondylitis. Assessments in Ankylosing Spondylitis Working Group. *J Rheumatol*, *24*, 2225–2229.

Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care*, *30*, 473–483.

Wiebe, S., Guyatt, G., Weaver, B., et al. (2003). Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol*, *56*, 52–60.

Wilson, P. H., Henry, J. L., Bowen, M., et al. (1991). Tinnitus reaction questionnaire: Psychometric properties of a measure of distress associated with tinnitus. *J Speech Hear Res*, *34*, 197–201.

Wood, J. M., Tataryn, D. J., Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with Varimax rotation. *Psychol Methods*, *1*, 354–365.

Zaugg, T. L., Schechter, M. A., Fausti, S. A., et al. (2002). Difficulties Caused by Patients' Misconceptions That Hearing Problems are due to Tinnitus. In R. Patuzzi (Ed). *Proceedings of the Seventh International Tinnitus Seminar* (pp. 226–228). Crawley: The University of Western Australia.