

OPEN ACCESS

EDITED BY

Grant Searchfield,
The University of Auckland, New Zealand

REVIEWED BY

Brian Richard Earl,
University of Cincinnati, United States
Tutomu Nakashima,
Nagoya University, Japan

*CORRESPONDENCE

Derek J. Hoare
✉ derek.hoare@nottingham.ac.uk

RECEIVED 20 October 2023

ACCEPTED 18 December 2023

PUBLISHED 19 January 2024

CITATION

Clarke NA, Hoare DJ and Trigg A (2024) A critical review of established tinnitus patient-reported outcomes as measures of Tinnitus Severity and Tinnitus Distress and exemplar analysis of the Tinnitus Handicap Inventory as a formative or reflective measure. *Front. Audiol. Otol.* 1:1325137. doi: 10.3389/fauot.2023.1325137

COPYRIGHT

© 2024 Clarke, Hoare and Trigg. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A critical review of established tinnitus patient-reported outcomes as measures of Tinnitus Severity and Tinnitus Distress and exemplar analysis of the Tinnitus Handicap Inventory as a formative or reflective measure

Nathan A. Clarke¹, Derek J. Hoare^{2,3*} and Andrew Trigg⁴

¹Adelphi Values, Macclesfield, United Kingdom, ²Hearing Sciences, Mental Health and Clinical Neurosciences, University of Nottingham, Nottingham, United Kingdom, ³National Institute for Health and Care Research (NIHR) Nottingham Biomedical Research Centre, Nottingham, United Kingdom, ⁴Bayer plc, Reading, United Kingdom

Tinnitus Severity and Tinnitus Distress are frequently referenced and conflated constructs in research, measured using established tinnitus patient-reported outcome measures (PROs). Confusion regarding these constructs and their relation to fundamental scientific conceptions of tinnitus represents a threat to the validity of PROs as applied in tinnitus research, the conclusions that are reached when applying them, and subsequent progress of theory and clinical interventions for those experiencing tinnitus. Therefore, we critically review relevant literature, providing the Severity of Symptoms (SoS) and Correlates of Complaint (CoC) framework to link tinnitus theory to these constructs. We provide researchers with an overview of latent variable fundamentals (including distinctions between formative and reflective measures, and psychometric and clinimetric measurement traditions). We then provide a synthesis of the relationship between Tinnitus Severity and Tinnitus Distress, the SoS/CoC framework, and latent variable measurement to elucidate their distinctions. Finally, we take the Tinnitus Handicap Inventory (THI) as an exemplar of established tinnitus PROs and use study data ($N = 200$) to empirically evaluate the appropriateness of the THI as a reflective measure of Tinnitus Distress. Subsequently, conceptual and criterion mediation tests provide evidence that the THI is not a reflective measure of Tinnitus Distress according to the CoC conception and should be considered as a formative measure. Researchers should therefore consider whether established tinnitus PROs, such as the THI, are congruent with the scientific conceptions and subsequent theories that they aim to evaluate.

KEYWORDS

tinnitus, patient-reported outcomes, psychometric, clinimetric, psychological measurement

Introduction

Subjective tinnitus (hereafter called tinnitus) is the perception of sound in the absence of a corresponding external sound source or identifiable internal mechanism (Clarke et al., 2020). Most individuals experiencing tinnitus do not report being distressed by the percept (McCormack et al., 2014); therefore, contemporary theorizing conceives of the tinnitus percept and distress related to the percept as distinct components (De Ridder et al., 2021; Mohan et al., 2022).

Many methods have been employed to try and measure aspects of the tinnitus experience. Broadly, these have been split into so-called “objective” and “subjective” methods. Objective methods include biomarkers and imaging methods, which have generally aimed to target the presence or magnitude of the tinnitus percept. Despite much research effort, no “objective” method of measuring the tinnitus percept has been successfully developed, which has been suggested as a contributing factor to there being no cure for tinnitus (McFerran et al., 2019). Subjective methods include psychoacoustic procedures, which aim to match the loudness, pitch, and spectral content of tinnitus to an external sound source (Moore, 2012; Henry, 2016). Self-reported methods (or patient-reported outcomes or PROs) measure aspects of the tinnitus experience directly.

There are many established PROs including the Tinnitus Functional Index (TFI), Tinnitus Questionnaire (TQ), and the Tinnitus Handicap Inventory (THI) but among the most frequently used is the THI (Newman et al., 1998; Hallam, 2008; Meikle et al., 2012; Haider et al., 2016; Hall et al., 2016). However, what established tinnitus PROs such as the THI measure and how the information they provide should be treated analytically is not a trivial issue. Nonetheless, an objective/subjective measurement dichotomy, and primacy of focus on the tinnitus percept as a driving component of distressing tinnitus has led to fundamental psychometric issues in the application of established tinnitus PROs such as the THI.

First among these issues is persistence of two constructs (Tinnitus Severity and Tinnitus Distress), which are frequently referenced and conflated in tinnitus research. Second, established tinnitus PROs are generally assumed to be valid and interchangeable measures of at least one of these constructs (Folmer, 2002; Haider et al., 2016; Boecking et al., 2021); this assumption is contingent on which of the aforementioned constructs is being measured, and therefore, may not be tenable. Third, although contemporary tinnitus theorizing now delineates the tinnitus percept and distress associated with it, established tinnitus PROs were developed within a conceptual framework that assumed all indicators were symptoms of a tinnitus percept (Baguley, 2002; Møller et al., 2010).

Taken in isolation, each of these issues would be problematic; however, an additional complicating factor comes in the form of a general failure within the tinnitus literature to distinguish between formative and reflective latent variable measurement. Taken together, uncritical continuation of the status quo represents a serious threat to the validity of PROs as applied in tinnitus research, the conclusions that are reached when applying them, and subsequent progress of theory and clinical interventions for those

experiencing tinnitus. To untangle and elucidate these fundamental issues, this review and empirical analysis:

- Provides a framework to link historical and contemporary tinnitus theories to tinnitus PROs and latent variable measurement models (the Severity of Symptoms [SoS] framework vs. the Correlates of Complaint [CoC] framework).
- Reviews fundamental concepts in latent variable measurement as applied to tinnitus PROs (i.e., distinction between clinimetric/psychometric measure development and formative/reflective measurement models).
- Explicates how the constructs of Tinnitus Severity and Tinnitus Distress relate to the SoS/CoC framework and latent variable fundamentals.
- Evaluates the THI as an exemplar of established tinnitus PROs to provide evidence that it is a formative measure of Tinnitus Severity, not a reflective measure of Tinnitus Distress.
- Discusses the implications of this evaluation for established tinnitus PROs in relation to validity evidence and continued use in tinnitus research.

Conceptions of tinnitus

Concepts must precede the constructs of Tinnitus Severity and Tinnitus Distress (see “Tinnitus Constructs” for further discussion) and should inform the PROs that operationalise aspects of them. Conceptual frameworks (or conceptions) concern the substantive scientific theories that researchers want to test, while constructs operationalise aspects of these conceptions. When using a PRO, these are operationalised with indicators that are subsequently combined according to a latent variable measurement model (Henseler and Schubert, 2021) to provide a resulting metric of Tinnitus Severity or Tinnitus Distress. The conception is the aspect of substantive scientific interest that each construct aims to represent. Over time, the scientific conceptions that underpin established tinnitus PROs have changed, but their rationale and application in tinnitus research is frequently tacitly assumed.

In the following sections, two essential conceptions of tinnitus are summarized, which link PROs as applied to the tinnitus experience with historic and contemporary scientific theorizing about distressing tinnitus. There are two essential conceptions that may be considered: Severity of Symptoms (SoS) and Correlates of Complaint (CoC). The following sections outline the key differences between SoS and CoC conceptions of tinnitus.

Severity of Symptoms (SoS)

The SoS conception of tinnitus postulates the tinnitus percept to cause symptoms that have a subsequent effect on a person’s function and, ultimately, their quality of life (Figure 1). The SoS conception is rooted in earlier “peripheral” models of tinnitus generation, which suggested that the tinnitus percept originated from a pathology localized to the auditory periphery (Baguley, 2002). The SoS conception of tinnitus arose in medicalised clinical

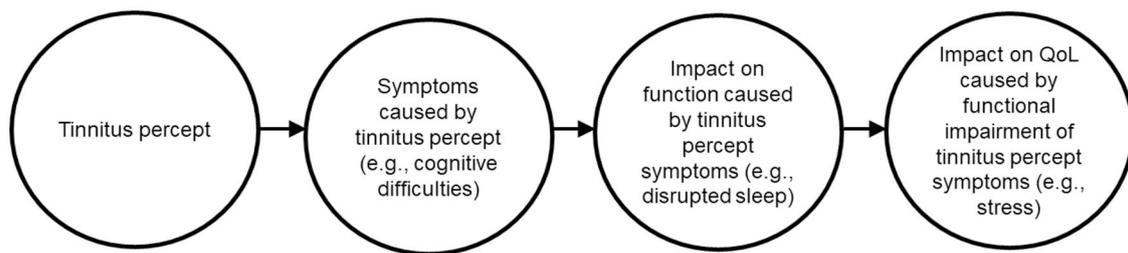


FIGURE 1

Severity of Symptoms (SoS) conception of tinnitus. This historic conception originates from earlier peripheral models of tinnitus and frames the tinnitus percept as having perceptual characteristics that cause observed symptoms on a PRO, which in turn impact function and quality of life.

audiology environments; therefore, development of established tinnitus PROs such as the THI have clinical origins (see “Psychometric vs. clinimetric measurement traditions” for further discussion). Peripheral models of tinnitus have been superseded by “central” models that assume tinnitus to be generated within the brain (Møller et al., 2010); however, the core assumption of early peripheral models of tinnitus generation remain tacit to SoS conceptions of tinnitus, and as such, are implicit to established tinnitus PROs such as the THI. The SoS conception frames the *tinnitus percept* as a causal agent that produces observed symptoms reported by those who respond to established tinnitus PROs.

The SoS conception is bound to the construct of Tinnitus Severity (see Tinnitus Constructs for further discussion). Importantly, SoS conceptions are intrinsic assumptions of various established tinnitus PROs such as the THI because they require the respondent to consider how *the tinnitus percept* affects presumed symptoms. Framing the tinnitus percept as the fundamental cause of symptoms historically led researchers to assume that a property of the percept (presumably its loudness) was driving distress; however, the association between loudness of the tinnitus percept (measured using psychoacoustic methods) and Tinnitus Severity (measured using established tinnitus PROs) was not as strong as initially presumed (Moore, 2012). This has led to an updated scientific conception among tinnitus researchers, and the construct of Tinnitus Severity being gradually replaced with Tinnitus Distress.

Correlates of Complaint (CoC)

The term “correlates of complaint” was used by Hallam et al. (1984) in the context of refuting assumptions intrinsic to the SoS conception (i.e., that tinnitus loudness was the sole driver of reported complaints). The CoC conception relaxes the strict causal assumption made within the SoS conception that the tinnitus percept causes responses to the indicators (i.e., questions on a PRO measure) reported by established tinnitus PROs, such as disrupted sleep, stress, anxiety, etc. The CoC conception postulates that indicators on established Tinnitus PROs are not necessarily symptoms of a tinnitus percept as suggested by the SoS conception, and that the tinnitus percept is associated with, but not necessarily the direct cause of the resultant tinnitus distress (Figure 2). The

CoC conception is therefore linked to the construct of Tinnitus Distress, which is becoming the dominant theoretical position within tinnitus research (De Ridder et al., 2021; Mohan et al., 2022). The CoC conception underpins recent theories that distinguish between causal and statistical links between tinnitus and its complaints (Clarke et al., 2023).

Summary

Essential conceptions of tinnitus were described that link historical and contemporary tinnitus theories to tinnitus PROs and latent variable measurement models. The SoS conception of tinnitus is based on historical peripheral theories of tinnitus generation and assumes that all indicators on PROs are symptoms of the tinnitus percept (or the resulting impact of symptoms on function and quality of life). The SoS conception is bound to the construct of Tinnitus Severity. The CoC conception of tinnitus was developed later than SoS conceptions and developed in parallel with central models of tinnitus generation (Figure 3). The CoC conception does not assume every indicator on established tinnitus PROs to be a symptom of the tinnitus percept. CoC conceptions are related to the construct of Tinnitus Distress and are becoming the dominant conception of what established tinnitus PROs measure.

Tinnitus PROs: latent variable fundamentals

This section introduces and contrasts important concepts in latent variable measurement in relation to established tinnitus PROs, including clinimetric vs. psychometric measurement traditions and reflective vs. formative measurement. Presently available guidance on the appropriateness of reflective and formative models as applied to PROs is also considered.

Psychometric vs. clinimetric measurement traditions

Clinimetrics and psychometrics are examples of different measurement traditions (de Vet et al., 2003). Psychometrics stems

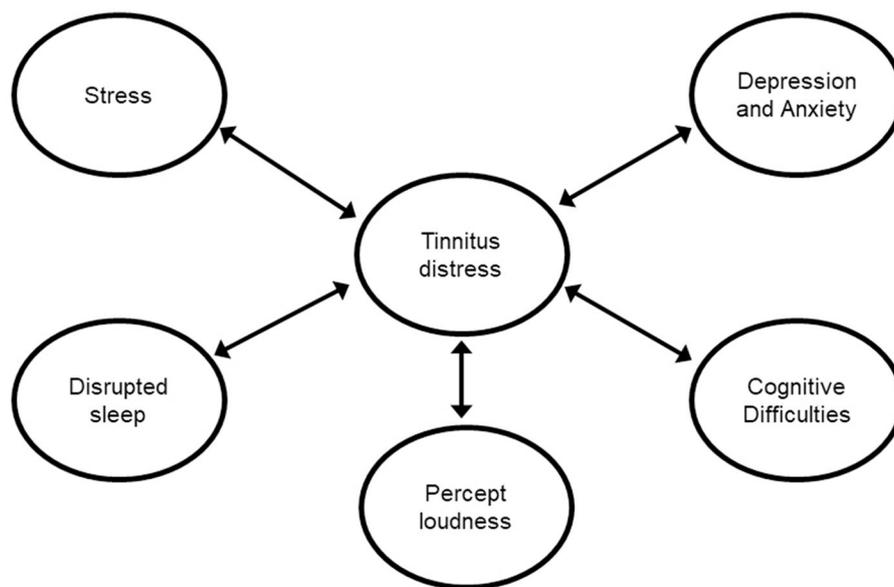


FIGURE 2

Correlates of Complaint (CoC) conception of tinnitus. This contemporary conception accounts for the observation that not everyone experiencing a tinnitus percept is distressed by it, and that tinnitus measured by PROs is not necessarily strongly correlated with psychoacoustic loudness of the tinnitus percept.

from psychological research and classical test theory, but its resultant methods are now frequently seen in healthcare research (Costa, 2015). The measurement philosophy and developmental history that underpins psychometrics and clinimetric disciplines have important implications.

Psychometrics is rooted in ability testing, with its earliest application being in “intelligence” testing. In this application, an individual’s intelligence was conceptualized as being a latent trait that caused their score on a test, with test items then being conceived as reflections of this ability (Spearman, 1904). This essential conception and its techniques were subsequently ported to other domains such as personality psychology, where its fundamentals were also valid. Later, psychometrics was applied in other fields (such as health, economics, and marketing), where the core assumptions of its primary export (i.e., items reflecting a latent trait) may be strained.

Clinimetrics relates to the construction of clinical indexes that emphasize clinical expertise and subject matter when developing measurement scales (Fava et al., 2012). Clinimetric measures describe “symptoms, physical signs, and other distinctly clinical phenomena in medicine” (Cappelleri et al., 2013, p. 13). Clinimetric scales have fundamentally different attributes to psychometric scales, and therefore the development and validation of such instruments should follow different protocols. It has been proposed that a good clinical scale or index comprises of items covering a variety of symptoms, while typical psychometric requirements are not always applicable (de Vet, 2011, p. 43). This relates to the fundamental purpose of a scale, which in many clinimetric applications is to develop a diagnostic, prognostic, or predictive index. An example of a clinimetric index is the Apgar score, used to evaluate the health of new-borns. It combines five symptoms (heart rate, respiratory rate, reflex responses, skin color, and muscle

tone), which may appear to be unrelated but actually provide a well-established and effective predictor of neonatal outcomes (Fayers and Machin, 2016).

This distinction between clinimetric and psychometric aims has not permeated the tinnitus literature. A PubMed search of the terms “tinnitus” and “clinimetric*” presently returns two relevant search results. In its fleeting mentions, clinimetrics as it relates to established tinnitus PROs has been described as “a methodological discipline with a focus on the quality of measurements in medical research and health-care practice” (Hall et al., 2015), which is a broader and more generic definition of clinimetrics than is generally understood in the wider PRO literature (Fayers and Machin, 2016). Nonetheless, a differentiation of clinimetrics from psychometrics and the relevance of this distinction to the development history of tinnitus PROs is crucial for tinnitus researchers. Fundamentally, established tinnitus PROs are congruent with the aims of the clinimetric development tradition. Meikle et al. (2008) note that “During the 1980s and 1990s, a number of questionnaires were designed to evaluate functional, emotional, and other effects of tinnitus”; while Haider et al. (2016) note that most tinnitus PROs have been developed for “doctor-patient decision making about treatment goals and options”.

As a result of the different measurement traditions, psychometrics and clinimetrics have different aims and validation methods, and the distinction between the traditions and their methodologies has been noted in relation to PROs in quality of life research (Cappelleri et al., 2013; Fayers and Machin, 2016). Fayers and Machin suggest that validation of clinimetric scales differs from psychometric scales because the outcomes of the patient (and therefore the predictive value of the scale) is often available. An important feature of clinimetric measures is that they “combine

multiple attributes into a single index”, which has subsequent diagnostic (categorization) or prognostic (predictive) value. This same aim clearly underpins the development of established tinnitus PROs such as the THI (Newman et al., 1998; Fackrell et al., 2018).

Fayers and Hand (2002) summarize the distinction between psychometrics and clinimetrics as follows: psychometrics attempts to measure a single attribute with multiple items, while clinimetrics attempt to measure multiple attributes with a single index. Fayers and Hand also suggest that this distinction is closely related to the distinction between causal and reflective indicators; historically, psychometric methods have assumed reflective-only indicators, while clinimetrics have both formative and reflective indicators.

In summary, many established tinnitus PROs have a development history rooted in the clinimetric tradition, with an associated aim of patient grouping in the context of future treatment efforts. This is particularly true of the THI, where its developers used tinnitus “case histories of patients with tinnitus” to create a measure “helpful in selecting those patients with tinnitus most in need of intervention” (Newman et al., 1996). However, the clinimetric foundations of the THI are not exclusive as subsequent now-established tinnitus PROs, such as the TFI, built on this foundation during development processes (Meikle et al., 2012). Furthermore, despite initial clinimetric development aims, subsequent validation morphed to the near exclusive use of reflective psychometric methodologies (e.g., factor analysis) that are not necessarily appropriate for clinimetric measures.

Reflective vs. formative measurement models

The difference between clinimetric and psychometric measures has been essentially cast as the difference between reflective and formative measurement techniques (Fayers and Hand, 2002). Measurement theories concern how scores generated by PROs represent the unobservable construct that they measure (de Vet, 2011). A measurement model allows a single quantitative value to be obtained “by an algebraic transformation of the responses to item(s)” (Vanier et al., 2021). Latent variable measurement models should follow from theory. Depending on the research interest, observed variables may form different constructs that rely on axioms from different auxiliary theories as justification.

The fundamental difference between reflective and formative measurement concerns whether indicators are “reflections” of the construct they operationalise, or “form” this construct. Formative indicators can be further divided into causal and composite indicators (Bollen and Bauldry, 2011; Henseler and Schuberth, 2021). Conceptual differences between how the various indicators represent a construct impacts expectations about the properties that indicators should exhibit in the context of the concept they are intended to operationalise, and ultimately, affects subsequent claims to validity and success of latent variable measurement.

Reflective measurement models include well-known measurement theories such as Classical Test Theory (CTT)

and Item Response Theory (IRT). Historically, formative models have had no readily equivalent theories, and have relied more on “common sense” (de Vet, 2011). This lack of formalization may explain the popularity and proliferation of reflective latent variable measurement models in healthcare research, including the field of tinnitus research; in fact, Henseler and Schuberth (2021, p. 51) suggest that the “*reflective measurement model is so strongly established... that hardly anyone questions its applicability, despite the fact empirical evidence almost always speaks against it*”. However, in recent years there has been increased recognition of formative measurement models (Henseler and Schuberth, 2021) and their use in areas that have, despite questionable appropriateness, relied solely on reflective psychometric methods (Atkinson and Lennox, 2006; Bollen and Bauldry, 2011; Costa, 2015).

Several authors have explicated key differences between reflective and formative measurement models (Bollen and Bauldry, 2011; Costa, 2015; Fayers and Machin, 2016; Murray and Booth, 2018; Hanafiah, 2020; Henseler and Schuberth, 2021). For formative models: the construct is forged based on the indicators and is considered as an emergent variable; causality flows from the indicators to the construct and a change in the indicators results in a change of the construct; given this, the construct is sensitive to the number and types of indicators representing it. For reflective models, the latent construct exists (in an absolute sense) independently of a measure; causality flows from construct to indicators and a change in the construct causes a change in the indicators; therefore, the construct is not sensitive to changes in the indicators as they are assumed to represent a sample of the infinite population of indicators that could represent the latent variable.

Consequently, the behavior of the indicators in formative and reflective models can be evaluated relative to the construct they are purported to represent [discussed further in “Tinnitus Handicap Inventory (THI): reflective or formative measure?”]. The fundament of this issue is shown in Figure 4 because the nature of the construct being measured, and the indicators that are used to measure it should dictate the measurement model. Figure 4A shows the emergent variable of “Tinnitus Severity”, formed of various items deemed to capture “severe tinnitus” under its initial clinimetric conception (the SoS conception). This is consistent with the development aims of the THI and similar instruments such as the TQ (Newman et al., 1996; Hallam, 2008). In contrast, Figure 4B highlights what happens when reflective measurement models (such as those inherent in factor analysis) are assumed for these instruments; a latent variable of “Tinnitus Severity” is assumed that causes responses to the items. Such an assumption is not only ambiguous regarding what “Tinnitus Severity” actually describes (discussed in Section Tinnitus Constructs), but also creates inherent theoretical inconsistencies with the items comprising “Tinnitus Severity”, and whether it is scientifically plausible that such a latent variable exists that causes tinnitus percept loudness (i.e., Figure 4B). This is a crucial issue for established tinnitus PROs because reflective and formative models have different validation requirements (Henseler and Schuberth, 2021), but evidence for validity of established tinnitus PROs such as the THI has generally been provided using paradigms built around reflective measurement models (e.g., structural validity through factor

analysis). It is therefore unclear as to what validity evidence such studies provide toward the measurement of “Tinnitus Severity”.

Summary

The THI (like many established tinnitus PROs) is most appropriately viewed as a clinimetric measure because its general aim was to measure multiple aspects forming the experience of tinnitus using a single index for use in a clinical context. Subsequently, some indicators are causal/formative in nature, raising doubts about the appropriateness of validity evidence for established tinnitus PROs provided using methods from the psychometric measurement traditions (i.e., reflective methods).

Tinnitus Constructs

A construct is the operationalisation of the concept (“Conceptions of tinnitus”). Ideally when developing a PRO, a concept of interest would be clearly defined; the appropriateness of a reflective or formative measurement model be subsequently considered and identified; and appropriate validation and application of the PRO would then proceed (Fayers and Machin, 2016; Murray and Booth, 2018). Despite the ideal, most researchers are left dealing with an existing PRO and deciding if the indicators should be modeled reflectively or formatively (Costa, 2015; Murray and Booth, 2018).

Fundamentally, “Tinnitus Severity” is an ill-defined construct. However, this issue has been compounded for established tinnitus PROs because there has been a progressive theoretical development in essential conceptions of tinnitus (i.e., a gradual shift from SoS to CoC conceptions). Over time, this has led to a subtle recasting of the construct of interest when using tinnitus PROs from Tinnitus Severity to Tinnitus Distress, which is becoming the dominant view of what established tinnitus PROs measure (De Ridder et al., 2021; Mohan et al., 2022).

Both Tinnitus Severity and Tinnitus Distress are still frequently seen in tinnitus research, but confusion persists regarding their natures, relations to essential conceptions of tinnitus, and implications for appropriate latent variable measurement models. Moreover, a recent trend within the literature has seen these constructs being equated; for example, Gos et al. (2020) define Tinnitus Severity “as the level of distress or impact that tinnitus has on the person”.

The following sections describe evolution of the constructs of Tinnitus Severity and Tinnitus Distress, detailing their relation to essential conceptions of tinnitus, and the implications for appropriate latent variable measurement models.

Tinnitus Severity

Tinnitus Severity as an ambiguous description

In a systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults, Hall et al. (2016) note that the construct of Tinnitus Severity “is not an adequate domain because it does not explain the dimension of

complaint on which severity should be considered.” However, Haider et al. (2016) note that established tinnitus PROs generally “*purport to measure tinnitus severity*”. This highlights the disjoint between this term as an ambiguous description and as a clinimetric index.

Nonetheless, Tinnitus Severity remains a common focus within the literature, with research having frequently investigated its “effects” and “impacts” (Folmer, 2002; Mohamad et al., 2016). Tinnitus Severity has even achieved acceptance as a distinct construct amongst experts and clinical policymakers; this is evidenced in a recent review by the National Institute for Health and Care Excellence (2020) that investigated the usefulness of psychoacoustic measurements in clinical settings, that noted distress, annoyance, and tinnitus severity as additional critical outcomes. But what exactly is tinnitus severity?

Fundamentally, “severity” may describe various aspects of the tinnitus experience:

- Perceptual qualities of tinnitus (typically loudness)
- Symptoms assumed to be caused by the tinnitus percept
- Functional impact caused by tinnitus-related symptoms (assuming a “severe” [loud] tinnitus percept is the cause)
- Impact on quality of life caused by tinnitus symptoms.

This polysemy alone should motivate researchers to define the specific aspect of the tinnitus experience that is viewed as severe. However, it is important to recognize that the term “Tinnitus Severity” necessarily invokes an SoS conception of tinnitus even when being used as an ambiguous descriptor. The only exception is when tinnitus severity is specifically being used to refer to the percept; however, in this instance the term “severity” is still too ambiguous because, even though most people would assume this to be loudness, some people may understand “severity” as referring to pitch or timbre (i.e., a harsh or abrasive sound).

Tinnitus Severity as a clinimetric index

Having described dimensions that “severity” can exist in relation to the tinnitus experience, it is important to note that established tinnitus PROs (such as the THI) in fact cover all aspects of this experience. This is because, as noted by Hallam (2008), complaints relating to tinnitus are “multifaceted”, yet researchers have “nevertheless sought to devise a single indicator of “Tinnitus Severity””. Despite researchers attempting to provide validation evidence for the THI (and similar measures) under a reflective measurement model, “Tinnitus Severity” was conceived under an SoS conception of tinnitus (which aimed to capture all elements of this the impact of the tinnitus percept). Therefore, Tinnitus Severity as measured by established tinnitus PROs should be viewed as a formative construct (or emergent variable). However, scientific conceptions of tinnitus have evolved since the initial development of established tinnitus PROs (Figure 3). Recent scientific discourse of the tinnitus experience stresses the key distinction between the tinnitus percept and related distress (De Ridder et al., 2021; Spankovich, 2021; Mohan et al., 2022).

Despite a desire to retrofit established tinnitus PROs to this new paradigm (i.e., the CoC conception), the difficulties inherent

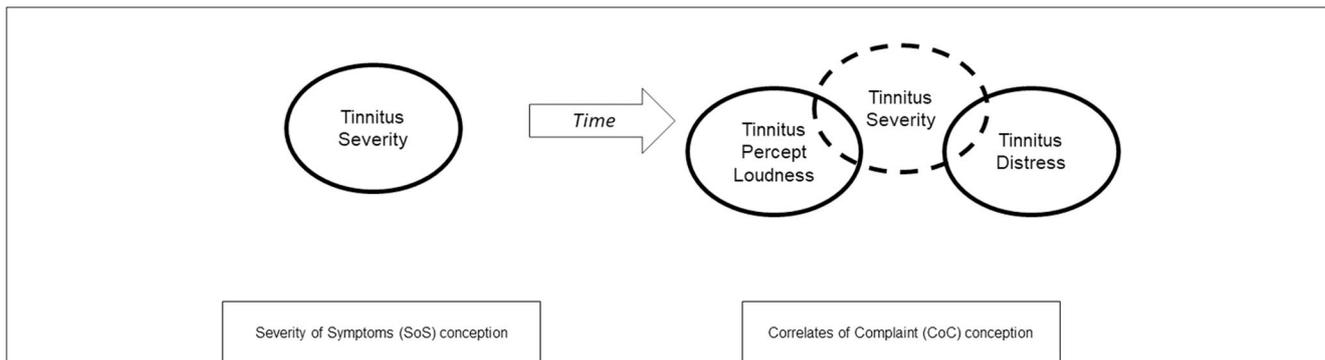


FIGURE 3 Figure depicts the evolution of understanding of the term “tinnitus severity” and its relation to concepts and constructs of tinnitus. Understanding of the term Tinnitus Severity has evolved since its initial usage under a Symptoms of Severity (SoS) conception its contemporaneous splitting into Tinnitus Percept *per se* and related Tinnitus Distress relating to a CoC conception. To accommodate this theoretical distinction, tinnitus researchers have generally adopted a contemporary understanding of Tinnitus Severity as a fractionated construct that encompasses loudness of the tinnitus percept and distress related to this percept.

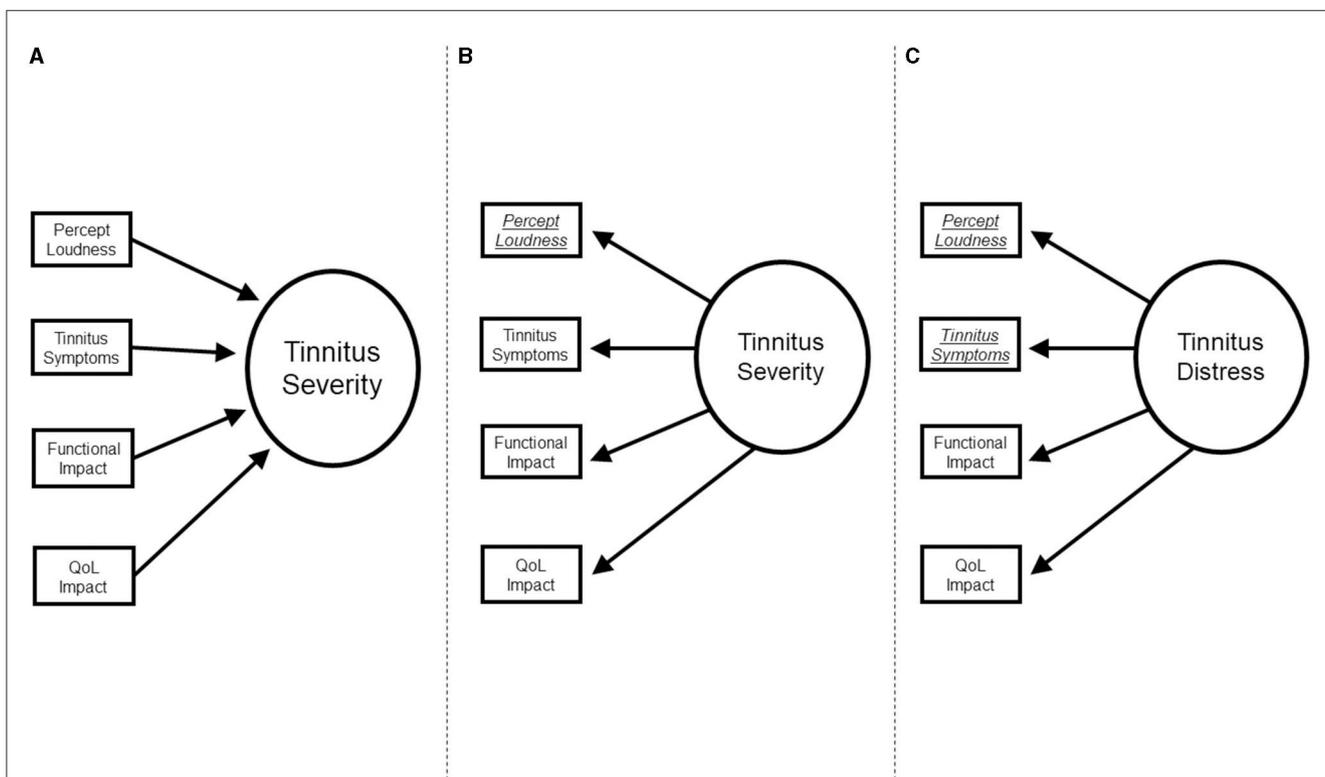


FIGURE 4 Illustrative path diagrams of Tinnitus Severity and Tinnitus Distress relating to formative and reflective measurement models. **(A)** Depicts heterogenous items of established tinnitus PROs contributing to a formative construct of “Tinnitus Severity”, congruent with initial clinimetric aims under a Severity of Symptoms conception. **(B)** Depicts the fundamental issues when applying a reflective measurement model to the same established tinnitus PROs (as has been implicitly assumed through use of techniques such as factor analysis and Cronbach’s alpha); an ambiguous and questionable latent construct must be theorized to cause variation in items. **(C)** Depicts the current practice, where established tinnitus PROs are purported to be measures of Tinnitus Distress under a Correlations of Complaint conception; this entails the same issues introduced in **(B)**, as well as obfuscating what is truly being measured.

in this fractionation and the congruence of established tinnitus PROs with the construct of Tinnitus Distress are evident in current validation efforts of the THI. For example, Gos et al. (2020) defined Tinnitus Severity “as the level of distress or impact that tinnitus has on the person”. However, “distress” may be associated with tinnitus without necessarily being caused by it (i.e., the

CoC conception); while the “impact” of tinnitus evokes a vestigial clinical view of a loud tinnitus percept and “severe” tinnitus (i.e., the SoS conception).

Semantics aside, it would be an error to assume established PROs intended to measure Tinnitus Severity are appropriate measures of Tinnitus Distress. Any attempt at redefinition of the

construct glosses over the fact that established tinnitus PROs were developed under an SoS conception of tinnitus, as measures of Tinnitus Severity, and aimed to provide a single index of “severity”. This essentially clinimetric aim suggests Tinnitus Severity is an emergent variable and that validity evidence provided using a reflective measurement model is questionable.

Tinnitus Distress

Tinnitus Distress is a construct that represents the emotional distress associated with, but not necessarily caused exclusively by, the tinnitus percept. This construct has emerged in contemporary theorizing, separate from loudness of the tinnitus percept, as one of the fractionated components of Tinnitus Severity. Fundamentally this conception removes some of the strong causal assumptions implicit in an SoS conception (i.e., that the tinnitus percept—presumably its loudness—causes observed symptoms of depression, stress, etc). Instead, tinnitus-related distress aligns with a CoC conception because it does not make causal assumptions and describes its exacerbation from “pre-existing” psychological stressors or external disruptive factors (Mohan et al., 2022; Clarke et al., 2023).

Undoubtedly, the switch from Tinnitus Severity to Tinnitus Distress has happened gradually, but its acceptance is crystallising through consensus as the construct of interest apparently measured by established tinnitus PROs (De Ridder et al., 2021; Spankovich, 2021; Mohan et al., 2022). However, while conceptual theorizing has evolved (i.e., SoS to CoC conceptions), it is important to note that established tinnitus PROs were designed under the SoS conception of tinnitus and “Tinnitus Severity”, which are aligned with formative measures (specifically clinimetric indexes) because of their clinical development based on earlier peripheral models.

Nonetheless, the switch of PROs to measures of Tinnitus Distress has seen psychometric methods being used to provide validation evidence and refinement of established tinnitus PROs such as the THI (Gos et al., 2020; Wakabayashi et al., 2020). However, the application of psychometric measurement models in the validation and application of tinnitus PROs is problematic. Fundamentally, this is because of assumptions that PRO items are reflective indicators of Tinnitus Distress.

The primary problem inherent in the underlying switch from Tinnitus Severity (SoS conception) to Tinnitus Distress (CoC conception) is illustrated in Figure 4. Figure 4A shows Tinnitus Severity as a formative model (with the types of indicators that are generally present across established tinnitus PROs). As a formative index, these can be viewed as the constituent parts of “severe tinnitus” and its impacts, which are congruent with initial clinimetric aims.

Although various factor models have been suggested for established tinnitus PROs such as the THI and TFI (Kuk et al., 1990; Baguley and Andersson, 2003; Fackrell et al., 2016; Chandra et al., 2018; Beukes et al., 2020; Wakabayashi et al., 2020), Figure 4B illustrates the fundamental psychometric issues introduced by using a reflective measurement model

to operationalise these (i.e., an ambiguous latent variable is assumed—see Section Tinnitus Constructs); simply swapping the construct of interest from Tinnitus Severity to Tinnitus Distress does nothing to alleviate the fundamental issue. Although the construct has nominally changed, various conflicts remain because the established tinnitus PRO indicators remain the same; only the theoretical assumptions of the latent variable measurement model have changed. Moreover, new conflicts are introduced because the items of the THI (and other established tinnitus PROs such as the TFI) explicitly specify the *tinnitus percept* to be the cause of these issues, not a reflectively operationalised latent variable (whether it be Tinnitus Severity or Tinnitus Distress).

This fundamental conflict is between the item content of established tinnitus severity PROs and the theoretical proposition of Tinnitus Distress being related to—but not necessarily caused by—the tinnitus percept (i.e., the switch from an SoS to CoC conception). The item content of established tinnitus PROs such as the THI states typically begins with the item stem: “Because of your tinnitus...”. Most patients would understand this item content to be enquiring about the direct causal effect of the tinnitus percept on symptoms, functions, and quality of life (i.e., the SoS conception). This raises a fundamental question regarding whether the content of such measures allows them to be considered as valid reflective indicators of Tinnitus Distress.

A second conflict concerns the specific assumptions entailed by a reflective measurement model as applied to established PRO indicators originally intended to measure Tinnitus Severity. A reflective model essentially switches the assumed direction of causality. Fundamentally, this switch suggests that the latent variable of Tinnitus Distress is causing responses to indicators; however, a patient would suggest that these are being caused by the tinnitus percept *per se* (and are being asked to respond based on this presupposition). This is problematic because it has been shown that the inappropriate application of a reflective (Common Factor) model may bias resulting parameter estimates (Rhemtulla et al., 2020); however, much research effort has been spent establishing the validity of established tinnitus PROs based on methods that are unique to reflective validation, such as structural validity through factor analysis, or internal consistency using Cronbach’s alpha (Chandra et al., 2018; Gos et al., 2020).

Additionally, interactions between the assumptions of a reflective measurement model and the indicators of established tinnitus severity PROs create difficult measurement propositions from a theoretical perspective. For example, indicators that inquire about the loudness of the tinnitus percept are still included in a reflective model. Conceptually, this implies that Tinnitus Distress causes tinnitus loudness. Although some researchers have proposed theoretical models that attention may result in a “distorted perception” (McKenna et al., 2014), it is likely that many patients would suggest that the loudness of their tinnitus percept causes their distress.

Finally, it should be noted that Tinnitus Distress may be measured formatively; however, while this is true, the indicators define the construct to be entirely synonymous with Tinnitus Severity. While this may provide semantic consensus, it entirely avoids the theoretical and substantive considerations that have been presented.

Summary

This section has discussed “tinnitus severity” and how the term can be an ambiguous description of elements of the tinnitus experience; however, established Tinnitus PROs such as the THI were intended as a clinimetric index of Tinnitus Severity under the SoS conception, and should be operationalised as a formative construct. Nonetheless, tinnitus researchers are generally moving toward a consensus that established tinnitus PROs are measures of Tinnitus Distress (based on a CoC conception). The construct switch from Tinnitus Severity to Tinnitus Distress raises several conflicts and challenges concerning both reflective measurement and attempting to retrofit established tinnitus PROs to operationalise Tinnitus Distress. Given these issues, it is necessary to evaluate whether established tinnitus PROs such as the THI are adequate reflective measures of Tinnitus Distress.

Tinnitus Handicap Inventory (THI): reflective or formative measure?

Given the ambiguity surrounding established tinnitus PROs as measures of Tinnitus Severity or Tinnitus Distress, and the mapping of these constructs to specific conceptual and latent variable models, it is crucial to adequately evaluate the appropriateness of the THI as a reflective measure of Tinnitus Distress. There are two main ways to evaluate the appropriateness of PROs as reflective or formative measures: conceptual tests and empirical analysis.

Conceptual tests essentially take the form of “thought experiments” (Bollen and Bauldry, 2011; Fayers and Machin, 2016). The fundamental aim of such conceptual tests is to evaluate aspects of expected behavior of indicators of a PRO when operationalizing a specific construct, as well as their theorized relationship to the construct. Empirical tests provide supportive evidence of the appropriateness of modeling a set of indicators in a reflective or formative manner.

As a formative measure of Tinnitus Severity, the THI is simply defined as a composite index of the chosen indicators (how the appropriate weights for such a composite index should be selected is not addressed here); therefore, empirical tests are more fruitfully employed in assessing whether the THI is an appropriate reflective measure of Tinnitus Distress. Moreover, framing the question in this manner provides direct assessment of an essential question: can the THI be retrofit to the emerging consensus view of established tinnitus PROs being reflective measures of Tinnitus Distress? If the THI is not an appropriate reflective measure, this raises questions regarding validation evidence based on the reflective assumption, and whether this evidence should continue to be considered permissive of established tinnitus PROs validity in a CoC conception.

To answer this question, this analysis assessed the appropriateness of the THI as a reflective measure of Tinnitus Distress. This was done using an assessment strategy proposed by Murray and Booth (2018). The conceptual test of the proposition was provided by using five evaluative criteria, combined with an empirical test that employed criterion mediation modeling.

TABLE 1 Summary overview of whether THI indicators (items) are reflective in nature.

Criteria	Summary explanation	Are the indicators reflective?
Does the direction of causality flow from construct to indicators or vice versa according to the theoretical definition of the construct?	The THI indicators cause the construct	No
Would a change in the indicator produce a change in the construct?	Changing the THI indicators would change construct	No
Is the construct the common cause of all the indicators?	Tinnitus Distress does not cause the loudness of the tinnitus percept	No
Do the indicators have the same antecedents and consequences?	Tinnitus Distress is not the antecedent and the consequences are known to be heterogeneous	No
Are the indicators interchangeable?	THI item content varies widely	No

Conceptual test of the THI as a reflective measure of Tinnitus Distress

Several conceptual criteria have been suggested that provide “mental experiments” to establish whether a collection of indicators should be modeled in a formative or reflective manner (Jarvis et al., 2003; Fayers and Machin, 2016; Murray and Booth, 2018). Murray and Booth (2018) present five criteria that distill essential aspects of these mental experiments:

1. Does the direction of causality flow from construct to indicators or vice versa according to the theoretical definition of the construct?
2. Would a change in the indicator produce a change in the construct?
3. Is the construct the common cause of all the indicators?
4. Do the indicators have the same antecedents and consequences?
5. Are the indicators interchangeable?

If the answer to these questions is “yes”, then this suggests the indicators follow a reflective model. The following sections review the indicators of the THI in relation to the construct of Tinnitus Distress. Table 1 provides a summary overview of whether the various elements of the conceptual test are supportive of the THI being a reflective measure.

Does the direction of causality flow from construct to indicators or vice versa according to the theoretical definition of the construct?

The first of the criteria underscores the conceptual and construct-related issues reviewed detailed previously. The indicators of the THI clearly suggest that the *tinnitus percept* causes variation in its indicators (i.e., an SoS conception), which in turn causes the construct of interest (whether framed as “severity” or

“distress”). The flow of causation therefore runs from indicators to construct, suggesting a formative and not a reflective model.

Would a change in the indicator produce a change in the construct?

In the case of the THI, removal of specific indicators would produce a change of the construct being measured because of the varied item content; moreover, this content is known to not be experienced by every person that is distressed by their tinnitus, which has led to the general conclusion that tinnitus is a heterogeneous condition (Mohan et al., 2022). Examples of item variation include items concerning attention (Item 18), trouble sleeping (Item 7), household responsibilities (Item 13), interfamilial stress (Item 17), and ability to cope (Item 23). It has already been noted in the literature that various purported established measures of Tinnitus Distress have different indicator content, likely impacting the construct that is being measured; Hall et al. (2019) note that “the TQ has proportionately more items asking about emotional distress than does the THQ (37% vs. 22%), while the THQ has proportionately more items asking about hearing than does the TQ (19% vs. 13%)”. This suggests generally that a unidimensional construct of Tinnitus Distress is not being reflected, and that a construct is being formed by the selection of indicators.

Is the construct the common cause of all the indicators?

Assuming a latent variable of Tinnitus Distress that causes tinnitus percept loudness (e.g., Item 2) is theoretically inconsistent, with most individuals intuition that a loud tinnitus percept to be the cause of Tinnitus Distress. Again, most indicators constituting the THI clearly describe the *tinnitus percept per se* (not Tinnitus Distress) as being the cause of responses to indicators. This is based on the SoS conception of tinnitus. Most indicators provide explicitly causal statements that implicate the percept, with item stems beginning: “Because of your tinnitus...?” or “Does your tinnitus make you...?”. The conceptual conflict is readily apparent when considering Tinnitus Distress as the common cause of indicators that most readers would infer to be asking about a percept.

Do the indicators have the same antecedents and consequences?

The antecedents of the THI are the *tinnitus percept* and not a latent variable of Tinnitus Distress. Moreover, consequences of the THI indicators are highly distinct. For example, an individual’s ability to enjoy their social life (Item 9) is a very different consequence than an inability to perform household activities (Item 13). Similarly, stressed relationships with family and friends (Item 17) are different consequences than trouble falling to sleep at night (Item 7) or feelings of desperation (Item 5).

Are the indicators interchangeable?

Finally, the indicators of the THI are clearly not interchangeable. For example, an indicator enquiring about difficulties when reading (Item 15) is not interchangeable as an indicator enquiring about feelings of insecurity (Item 25). Similarly, an inability to perform household activities (Item 13) is not interchangeable with an indicator of an individual’s ability to enjoy their social life (Item 9).

In summary, the criteria provided by Booth and Murray (2018) clearly highlight the difficulties of treating the indicators of the THI as reflective of the construct of Tinnitus Distress; this is at odds with the emerging consensus-driven notion that established tinnitus PROs may be treated as reflective measures of Tinnitus Distress.

Criterion mediation test of the THI as a reflective measure of Tinnitus Distress

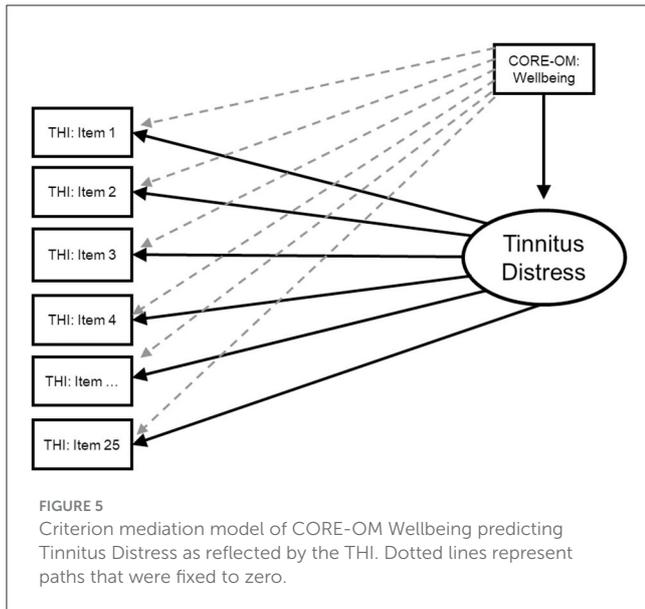
Rationale

Murray and Booth (2018) have proposed a criterion mediation test to assess whether items on a PRO should be treated as formative or reflective indicators. They suggest that if the effect of an external criterion variable on the indicators is completely mediated by the latent variable, then it is defensible to treat them as reflective indicators of the construct. Conversely, if the effect of the external criterion variable on the indicators is not completely mediated by the latent variable, then this can be viewed as evidence of the indicators being inappropriate within a reflective measurement model.

A criterion mediation test can therefore be used to evaluate whether the THI is an appropriate reflective measure of Tinnitus Distress by fitting a confirmatory factor analysis (CFA) model to the 25 indicators of the THI. Tinnitus Distress can then be predicted by a relevant criterion variable. In this analysis, the Wellbeing scale of the CORE-OM (Evans et al., 2002) was selected as a relevant criterion variable; this was chosen because the CORE-OM has been suggested as a measure of emotional distress for people with tinnitus (Handscorn et al., 2016) and the psychological theory that underpins this rationale is consistent with the CoC conception, and suggests that an individual’s emotional distress should predict their subsequent Tinnitus Distress (i.e., the latent variable as operationalised by the indicators of the THI). A structural equation model can be specified that estimates the effect of Wellbeing on Tinnitus Distress and fixes direct path coefficients from the criterion to each indicator to zero (Figure 5). Modification Indices (MIs) can then be used to ascertain if the fit of the model would be improved by allowing Wellbeing to directly impact the indicators. If this is the case, the latent variable of tinnitus distress does not empirically mediate the effect of Wellbeing, which would suggest that the indicators are not adequate reflective indicators of the latent variable Tinnitus Distress and may not represent this construct (Flake and Fried, 2020).

Materials and methods

A dataset of 200 participants who completed the THI was used for this analysis; participants were part of a study assessing



attention and working memory performance for individuals with tinnitus (Mohamad, 2015). Participants were recruited from both communal and clinical settings to obtain a broadly representative sample of the population with tinnitus. The prospective, cross-sectional study design included English speaking participants, aged 18–80-years old who experienced tinnitus (age $M = 59.49$, $SD = 12.71$). All participants provided informed consent. Ethical permission for the study was provided by Derby Research Ethics Committee (Reference: 13/EM/0192) on 18 June 2013, and the study was sponsored by the Nottingham University Hospitals NHS Trust Research and Development (Reference: 13IH001).

Statistical analyses were undertaken using the R programming language Version 4.1.1 and Lavaan package (Rosseel, 2012; R Core Team, 2022).

A structural equation model was fit of Tinnitus Distress (operationalised by the THI) being regressed on the criterion variable of CORE-OM Wellbeing. The direct paths from the criterion variable (CORE-OM Wellbeing) to each indicator of the THI were fixed to zero. A mean- and variance-adjusted weighted least squares (WLSMV) estimator was used given the THI items are categorical variables. The MIs were subsequently evaluated to assess whether the effect of the criterion was entirely mediated through the latent variable. The effect of the criterion variable was considered empirically mediated by the latent variable if all MI values were below 3.841 [i.e., the critical value for chi-square with an alpha level of 0.05 on one degree of freedom (Murray and Booth, 2018)].

Results

Table 2 shows global model fit statistics from the structural equation model. Table 3 displays several MIs that exceed the critical threshold (i.e., >3.841), where the fit of the criterion mediation model would be significantly improved by allowing THI indicators to be regressed directly onto the CORE-OM Wellbeing. The THI indicators include Item 2, $MI = 6.388$; Item 8, $MI = 5.063$; and

TABLE 2 Global fit statistics from criterion mediation model of CORE-OM Wellbeing predicting Tinnitus Distress as reflected by the THI.

Global fit statistic	Statistic value
Chi-square (p -value)	6.388 ($p < 0.001$)
Comparative fit index (CFI)	0.984
Root mean square error of approximation (RMSEA)	0.059

TABLE 3 Modification indices (MI) that are not moderated according to the Murray and Booth (2018) critical threshold (i.e., $MI > 3.841$) from criterion mediation model of CORE-OM Wellbeing predicting Tinnitus Distress as reflected by the THI.

Criterion mediation model suggested MI	MI value
THI Item 2 regressed on CORE-OM WellBeing	6.388
THI Item 8 regressed on CORE-OM WellBeing	5.063
THI Item 25 regressed on CORE-OM WellBeing	4.106

THI, Tinnitus Handicap Inventory; CORE-OM, Clinical Outcomes in Routine Evaluation-Outcome Measure; MI, Modification Indices.

Item 25, $MI = 4.106$. This provides evidence that the CORE-OM Wellbeing is not empirically mediated by a measurement model that considers the THI as a reflective measure of Tinnitus Distress. This suggests the items may be formative in nature.

Summary

A conceptual test of the THI evaluated it against five criteria to judge whether its indicators should be considered reflective of Tinnitus Distress. For each of these criteria, indicators of the THI clearly do not possess reflective qualities (i.e., they are not interchangeable, and would produce a change in the construct if swapped); furthermore, a common cause of Tinnitus Distress is not obviously shared by all the indicators, and if Tinnitus Distress is being measured, the flow of causality is from indicators to Tinnitus Distress. However, if a common cause related to the tinnitus experience is present, then the item content of the THI makes clear that it is the *tinnitus percept per se* (not Tinnitus Distress), which therefore binds the THI to the SoS conception of tinnitus. A criterion mediation test also provided empirical evidence that the latent variable of Tinnitus Distress as operationalised by the THI indicators did not reflectively mediate the effect of Wellbeing (operationalised by the CORE-OM). This provides further supportive evidence that the THI is not a reflective measure of Tinnitus Distress according to the CoC conception.

Discussion

In this critical review, we detailed essential conceptions linking current tinnitus theorizing to established tinnitus PROs. We then used the THI as an exemplar of an established tinnitus PRO and demonstrated that it cannot be retrofit as a reflective measure of Tinnitus Distress. Crucially, we have also highlighted that established tinnitus PROs (such as the THI) are bound to

specific theoretical assumptions (SoS conceptions) that are not congruent with contemporary theorizing (CoC conceptions). This has fundamental consequences for the use of established tinnitus PROs in contemporary research that cannot be ignored.

Tinnitus research is currently plagued by the “jingle jangle fallacy” (Flake and Fried, 2020). Hall et al. (2019) have noted that “the most popular measurement instruments are those that assess tinnitus as a composite multidomain construct, and these [prominent tinnitus PROs] are used somewhat interchangeably...”, while cautioning that “the tinnitus domains captured by each [PRO] can dramatically differ across instruments, and few conventions apply”. In the psychometric literature, this is known as the jingle jangle fallacy (Flake and Fried, 2020). Jingle occurs when two measures are assumed to measure the same construct because they have similar names. Jangle occurs when two constructs are assumed to measure different constructs because they have different names. A major source of such confusion is due to the historic success and proliferation of the term “tinnitus severity”. Continued use of the term “tinnitus severity” within the literature and subsequent empirical studies should be subject to increased scrutiny and justification as to what aspects of the tinnitus experience are thought to be “severe” (see Section “Tinnitus Severity” in this manuscript). Uncritical use of “tinnitus severity” persists as a source of continued opportunity for researchers, clinicians, patients, and policy makers to miscommunicate. It should be replaced with a term like “Impact of Tinnitus”. As outlined (see “Tinnitus Constructs” section), numerous distinct theoretical concepts can be sheltered under the umbrella term of “tinnitus severity”. At a conceptual level alone, this should raise serious concerns regarding the soundness of persisting with a term that obscures so many theoretically important details.

Beyond the semantics, various technical issues also exist in contemporary tinnitus research with regards to appropriate measurement models for established tinnitus PROs. All established tinnitus PROs feature a sum score that was intended as a clinimetric index (Hallam, 2008), and the score it produces is an emergent variable that should be modeled formatively. This review and analysis provide evidence that continued use of sum scores of established tinnitus PROs as measures of ‘Tinnitus Distress’ in applied tinnitus research is inappropriate. Established tinnitus PROs are more appropriately considered as clinimetric indexes of “Tinnitus Severity”. However, this raises issues beyond the construct. This is because the use of sum scores implicitly assumes equal contribution (or unit-weighting) of all tinnitus symptoms to the resulting Tinnitus Severity index (McNeish and Wolf, 2020). This is known to not be the case in tinnitus, which is consistently described as a heterogeneous condition (Genitsaridi et al., 2020; Mohan et al., 2022). Various methods exist to provide weights to formative indexes, and the continued application of the THI (and established tinnitus PROs in general) as a Tinnitus Severity index requires further investigation to justify the questionable unit-weighting assumption (de Vet, 2011; Murray and Booth, 2018).

A further important implication of this work is that current validity evidence of established tinnitus PROs needs to be reappraised, and the generally accepted position that established tinnitus PROs are universally “validated measures” needs to be seriously questioned. As we have demonstrated, established

tinnitus PROs are essentially formative measures, however, many aspects of reflective methodology have been used to provide “structural validity” and evidence of “internal consistency” (e.g., factor analysis and Cronbach’s alpha); these are inappropriate and simply not relevant aspects of validity to formative measures. The application of psychometric methodology (i.e., the common factor model) to validate quality of life PROs has been described as persisting because the results it provides “make sense” to the researcher (Fayers and Machin, 2016). However, modeling formative constructs with an unjustified common factor model has been criticized and shown to lead to biased relationships among psychological constructs (Rhemtulla et al., 2020). Furthermore, validity is a complex topic and not simply a function of parameters generated by statistical techniques (Markus and Borsboom, 2013; Murray and Booth, 2018). A key finding of our review and analysis is that established tinnitus PROs are likely to not be optimal measures for testing scientific theories based on CoC conceptions of tinnitus.

Our conclusions are at odds with the current zeitgeist in tinnitus research, which views tinnitus PROs as measures of Tinnitus Distress (Mohan et al., 2022). In terms of substantive scientific theory, established tinnitus PROs are bound to an SoS conception and retrofitting them to act as measures of Tinnitus Distress causes theoretical issues in modeling and interpretation. This extends beyond the THI, as even the individual scales of more recent instruments such as TFI are bound to the SoS conception because of their item content (which states “because of your tinnitus”—e.g., the sleep scale of TFI). Although use of such scales is not necessarily a problem if it captures the scientific conception and theory that a researcher intends to test (i.e., the SoS conception), it is crucial for researchers to be aware of the scientific assumptions intrinsic to the measure (i.e., the tinnitus percept causing the sleep problem). In contrast, if a test of CoC conception was intended, then an appropriate measure should be considered to operationalise that scientific theory. For example, if someone wanted to test whether poor sleep increases perception of tinnitus loudness the next day, they may want to use a measure of sleep disruption *per se*, not the TFI subscale.

This work has highlighted the importance of individual items that constitute a PRO when attempting to understand relationships between the construct it measures (i.e., Tinnitus Severity or Tinnitus Distress) and other variables. In the criterion mediation analysis accompanying this review, THI items 2 (Does the loudness of your tinnitus make it difficult for you to hear people?), 8 (Do you feel as though you cannot escape your tinnitus?), and 25 (Does your tinnitus make you feel insecure?) were shown to improve model fit if regressed directly on CORE-OM Wellbeing; however, the appropriateness and relevance of specific items when measuring Tinnitus Severity or Tinnitus Distress may be different depending on the context and aims of the study.

Crucially, despite the ubiquity of tinnitus PROs in research studies, there is no consensus regarding what constitutes direct symptoms of the tinnitus percept (i.e., an SoS conception). For example, Tyler et al. (2014) have distinguished between primary and secondary effects of tinnitus and have claimed that only four domains are impaired by tinnitus (emotions, hearing, sleep,

and concentration), and are therefore relevant when measuring therapeutic change attributed to a tinnitus-specific intervention. Ultimately, such suggestions are conjecture and highlight that it is crucial to explicate the theory underlying a researcher's conceptions, because one researcher may view something as a "symptom of severity" while another may view it as a "correlate of complaint". The Severity of Symptoms [SoS] framework vs. the Correlates of Complaint [CoC] framework provides a means for researchers to align their measurement choices with their substantive scientific intent.

This work also has implications for future research and is particularly relevant to endpoint selection in clinical trials assessing tinnitus interventions. Hall et al. (2019) have recommended core outcome sets for use in clinical trials investigating tinnitus that are categorized according to the intervention being assessed. However, selection of appropriate clinical outcome assessment for such trials also needs to consider the underlying scientific conception for appropriate measurement that is tied to established tinnitus PROs. This work has shown that established tinnitus PROs may not be appropriate for assessing theories based on CoC conceptions of tinnitus; this is an important finding, as clinical trials continue to consider the construct of Tinnitus Distress using the THI (Simoes et al., 2023). This issue will become increasingly relevant in clinical trials evaluating interventions for tinnitus. The relevance of the patient experience is already being increasingly recognized in clinical research (Fayers and Machin, 2016). For example, in oncology, where objective measures already exist (e.g., tumor growth in cancer clinical trials), there is increased recognition that PROs provide important data that is attainable only from the patient (e.g., pain and health-related quality of life). Contrasted with oncology, the issue of accurate PRO measurement related to tinnitus is at least as pressing given the absence of reliable "objective" measures.

Moreover, contemporary opinion among tinnitus researchers suggests that PROs are linked to subjective experience and emotional distress, while imaging and biomarkers are linked more directly to the tinnitus percept and may be considered more "objective" (McFerran et al., 2019). However, direct measurement of the patient experience of tinnitus will always be required, and currently risks being overlooked as "merely subjective" evidence. In tinnitus research, imposing an objective/subjective dichotomy runs the risk of inaccurately communicating to researchers and clinicians that PROs are fundamentally less reliable (de Vet, 2011), potentially casting doubt on patient-reported information and undermining this important direct source of data. If established tinnitus PROs are to continue to feature in clinical trials, they should be held to foremost evidential standards. Ultimately, this includes showing the fundamental validity and application of established tinnitus PROs in the context of the scientific theories that undergird them.

Conclusion

Researchers report established tinnitus PROs to be measures of both "Tinnitus Severity" and "Tinnitus Distress". Despite emerging consensus that established tinnitus PROs measure "Tinnitus Distress", this review detailed essential scientific

conceptions of tinnitus in relation to established tinnitus PROs, arguing that they are clinimetric measures of a formative construct "Tinnitus Severity". Therefore, the validity evidence of established tinnitus PROs that has relied on techniques assuming a reflective measurement model is questionable (i.e., structural validity through factor analysis and reliability using Cronbach's alpha). Subsequently, we evaluated the THI (an exemplar of established tinnitus PROs) as a reflective measure of Tinnitus Distress using both conceptual and empirical tests, with evidence from both showing that THI items are likely formative in nature. Researchers should therefore consider whether established tinnitus PROs, such as the THI, are congruent with the scientific conceptions and subsequent theories that they aim to evaluate.

Author contributions

NC: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. DH: Conceptualization, Data curation, Funding acquisition, Writing – review & editing. AT: Methodology, Validation, Conceptualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This article reports independent research supported by the National Institute for Health and Care Research (NIHR) Biomedical Research Center Funding Programme (BRC-1215-20003).

Conflict of interest

NC was an employee of Adelphi Values at the time of writing. AT is an employee and shareholder in Bayer plc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care (DHO).

References

- Atkinson, M. J., and Lennox, R. D. (2006). Extending basic principles of measurement models to the design and validation of Patient Reported Outcomes. *Health Qual. Life Outcom.* 4, 65. doi: 10.1186/1477-7525-4-65
- Baguley, D. M. (2002). Mechanisms of tinnitus. *Br. Med. Bull.* 63, 195–212. doi: 10.1093/bmb/63.1.195
- Baguley, D. M., and Andersson, G. (2003). Factor analysis of the Tinnitus Handicap Inventory. *Am. J. Audiol.* 12, 31–34. doi: 10.1044/1059-0889(2003/007)
- Beukes, E. W., Frumkin, M., Kallogjeri, D., Piccirillo, J., Manchaiah, V., Andersson, G., et al. (2020). *Measurement of Tinnitus Distress: Confirmatory Factor Analysis of the Tinnitus Handicap Inventory and Tinnitus Functional Index and Development of a Combined Short Form*. Baltimore, MD: Association for Behavioral and Cognitive Therapy 54th Annual Convention (ACBT 2020). Available online at: <http://arro.anglia.ac.uk/id/eprint/706001/>
- Boecking, B., Brueggemann, P., Kleinjung, T., and Mazurek, B. (2021). All for one and one for all? - Examining convergent validity and responsiveness of the German Versions of the Tinnitus Questionnaire (TQ), Tinnitus Handicap Inventory (THI), and Tinnitus Functional Index (TFI). *Front. Psychol.* 12, 596037. doi: 10.3389/fpsyg.2021.596037
- Bollen, K. A., and Bauldry, S. (2011). Three Cs in measurement models: causal indicators, composite indicators, and covariates. *Psychol. Methods* 16, 265–284. doi: 10.1037/a0024448
- Cappelleri, J. C., Zou, K. H., Bushmakina, A. G., Alvir, J. M. J., Alemayehu, D., and Symonds, T. (2013). *Patient-Reported Outcomes: Measurement, Implementation and Interpretation: 64 (Chapman and Hall/CRC Biostatistics Series)* (1st ed.). Boca Raton, FL: Chapman and Hall/CRC. Available online at: <https://www.amazon.co.uk/Patient-Reported-Outcomes-Implementation-Interpretation-Biostatistics/dp/1439873674>
- Chandra, N., Chang, K., Lee, A., Shekhawat, G. S., and Searchfield, G. D. (2018). Psychometric validity, reliability, and responsiveness of the tinnitus functional index. *J. Am. Acad. Audiol.* 29, 609–625. doi: 10.3766/jaaa.16171
- Clarke, N. A., Akeroyd, M. A., Henshaw, H., Hall, D. A., Mohamad, W. N. W., and Hoare, D. J. (2023). The executive disruption model of tinnitus distress: model validation in two independent datasets using factor score regression. *Front. Psychol.* 14, 1006349. doi: 10.3389/fpsyg.2023.1006349
- Clarke, N. A., Henshaw, H., Akeroyd, M. A., Adams, B., and Hoare, D. J. (2020). Associations between subjective tinnitus and cognitive performance: systematic review and meta-analyses. *Trends Hear.* 24, 2331216520918416. doi: 10.1177/2331216520918416
- Costa, D. S. J. (2015). Reflective, causal, and composite indicators of quality of life: a conceptual or an empirical distinction? *Qual. Life Res.* 24, 2057–2065. doi: 10.1007/s11136-015-0954-2
- De Ridder, D., Schlee, W., Vanneste, S., Londero, A., Weisz, N., Kleinjung, T., et al. (2021). "Tinnitus and tinnitus disorder: Theoretical and operational definitions (an international multidisciplinary proposal)," in *Progress in Brain Research*. Amsterdam: Elsevier.
- de Vet, H. C. W., Terwee, C. B., and Bouter, L. M. (2003). Clinimetrics and psychometrics: two sides of the same coin. *J. Clin. Epidemiol.* 56, 1146–1147. doi: 10.1016/j.jclinepi.2003.08.010
- de Vet, H. C. W. W. (2011). *Measurement in Medicine: A Practical Guide (Practical Guides to Biostatistics and Epidemiology)* (1st ed.). Cambridge University Press. Available online at: https://www.amazon.co.uk/Measurement-Medicine-Practical-Biostatistics-Epidemiology/dp/0521133858/ref=asc_df_0521133858/?tag=googshopuk-21&linkCode=df0&hvadid=310805565966&hvpos=&hvnwt=g&hvrand=10626868328435984570&hvpone=&hvpstwo=&hvqmt=&hvddev=c&hvdcmldl=&hvlcint=&hvlcophy=1006483&hvtargid=pla-493171047313&psc=1&th=1&psc=1
- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., et al. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *Br. J. Psychiat.* 180, 51–60. doi: 10.1192/bjp.180.1.51
- Fackrell, K., Hall, D. A., Barry, J. G., and Hoare, D. J. (2016). Psychometric properties of the Tinnitus Functional Index (TFI): assessment in a UK research volunteer population. *Hear. Res.* 335, 220–235. doi: 10.1016/j.heares.2015.09.009
- Fackrell, K., Hall, D. A., Barry, J. G., and Hoare, D. J. (2018). Performance of the Tinnitus Functional Index as a diagnostic instrument in a UK clinical population. *Hear. Res.* 358, 74–85. doi: 10.1016/j.heares.2017.10.016
- Fava, G. A., Tomba, E., and Sonino, N. (2012). Clinimetrics: the science of clinical measurements. *Int. J. Clin. Pract.* 66, 11–15. doi: 10.1111/j.1742-1241.2011.02825.x
- Fayers, P. M., and Hand, D. J. (2002). Causal variables, indicator variables and measurement scales: an example from quality of life. *J. Royal Statist. Soc. A* 165, 233–253. doi: 10.1111/1467-985X.02020
- Fayers, P. M., and Machin, D. (2016). *Quality of Life: The Assessment, Analysis and Reporting of Patient-Reported Outcomes* (3rd ed.). Hoboken, NJ: Wiley-Blackwell. Available online at: https://www.amazon.co.uk/Quality-Life-Assessment-Reporting-Patient-Reported-dp-1444337955/dp/1444337955/ref=dp_ob_title_bk
- Flake, J. K., and Fried, E. I. (2020). Measurement schmeasurement: questionnaire measurement practices and how to avoid them. *Adv. Methods Pract. Psychol. Sci.* 3, 456–465. doi: 10.1177/2515245920952393
- Folmer, R. L. (2002). Long-term reductions in tinnitus severity. *BMC Ear Nose Throat Disord.* 2, 3. doi: 10.1186/1472-6815-2-3
- Genitsaridi, E., Hoare, D. J., Kyraios, T., and Hall, D. A. (2020). A review and a framework of variables for defining and characterizing tinnitus subphenotypes. *Brain Sci.* 10, 12. doi: 10.3390/brainsci10120938
- Gos, E., Sagan, A., Skarzynski, P. H., and Skarzynski, H. (2020). Improved measurement of tinnitus severity: study of the dimensionality and reliability of the Tinnitus Handicap Inventory. *PLoS ONE* 15, e0237778. doi: 10.1371/journal.pone.0237778
- Haider, H., Fackrell, K., Kennedy, V., and Hall, D. A. (2016). Dimensions of tinnitus-related complaints reported by patients and their significant others: protocol for a systematic review. *BMJ Open* 6, e009171. doi: 10.1136/bmjopen-2015-009171
- Hall, D. A., Haider, H., Kikidis, D., Mielczarek, M., Mazurek, B., Szczepke, A. J., et al. (2015). Toward a global consensus on outcome measures for clinical trials in tinnitus: report from the first international meeting of the COMiT initiative, November 14, 2014, Amsterdam, The Netherlands. *Trends Hear.* 19, 2331216515580272. doi: 10.1177/2331216515580272
- Hall, D. A., Haider, H., Szczepke, A. J., Lau, P., Rabau, S., Jones-Diette, J., et al. (2016). Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults. *Trials* 17, 270. doi: 10.1186/s13063-016-1399-9
- Hall, D. A., Hibbert, A., Smith, H., Haider, H. F., Londero, A., Mazurek, B., et al. (2019). One size does not fit all: developing common standards for outcomes in early-phase clinical trials of sound-, psychology-, and pharmacology-based interventions for chronic subjective tinnitus in adults. *Trends Hear.* 23, 2331216518824827. doi: 10.1177/2331216518824827
- Hallam, R. (2008). *Manual of the Tinnitus Questionnaire*. London: Polpresa Press. Available online at: <https://richardhallam.co.uk/Downloads/TinManREV5.pdf> (accessed January 02, 2024).
- Hallam, R., Rachman, S., and Hinchcliffe, R. (1984). *Psychological Aspects of Tinnitus*, 31–53. Available online at: https://www.researchgate.net/publication/306164435_Psychological_aspects_of_tinnitus
- Hanafiah, M. H. (2020). Formative Vs. Reflective Measurement Model: Guidelines for Structural Equation Modeling Research. *Int. J. Anal. Appl.* 18, 876–889. doi: 10.28924/2291-8639-18-2020-876
- Handscorn, L., Hall, D. A., Hoare, D. J., and Shorter, G. W. (2016). Confirmatory factor analysis of clinical outcomes in routine evaluation (CORE-OM) used as a measure of emotional distress in people with tinnitus. *Health Qual. Life Outc.* 14, 124. doi: 10.1186/s12955-016-0524-5
- Henry, J. A. (2016). "Measurement" of Tinnitus. *Otology Neurotol.* 37, e276–e285. doi: 10.1097/MAO.0000000000001070
- Henseler, J., and Schubert, F. (2021). *Composite-Based Structural Equation Modeling: Analyzing Latent and Emergent Variables (Methodology in the Social Sciences)* (1st ed.). New York: Guilford Press. Available online at: <https://www.amazon.co.uk/Composite-Based-Structural-Equation-Modeling-Methodology/dp/1462545602>
- Jarvis, C. B., MacKenzie, S., and Podsakoff, P. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *J. Consum. Res.* 30, 199–218. doi: 10.1086/376806
- Kuk, F. K., Tyler, R. S., Russell, D., and Jordan, H. (1990). The psychometric properties of a tinnitus handicap questionnaire. *Ear Hear.* 11, 434–445. doi: 10.1097/00003446-199012000-00005
- Markus, K. A., and Borsboom, D. (2013). *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning (Multivariate Applications Series)* (1st ed.). London: Routledge. Available online at: <https://www.amazon.co.uk/Frontiers-Test-Validity-Theory-Multivariate/dp/1841692204>
- McCormack, A., Edmondson-Jones, M., Fortnum, H., Dawes, P., Middleton, H., Munro, K. J., et al. (2014). The prevalence of tinnitus and the relationship with neuroticism in a middle-aged UK population. *J. Psychosom. Res.* 76, 56–60. doi: 10.1016/j.jpsychores.2013.08.018
- McFerran, D. J., Stockdale, D., Holme, R., Large, C. H., and Baguley, D. M. (2019). Why is there no cure for tinnitus? *Front. Neurosci.* 13, 802. doi: 10.3389/fnins.2019.00802
- McKenna, L., Handscorn, L., Hoare, D. J., and Hall, D. A. (2014). A scientific cognitive-behavioral model of tinnitus: novel conceptualizations of tinnitus distress. *Front. Neurol.* 5, 196. doi: 10.3389/fneur.2014.00196

- McNeish, D., and Wolf, M. G. (2020). Thinking twice about sum scores. *Behav. Res. Methods* 52, 2287–2305. doi: 10.3758/s13428-020-01398-0
- Meikle, M. B., Henry, J. A., Griest, S. E., Stewart, B. J., Abrams, H. B., McArdle, R., et al. (2012). The tinnitus functional index: development of a new clinical measure for chronic, intrusive tinnitus. *Ear Hear.* 33, 153–176. doi: 10.1097/AUD.0b013e31822f67c0
- Meikle, M. B., Stewart, B. J., Griest, S. E., and Henry, J. A. (2008). Tinnitus outcomes assessment. *Trends Amplificat.* 12, 223–235. doi: 10.1177/1084713808319943
- Mohamad, N. (2015). *Tinnitus: Priority Research Questions and the Impact of the Condition for Working Memory and Attention* (Doctor of Philosophy). Nottingham: University of Nottingham.
- Mohamad, N., Hoare, D. J., and Hall, D. A. (2016). The consequences of tinnitus and tinnitus severity on cognition: a review of the behavioural evidence. *Hear. Res.* 332, 199–209. doi: 10.1016/j.heares.2015.10.001
- Mohan, A., Leong, S. L., De Ridder, D., and Vanneste, S. (2022). Symptom dimensions to address heterogeneity in tinnitus. *Neurosci. Biobehav. Rev.* 134, 104542. doi: 10.1016/j.neubiorev.2022.104542
- Møller, A. R., Langguth, B., DeRidder, D., and Kleinjung, T. (2010). *Textbook of Tinnitus*. Cham: Springer Science and Business Media. doi: 10.1007/978-1-60761-145-5
- Moore, B. C. J. (2012). “The Psychophysics of Tinnitus,” in *Tinnitus*, eds J. J. Eggermont, F.-G. Zeng, A. N. Popper, and R. R. Fay (Cham: Springer New York), 187–216.
- Murray, A. L., and Booth, T. (2018). “Causal indicators in psychometrics,” in *The Wiley Handbook of Psychometric Testing* (Hoboken, NJ: JohnWiley and Sons, Ltd).
- National Institute for Health and Care Excellence (2020). *Tinnitus: Assessment and Management. Evidence Review for Questionnaires to Assess Tinnitus*. London: National Institute for Health and Care Excellence. Available online at: <https://www.nice.org.uk/guidance/ng155/evidence/e-evidence-review-for-questionnaires-to-assess-tinnitus-pdf-7089705760>
- Newman, C. W., Jacobson, G. P., and Spitzer, J. B. (1996). Development of the tinnitus handicap inventory. *Arch. Otolaryng.* 122, 143–148. doi: 10.1001/archotol.1996.01890140029007
- Newman, C. W., Sandridge, S. A., and Jacobson, G. P. (1998). Psychometric adequacy of the Tinnitus Handicap Inventory (THI) for evaluating treatment outcome. *J. Am. Acad. Audiol.* 9, 153–160.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing* (4.1.1). Available online at: <https://www.r-project.org/>
- Rhemtulla, M., van Bork, R., and Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychol. Methods* 25, 30–45. doi: 10.1037/met0000220
- Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Statist. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02
- Simoes, J. P., Schoisswohl, S., Schlee, W., Basso, L., Bernal-Robledano, A., Boecking, B., et al. (2023). The statistical analysis plan for the unification of treatments and interventions for tinnitus patients randomized clinical trial (UNITI-RCT). *Trials* 24, 472. doi: 10.1186/s13063-023-07303-2
- Spankovich. (2021). Psychoacoustics of tinnitus: lost in translation. *Acoust. Today* 15, 12. doi: 10.1121/AT.2021.17.1.35
- Spearman, C. (1904). “general intelligence,” objectively determined and measured. *Am. J. Psychol.* 15, 201. doi: 10.2307/1412107
- Tyler, R., Ji, H., Perreau, A., Witt, S., Noble, W., and Coelho, C. (2014). Development and validation of the tinnitus primary function questionnaire. *Am. J. Audiol.* 23, 260–272. doi: 10.1044/2014_AJA-13-0014
- Vanier, A., Sébille, V., Blanchin, M., and Hardouin, J.-B. (2021). The minimal perceived change: a formal model of the responder definition according to the patient’s meaning of change for patient-reported outcome data analysis and interpretation. *BMC Med. Res. Methodol.* 21, 128. doi: 10.1186/s12874-021-01307-9
- Wakabayashi, S., Oishi, N., Shinden, S., and Ogawa, K. (2020). Factor analysis and evaluation of each item of the tinnitus handicap inventory. *Head Face Med.* 16, 4. doi: 10.1186/s13005-020-00217-3