AMA Guides Proposal 100160:

Proposed Revisions, Ear, Nose, Throat, and Related Structures

Public Comments

Table of Contents:

Arranged by alphabetically by organization/society to individual

Comments Submitted by Interested Parties on a Pending AMA Guides® Editorial Change Proposal

Instructions: By submitting comments on this form regarding an Editorial Change Proposal, I attest that I have read AMA Guides® Editorial Change Proposal and Submission Requirements and will use them as the primary points of consideration when submitting the Comment Form. As an interested party, I understand that my comments are limited to the original editorial change proposal.

Name or Topic of Proposal: ENT Chapter

Individual or Organization Submitting Comments: American Psychological Association

Date: 7/15/2022

I. General Criteria for Guides Editorial Changes

- The proposed change is carefully drafted and conforms to the prevailing style of the *AMA Guides 6th Edition*;
- The terminology and the analytical frameworks used in the proposal are consistent with the World Health Organization's International Classification of Functioning, Disability, and Health (ICF);
- The structure and content of the proposed editorial change ensures that impairment ratings are transparent, clearly stated, and reproducible, to insure physician interrater reliability;
- The clinical soundness of the proposed editorial change is demonstrated with the best available evidence except in the case of minor editorial changes.

1. **Does the requested procedure meet the AMA Guides® Editorial Change Proposal and Submission Requirements?**

   ☒Yes
   ☐No
   If No, please explain. (1500 character limit)

2. **Does the submitted literature adequately support the Editorial Change Proposal?**
   ☐Yes
   ☒No
   ☐N/A

   If No, please explain. (1500 character limit)

   Please see below

3. **Are you aware of contradictory literature related to the Editorial Change Proposal?**

   ☒Yes
   ☐No
   ☐N/A

   If Yes, please include a maximum of five (5) articles when submitting this form. Articles in full text or PDF formats are required. Citations only will not be considered.

   See attached articles (Gos, 2020; Wakabayashi, 2020).

4. **Do you support this Editorial Change Proposal?**

   ☐Yes
   ☒No

   If No, please provide the rationale for lack of support, citing the specific criteria not met shown at the top of this form. (1500 character limit)

   Please see below

5. **Does the Editorial Change Proposal have any impact on other *AMA Guides* content that may not have been recognized or considered, or conflict with other precedents in the AMA Guides that might affect usage?**

☒Yes

☐No

If Yes, please explain. (1500 character limit)

Guides chapter 1 should have a more detailed discussion of standards for measures of functioning to be used in the Guides. These criteria would be used to select measures for Guides use. We would recommend the following:

1. Test development & standardization
    a. Were all stakeholders represented in measure development?
    b. Is there a test manual?
    c. Is there a standardized method of administration and scoring
    d. Does the test use standardized scores?
2. Reliability
    a. Test-retest reliability
    b. Test internally consistency
3. Validity
    a. Content validity
        i. Do the items of the instrument cover all aspects of the target condition?
        ii. Scales are statistically one-dimensional?
        iii. Do items range from mild to severe?
    b. Construct validity
        i. Correlation with other measures
        ii. Findings consistent with current understanding of the disease
    c. Bias and fairness
        i. No gender, race, age, education bias
        ii. Reading level ≤ 6^{th} grade
        iii. Minimally English & Spanish versions?
4. Norms
    a. Community norms
        i. Necessary to determine what a “normal” score is
    b. Patient norms
        i. Necessary to determine what a typical is for a patient
    c. Norms groups should be representative of the US population
5. Item response theory analysis was performed

Please provide additional commentary related to the editorial change proposal.

1. The Tinnitus Handicap Inventory (THI) was previously proposed as a functional assessment measure for tinnitus in the Guides. It appears that this has been removed and we (American Psychological Association (APA), and the APA-affiliated Interdivisional Healthcare Committee support this decision.

2. The THI is a "diagnosis-specific measure" designed to be used only with a specific diagnosis. The use of general versus specific measures point to two separate paths forward for the Guides.

    a. It was our conclusion that in the context of the Guides, diagnosis-specific measures were problematic because the use of this assessment approach would either require a very large number of measures (one fPROM for every diagnosis), or a decision to assess only specific conditions (e.g. tinnitus and low back pain), and <u>not</u> to offer any assessment of function for most other conditions.

    b. We could think of no justification for saying "Conditions A, B & C were judged to be in need of fPROM assessment, while conditions X, Y & Z were judged to be <u>not</u> in need of fPROM assessment." If the Guides wishes to implement the ICF and biopsychosocial model, assessing patient reports would be a necessary component for all impairment ratings.

3. The THI was examined in two recent psychometric reviews (Gos, 2020; Wakabayashi, 2020). Our concerns about the THI include that:

    a. The THI does not measure function, but instead is strongly associated with psychological distress attributed to tinnitus.

    b. When the THI underwent IRT analysis, a number of concerns were identified.

        i. IRT analysis determined that the three THI subscales (functional, emotional, catastrophic/catastrophizing with 11/9/5 items respectively) are invalid, as they don't measure separate domains.

        ii. The IRT analysis determined that 20 of the THI's 25 items formed a coherent scale, but concluded that 5 of the 25 THI items should be removed. These 5 items didn't "belong" with the others because they

are distinctly different, and recommended that a future version of the THI eliminate these items to address this validity problem.

iii. IRT analysis determined that the THI is most closely associated with tinnitus-related psychosocial distress, as opposed to functioning

iv. IRT analysis determined that the THI items that should be removed included items about hearing loss and difficulties with working. Overall, the THI seems to be measuring emotional distress and not physical functioning. This seemed inconsistent with the philosophy of the ICF and the Guides.

c. Another shortcoming is that the THI lacks community norms. Community norms are necessary to determine what a "normal" score is versus a high or low score.

i. Like back pain, a little bit of tinnitus may be "normal", and some studies have found that up to 15% of healthy people have this symptom. Consequently, unless you test normal subjects you don't know what the "normal" range of these symptoms is, and can't say what a high score is.

4. Overall, due to our concerns about the THI, we support that its use *for the purpose of determining the functional grade* is not recommended in the chapter.

PLOS ONE

RESEARCH ARTICLE

Improved measurement of tinnitus severity: Study of the dimensionality and reliability of the Tinnitus Handicap Inventory

Elżbieta Gos[1], Adam Sagan[2], Piotr H. Skarzynski[1,3,4], Henryk Skarzynski[5]*

1 Department of Teleaudiology and Screening, World Hearing Center, Institute of Physiology and Pathology of Hearing, Warsaw, Poland, **2** Department of Market Analysis and Marketing Research, Faculty of Management, Cracow University of Economics, Cracow, Poland, **3** Heart Failure and Cardiac Rehabilitation Department, Faculty of Medicine, Medical University of Warsaw, Warsaw, Poland, **4** Institute of Sensory Organs, Kajetany, Poland, **5** Department of Otorhinolaryngosurgery, World Hearing Center, Institute of Physiology and Pathology of Hearing, Warsaw, Poland

* p.skarzynski@ifps.org.pl

Check for updates

OPEN ACCESS

Citation: Gos E, Sagan A, Skarzynski PH, Skarzynski H (2020) Improved measurement of tinnitus severity: Study of the dimensionality and reliability of the Tinnitus Handicap Inventory. PLoS ONE 15(8): e0237778. https://doi.org/10.1371/journal.pone.0237778

Editor: Karl Bang Christensen, University of Copenhagen, DENMARK

Received: November 13, 2019

Accepted: August 3, 2020

Published: August 25, 2020

Copyright: © 2020 Gos et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors declare that no competing interests exist.

Abstract

Objective

The Tinnitus Handicap Inventory (THI) is widely used in clinical practice and research as a three-dimensional measure of tinnitus severity. Despite extensive use, its factor structure remains unclear. Furthermore, THI can be considered a reliable measure only if Cronbach's alpha coefficient and Classical Test Theory is used. The more modern and robust Item Response Theory (IRT) has so far not been used to psychometrically evaluate THI. In theory, IRT allows a more precise evaluation of THI's factor structure, reliability, and the quality of individual items.

Method

There were 1115 patients with tinnitus (556 women and 559 men), aged 19–84 years (M = 51.55; SD = 13.28).

The dimensionality of THI was evaluated using several models of Confirmatory Factor Analysis and an Item Response Theory approach. Exploratory non-parametric Mokken scaling was applied to determine a unidimensional and robust scale. Several IRT polytomous models were used to assess the overall quality of THI.

Results

The bifactor model had the best fit (RMSEA = 0.055; CFI = 0.976; SRMR = 0.040) and revealed one strong general factor and several weak specific factors. Mokken scaling generated a reliable unidimensional scale (Loevinger's $H = 0.463$). In order to refine THI we propose that five items be removed. The IRT Generalized Partial Credit Model generated good parameters in terms of item location (difficulty), discrimination, and information content of items.

Conclusion

Our findings support the use of THI to evaluate tinnitus severity in terms of it being a reliable unidimensional scale. However, clinicians and researchers should rely only on its overall score, which reflects global tinnitus severity. To improve its psychometric quality, several refinements of THI are proposed.

Introduction

In recent years there has been increasing interest among clinicians and healthcare providers in assessing patients' health status using Patient Reported Outcome Measures (PROMs). A PROM instrument is any report of the status of a patient's health that originates directly from the patient [1]. PROMs have been defined as health questionnaires which evaluate aspects of a patient's health from the patient's perspective [2].

PROMs are useful in clinical practice for diagnosis, choice of treatment, and monitoring changes. There is evidence that the use of PROMs improves patients' satisfaction, allows monitoring of response to treatment, and detects unrecognized problems [3]. In clinical trials they serve as primary or secondary endpoints [4], and they are used in health systems and in health policymaking for assessing and improving quality of care [5, 6]. The scope of PROMs' application is still expanding [7], and efforts have been made recently to ensure that the methodology of PROM use is clinically meaningful, valid, and reliable [8–11]; only then can they serve as effective instruments in enhancing healthcare quality.

PROMs are particularly useful in assessing subjective disorders which are not apparent to others but which are registered only through the complaints of the sufferers–and tinnitus is one such disorder. Tinnitus is the subjective perception of sound without any external acoustic stimulation, and is perceived as ringing in the ears, hissing, chirping, buzzing, or other sounds [12–14]. Its prevalence ranges from 4% to 15% in adults [15], and 6% to 34% in children [16–18]. Tinnitus is accompanied by a broad range of negative emotional symptoms, and significantly impacts on quality of life [19, 20]. Because of the limited effectiveness of audiological assessment and psychoacoustic measurement, self-reported rating scales and questionnaires are widely used in evaluating the severity of individually perceived tinnitus [21–23], where severity is defined as the level of distress or impact that tinnitus has on the person [24]. There is no other option for measuring tinnitus severity other than with self-reporting measures (primarily multi-item questionnaires), which need to have acceptable psychometric quality.

There are many questionnaires used for assessing tinnitus severity [23]. The Tinnitus Handicap Inventory (THI) stands out among them–it is the most commonly used tool which has been validated in the largest number of languages [25]. THI was created to evaluate the impact of tinnitus on daily living [26], and is used as a screening tool for psychiatric disorders [27], and as an outcome measure for evaluating treatment effects in clinical trials [21, 28–30]. There is a brief, time-efficient screening version which consists of only 10 items, and this has greatly increased the use of THI [31].

Although THI is a widespread tool, its factor structure remains unclear. Newman and colleagues originally postulated three factors–the Emotional, Functional, and Catastrophic subscales–but they were based on item content, not on factor analysis [26]. Factor analysis for THI was first reported for a Danish version of THI, but the study sample comprised only 50 tinnitus patients [32]. Exploratory factor analysis did not confirm a three-factor solution, indicating that only the THI total score should be used as a valid measure of tinnitus severity (not

the scores on the three subscales). In 2003, Baguley and Andersson conducted exploratory factor analysis of THI in a group of 196 patients, and the analysis gave strong support for a unifactorial structure [33]. To date, more than a dozen factor structure validation studies of THI have been published, with study groups ranging from 50 [32, 34] to 373 patients [35]. The majority of these studies failed to demonstrate a three-factor structure [36–38], although two of them did support the original three-factor solution [35, 39]. In particular, the German study seems very strong: in its confirmatory factor analysis it used a large sample of 373 tinnitus patients [35]. The findings showed that a three-factor model gave a better fit than a unidimensional model, and indicated that the three subscales of THI (Functional, Emotional, Catastrophic) were each valid and provided three distinct dimensions of tinnitus severity.

It is worth noting that work so far has used only a Classical Test Theory (CTT) approach, whereas a more modern and robust approach is now available–Item Response Theory (IRT). In this context, the factor structure of THI is not just an academic exercise but an important problem in clinical practice. It is crucial for a clinician or researcher to know which factor structure (three- or unidimensional) is appropriate to the situation and be confident they can rely on each subscale score or only on the total THI score.

The second issue which is critical to psychometric quality is reliability. The most popular index of reliability is Cronbach's alpha coefficient, which is based on CTT [40]. All studies concerning psychometric properties of THI report alpha for both total scale and for subscales. Reliability across studies appears to be very high, mostly above 0.90. Across almost all studies, alpha for the Functional and Emotional subscales ranged from 0.8 to 0.9, while for the Catastrophic subscale it was lower, about 0.6–0.7. However, Cronbach's alpha coefficient has numerous limitations [41–43], and other more robust model-based indices of reliability have recently been proposed. Reliability estimates within CTT has some limitations–they are dependent on the particular sample and measurement error is the same across all level of the ability. IRT overcomes these limitations treating reliability as precision of measurement independent of the particular sample and enabling estimation of measurement error at any given level of a latent trait.

The present study has three goals:

1. To examine the theoretical structure of tinnitus severity as measured by THI. Our starting hypothesis is that a unidimensional model best accounts for the structure of a measured construct.
2. To determine the reliability of THI in a model-based approach which has so far not been used in psychometric studies of THI.
3. To give guidance for a potential refinement of THI using Item Response Theory.

Method

Design

Our retrospective study used data from patients admitted to a tertiary referral ENT center in Poland over the period July 2015 to September 2018. Patients had reported problems with tinnitus as a primary complaint or secondary to hearing loss, and filling in THI was part of the standard diagnostic evaluation. Records of patients were retrospectively screened to check compliance with the eligibility criteria: age above 18 years, duration of tinnitus at least 1 month, documented hearing thresholds based on clinical pure tone audiometry, and a completed Tinnitus Handicap Inventory. The Institutional Review Board approved the protocol of

the study (approval no. KB IFPS 18/2018). Due the retrospective nature of our evaluation, no written consent from the participants were gathered.

Measures

The Tinnitus Handicap Inventory (THI) comprises 25 items grouped into three subscales: Functional, Emotional, and Catastrophic. The Functional subscale (11 items) deals with limitations caused by tinnitus in the areas of mental, social, and physical functioning. The Emotional subscale (9 items) concerns affective responses to tinnitus, e.g. anger, frustration, depression, anxiety. The Catastrophic subscale (5 items) probes the most severe reactions to tinnitus, such as loss of control, inability to escape from tinnitus, and fear of having a terrible disease. For each item a patient can respond with a "yes" (scored 4 points), "sometimes" (2 points), or "no" (0 points). The responses are summed within each subscale and for the total scale. The higher the score, the greater the perceived tinnitus severity [26]. The Polish version of THI validated by Skarzynski et al. [38] was used in this study.

Participants

There were 1115 individuals (556 women and 559 men); their mean age was 51.6 years (SD = 13.3) and ranged from 19 to 84 years. The period of suffering from tinnitus varied from 1 month to 50 years (M = 6.6; SD = 7.7). Most frequently, the tinnitus was bilateral (57%), while 26% of the patients reported tinnitus in the left ear and 17% in the right.

Data analysis

The first step was to evaluate the dimensionality of THI, and here four CFA models were used: a unidimensional CFA, a second-order CFA, a bifactor CFA, and a three-dimensional CFA model with correlated factors. Weighted Least Square estimation with means and variance adjustment of Chi-square statistics (WLSMV) and Theta and Delta parameterization were applied. Taking into account that the THI items are ordinal categorical variables, polychoric correlation coefficients were used. The overall fit of a CFA model was considered adequate if its Root Mean Square Error of Approximation (RMSEA) was < 0.05, the Comparative Fit Index (CFI) was > 0.95, and the Standardised Root Mean Square Residual (SRMR) < 0.05 [44].

Model-based reliability was assessed by McDonald's omega and the *H*-index, and the average variance extracted [45]. McDonald's omega was calculated as both omega total (ω) and omega hierarchical (ω_H), and for the bifactor model omega hierarchical of the subscales (ω_{HS}) and Percentage of Reliable Variance (PRV) were also calculated [46]. An omega value above 0.80 was considered high [47]. Omega hierarchical above 0.75, in conjunction with a PRV above 75%, indicates a scale's unidimensionality. Omega hierarchical subscale reflects the reliability of a subscale after controlling for the variance due to the general factor [48]. Average Variance Extracted (AVE) refers to the variance explained by a construct due only to measurement error. Fornel and Larcker stated it should be at least 0.5 [49]. The *H*-index is a measure of maximal reliability for an optimally-weighted scale, i.e. when each item contributes different information to the global score [50, 51]. The *H*-value was expected to have a minimum of 0.7.

Additional measures of dimensionality were applied in the bifactor model. Explained Common Variance (ECV) is an indicator of unidimensionality, with high ECV indicating a strong general factor compared to group factors [52]. Item Explained Common Variance (IECV) shows item-level variation attributed to a general factor [53]. ECV was used in conjunction with Percent of Uncontaminated Correlations (PUC). ECV > 0.70 and PUC > 0.70 suggest that a given construct is unidimensional [47]. Average Relative Parameter Bias (ARPB) occurs

when multidimensionality is ignored and a unidimensional model is specified [47]. An ARPB less than 10–15% is considered acceptable [54].

The second step involved exploring non-parametric Mokken scaling to check for the monotonicity of items. Selection of the best items for unidimensional parametric IRT modeling was carried out via an automated item selection procedure using a genetic algorithm. In terms of the IRT approach, the scalability of the THI scale was measured using Loevinger's H [55]. If the item scalability coefficients $H_{ij} > 0$, $H_i > 0.3$, and $H > 0.3$ then this suggests a reliable, cumulative scale.

In the third step, three IRT polytomous models were used to assess unidimensional THI scale quality: the Rasch Model for polytomous items, the Generalized Partial Credit Model (GPCM, an extension of the Rasch model) with parameters for item discrimination and adjacent-category response functions [56], and the Graded Response Model (for ordered polytomous categories of a Likert scale and with cumulative category response functions) [57]. The overall fit was checked using the M2 statistic [58]. Marginal reliability was computed, given an estimated model and a prior density function; marginal reliability above 0.7 suggests an acceptable scale. The local independence assumption was checked using Yen's Q_3 statistic based on correlation of the residuals for a pair of items [59]. The final scale was developed on the basis of model-based reliability, item goodness of fit, and item information functions.

The sample size was calculated using power 0.80 and alpha level 0.05, assuming 3 latent variables, 25 observed variables, and an anticipated effect size of 0.1. The required minimum sample was 823 individuals. Statistical analyses were performed with IBM SPSS Statistics v.24, Mplus 8.2, and the mirt, ltm, eRm, and mokken libraries of the R package.

Results

Basic statistics

Descriptive statistics for the THI items and its subscales are summarized in Table 1. The majority of correlations between individual items and the total score were above 0.5, making the whole scale seem reliable.

Dimensionality of CTT- and IRT-based measurement models

Before testing multidimensional models, CFA unidimensional analyses of the Functional, Emotional and Catastrophic subscales were conducted using WLSMV method.

For Functional subscale: χ^2 (44) = 295.14; $p < 0.001$; RMSEA (Root Mean Square Error Of Approximation) = 0.072; CFI (Comparative Fit Index) = 0.978; SRMR (Standardized Root Mean Square Residual) = 0.042. After controlling for correlated errors (based on modification index) items THI7 with THI20, and THI7 with THI2, χ^2 (42) = 247.06; $p < 0.001$; RMSEA = 0.066; CFI = 0.982; SRMR = 0.038.

For Emotional subscale: χ^2 (27) = 234.24; $p < 0.001$; RMSEA = 0.083; CFI = 0.983; SRMR = 0.035.

After controlling for correlated errors items THI3 with THI14, THI25 with THI17 and THI25 with THI22, χ^2 (24) = 111.02; $p < 0.001$; RMSEA = 0.057; CFI = 0.993; SRMR = 0.023.

For Catastrophic subscale: χ^2 (5) = 82.81; $p < 0.001$; RMSEA = 0.118; CFI = 0.967; SRMR = 0.045.

After controlling for correlated errors items THI8 with THI19, the fit drastically has been improved: χ^2 (4) = 5.83; $p < 0.001$; RMSEA = 0.020; CFI = 0.999; SRMR = 0.012.

Afterwards, four CFA models for the whole THI were tested and they are set out in Figs 1–4.

Results of dimensionality analysis and comparison of models of goodness of fit are shown in Table 2.

Table 1. Descriptive statistics for THI.

	Yes (%)	Sometimes (%)	No (%)	M	SD	Corrected item–total correlation	Cronbach's alpha if item deleted
THI 1 F	33.3	43.5	23.2	2.20	1.49	0.68	0.939
THI 2 F	33.5	30.8	35.7	1.96	1.66	0.37	0.943
THI 3 E	29.8	39.0	31.2	1.97	1.56	0.65	0.939
THI 4 F	18.0	31.7	50.3	1.35	1.52	0.63	0.940
THI 5 C	17.2	30.6	52.2	1.30	1.51	0.69	0.939
THI 6 E	42.7	42.2	15.1	2.55	1.42	0.59	0.940
THI 7 F	35.2	33.1	31.7	2.07	1.64	0.49	0.941
THI 8 C	66.5	20.3	13.2	3.07	1.43	0.48	0.941
THI 9 F	27.3	27.7	45.0	1.64	1.66	0.66	0.939
THI 10 E	26.0	38.2	34.9	1.84	1.57	0.69	0.939
THI 11 C	24.0	33.5	42.5	1.63	1.59	0.46	0.942
THI 12 F	30.5	34.9	34.6	1.92	1.61	0.75	0.938
THI 13 F	22.5	34.0	43.5	1.58	1.57	0.67	0.939
THI 14 E	30.4	43.6	26.0	2.09	1.50	0.73	0.938
THI 15 F	28.3	32.5	39.2	1.78	1.63	0.57	0.940
THI 16 E	35.2	36.5	28.3	2.14	1.59	0.64	0.939
THI 17 E	17.1	25.2	57.7	1.19	1.53	0.61	0.940
THI 18 F	17.7	42.0	40.3	1.55	1.46	0.69	0.939
THI 19 C	69.3	19.7	11.0	3.17	1.36	0.42	0.942
THI 20 F	35.9	39.4	24.7	2.22	1.54	0.69	0.939
THI 21 E	31.2	37.4	31.4	2.00	1.58	0.75	0.938
THI 22 E	20.8	25.9	53.3	1.35	1.60	0.63	0.940
THI 23 C	22.2	37.6	40.2	1.64	1.54	0.72	0.938
THI 24 F	45.5	22.0	32.5	2.26	1.75	0.37	0.943
THI 25 E	29.2	27.3	43.5	1.71	1.68	0.67	0.939
	Range			M	SD	Cronbach's alpha	Cronbach's alpha if items 2,8,13,19,24 were deleted
Functional	0–44			20.53	11.71	0.875	0.865
Emotional	0–36			16.84	10.30	0.893	0.893
Catastrophic	0–20			10.81	5.17	0.731	0.696
THI total	0–100			48.18	25.27	0.942	0.942

Capital letters represent items contained on the subscales: F–Functional, E–Emotional, C- Catastrophic.

Corrected item-total correlation is a correlation between the item and the scale score that excludes this item.

Items excluded in subsequent analysis are in bold.

https://doi.org/10.1371/journal.pone.0237778.t001

All the CTT models had acceptable goodness of fit, taking into account the values of fit indices. However, the bifactor model had a significantly better fit in comparison with the correlated factor model, which was slightly superior to the unidimensional model. In the family of IRT models, bifactor GPCM and unidimensional GPCM had the best fit (M2 statistic); however the SRMR of bifactor GPCM appeared to be too high. In summary, both CTT and IRT confirmatory models suggest a more detailed elaboration of the unidimensional and bifactor models is needed in order to verify the unidimensionality of THI.

Model reliability

Reliability was evaluated for the two best models: unidimensional and bifactor. Results are gathered together in Table 3.

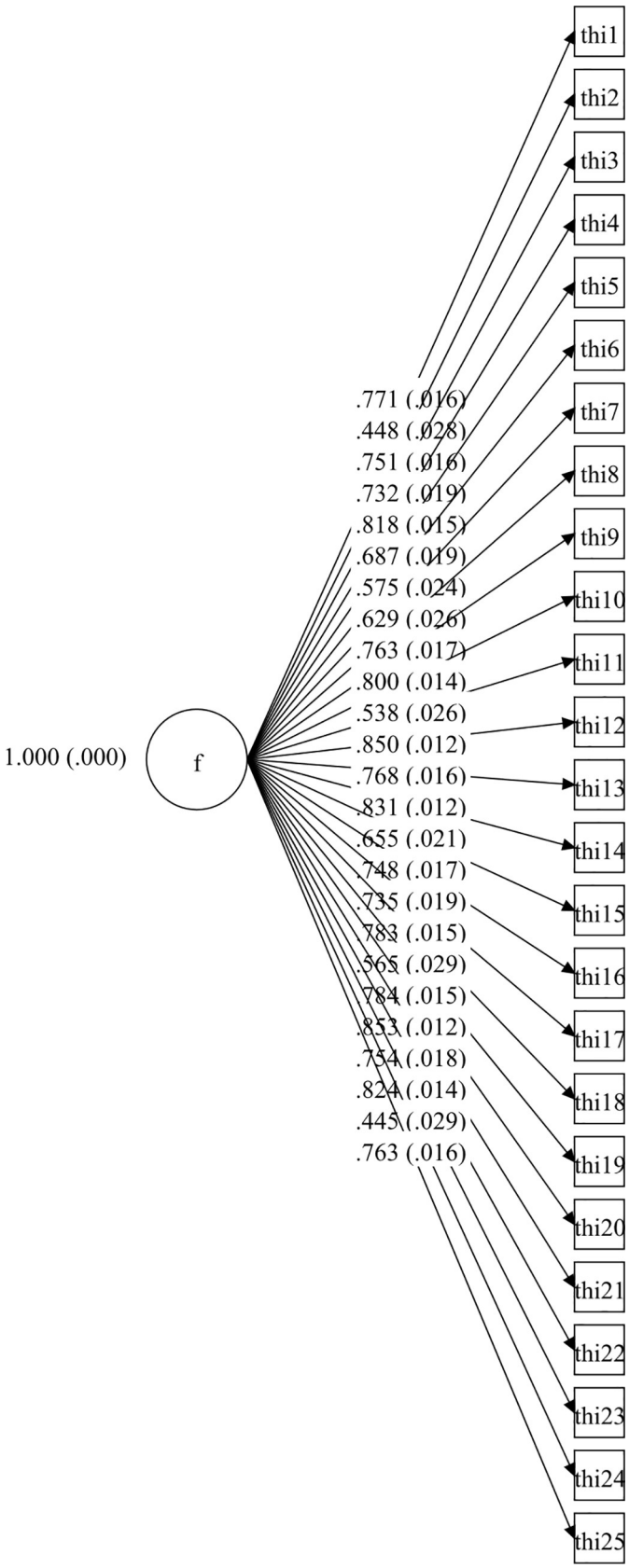


Fig 1. Unidimensional CFA model of Tinnitus Handicap Inventory.

https://doi.org/10.1371/journal.pone.0237778.g001

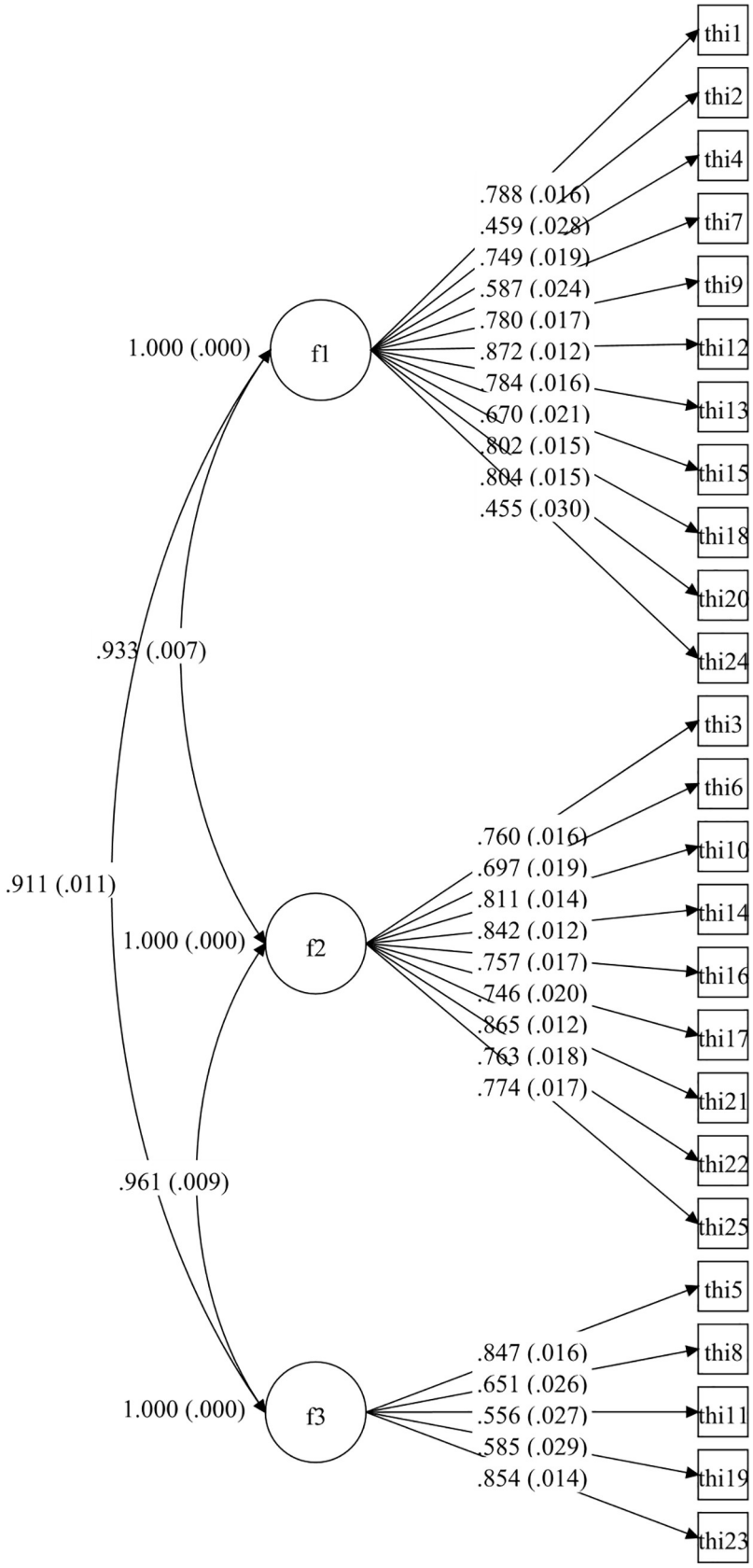


Fig 2. Three-dimensional CFA model of Tinnitus Handicap Inventory.

https://doi.org/10.1371/journal.pone.0237778.g002

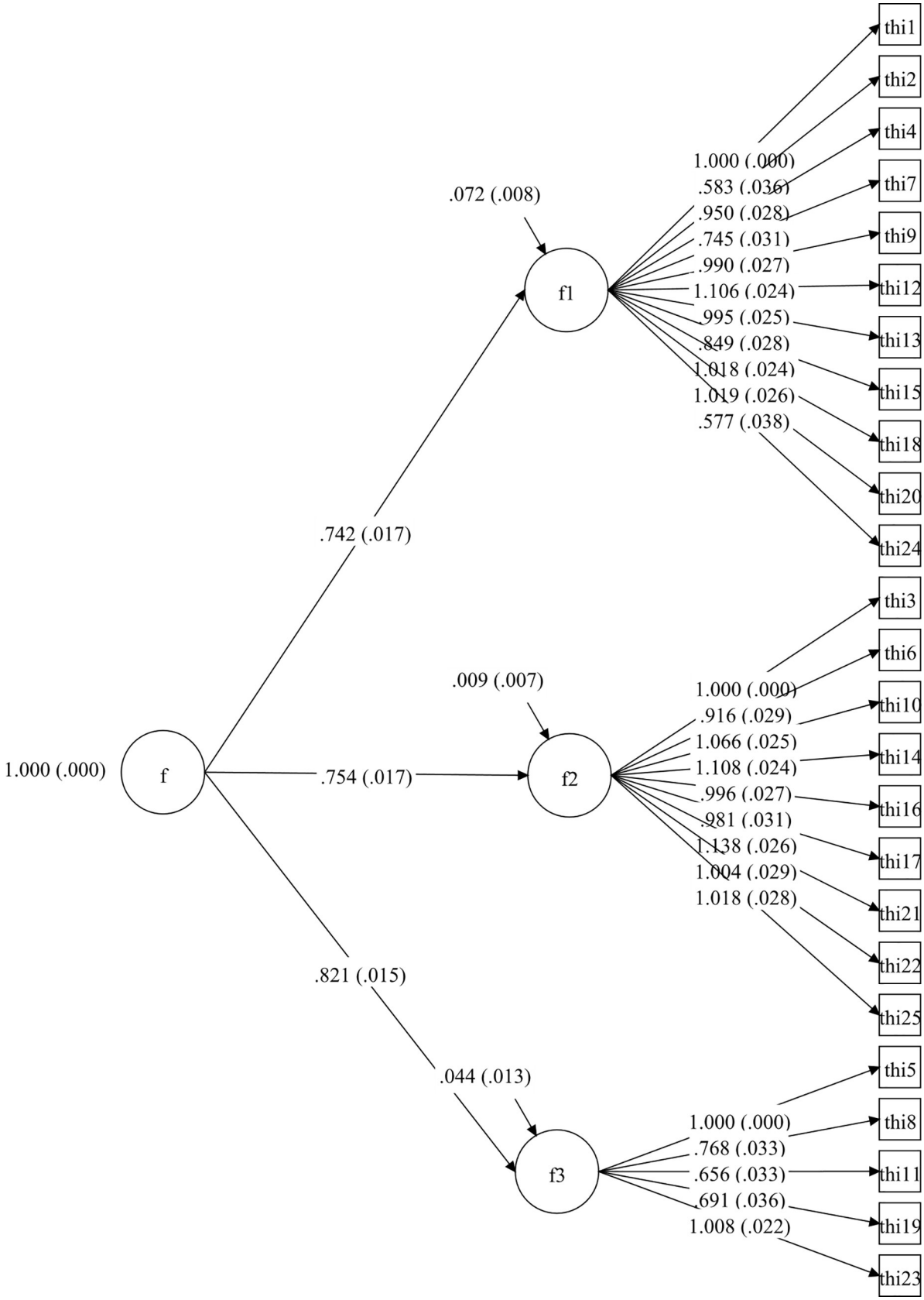


Fig 3. Second-order CFA model of Tinnitus Handicap Inventory.

https://doi.org/10.1371/journal.pone.0237778.g003

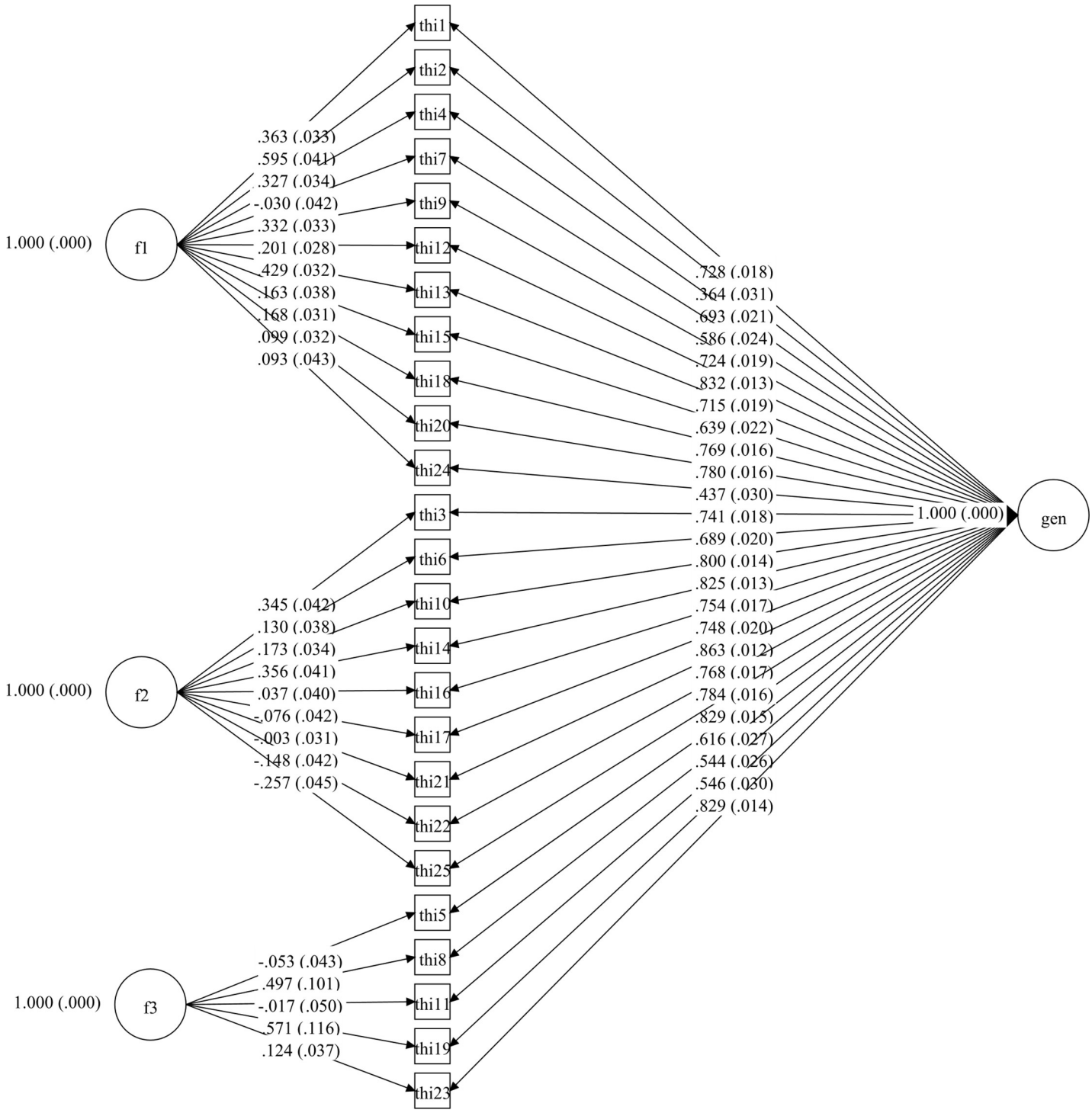


Fig 4. Bifactor CFA model of Tinnitus Handicap Inventory.

https://doi.org/10.1371/journal.pone.0237778.g004

The unidimensional model had acceptable reliability. The bifactor model showed high overall and sub-dimension reliability; however a unidimensional solution was most strongly supported. $Omega_H = 0.945$ showed that total score predominantly reflects a single general

Table 2. Goodness of fit for various THI models.

CTT factor models					
Model	**Chi-square**	**Df**	**RMSEA**	**CFI**	**SRMR**
Unidimensional	1701.362	275	0.068	0.960	0.051
Correlated factors	1522.900	272	0.064	0.965	0.048
Second-order	1493.146	272	0.064	0.964	0.049
Bifactor	1101.318	250	0.055	0.976	0.040
Model comparison	**Delta Chi-square**	**Delta df**	**Delta RMSEA**	**Delta CFI**	**Delta SRMR**
Bifactor vs Correlated factors	391.752 ($p<0.001$)	22	0.009	0.011	0.008
Correlated factors vs Unidimensional	148.837 ($p<0.001$)	3	0.013	0.016	0.003
IRT models					
Model	**M2**	**df**	**RMSEA**	**CFI**	**SRMR**
Unidimensional					
GPCM	1637.937	248	0.071	0.967	0.046
Rasch	1923.624	273	0.074	0.961	0.111
Correlated factors					
GPCM	-	-	-	-	-
Rasch	1750.035	266	0.071	0.965	0.113
Second-order					
GPCM	3496.855	245	0.109	0.9223	0.327
Rasch	-	-	-	-	-
Bifactor					
GPCM	822.086	193	0.054	0.984	0.164
Rasch	3075.998	265	0.094	0.933	0.409
Model comparisons	**Delta Chi^2**	**Delta df**	**Delta RMSEA**	**Delta CFI**	**Delta SRMR**
Bifactor vs correlated factors Rasch (ANOVA)	-1527.237 ($p>0.99$)	1	0.023	0.032	0.296
Correlated factors vs unidimensional Rasch (ANOVA)	99.684 ($p<0.001$)	7	0.003	0.004	0.002
Bifactor vs unidimensional Rasch (ANOVA)	-1427.554 ($p>0.99$)	8	0.020	0.028	0.298

- Model failed to converge

RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index; SRMR, Standardised Root Mean Square Residual

https://doi.org/10.1371/journal.pone.0237778.t002

factor. Omegas for the subscales scores seemed to demonstrate high reliability for the THI sub-factors, but low values of ω_{HS} indicated that almost all sub-scale score variance is due to the general factor and almost no variance is due to specific factors. It also indicated the heavy confounding of sub-scale reliability (reliabilities of sub-scales were overwhelmingly inflated). Also PRV values confirmed that the three subdimensions of the THI scale are questionable and suggest that the scale is undimensional. General ECV values also suggested the scale is unidimensional, with ECVs for sub-scales meaningless. The Difference ARP bias between the unidimensional scale and the general factor in the bifactor model was acceptable. Only PUC = 0.66 showed that there might be some multi-dimensionality in THI; however, it was not severe enough to disqualify the interpretation of the instrument as being primarily unidimensional. The individual explained common variance (IECV) indicated that almost all items well represent the unidimensional THI scale except items THI2 and THI19, which were less than 0.50. The best items for unidimensional THI scale having the highest IECV were THI21, THI11, THI16, THI7, THI5, THI6, THI17, THI20, THI23, THI22, and THI24.

In general, all criteria of dimensionality analysis (ω_H, ω_{HS}, PRV, ECV, PUC, and ARPB) gave sufficient support for scale unidimensionality. In the subsequent analysis, unidimensional IRT-based models are adopted to assess the monotonicity and quality of each THI item.

Table 3. Reliability of unidimensional and bifactor models.

	Uni-dimensional model	Bifactor model			
		F	E	C	THI total
ω	0.967	0.930	0.941	0.856	0.971
ω_H		0.114	0.005	0.079	0.945
ω_{HS}		0.021	0.001	0.003	-
H	0.970	0.570	0.300	0.460	0.950
AVE	0.550	0.187	0.166	0.199	0.581
PRV		0.123	0.006	0.093	0.973
ECV		0.066	0.029	0.033	0.872
IECV		THI1 = 0.801, THI2 = 0.272, THI3 = 0.822, THI4 = 0.818, THI5 = 0.996, THI6 = 0.966, THI7 = 0.997, THI8 = 0.605, THI9 = 0.826, THI10 = 0.955, THI11 = 0.999, THI12 = 0.945 THI13 = 0.735, THI14 = 0.843, THI15 = 0.939, THI16 = 0.998, THI17 = 0.990, THI18 = 0.955, THI19 = 0.478, THI20 = 0.984, THI21 = 1.000, THI22 = 0.964, THI23 = 0.978, THI24 = 0.956, THI25 = 0.903			
ARPB		0.03			
PUC		0.660			

F, Functional subscale; E, Emotional subscale; C, Catastrophic subscale; THI total, THI total score; ω, McDonald's omega; ω_H, omega hierarchical; ω_{HS}, omega hierarchical subscale; H, Bentler's index; AVE, Average Variance Extracted; PRV, Percentage of Reliable Variance; ECV, Explained Common Variance; IECV, Individual Explained Common Variance; ARPB, Average Relative Parameter Bias; PUC, Percent of Uncontaminated Correlations.

https://doi.org/10.1371/journal.pone.0237778.t003

Exploratory Mokken model of the unidimensional THI scale

Having verified unidimensionality and the cumulative character of the THI scale, an exploratory nonparametric Mokken model was used to evaluate the scale's monotonicity and to select items. All the item scalability coefficients H_{ij} between pairs of items were positive ($H_{ij} > 0$) and ranged between 0.127 (THI2–THI7) and 0.733 (THI5–THI10). THI2 and THI24 were regarded as the weakest items ($H_i < 0.3$). The Loevinger H for the total scale was 0.463 (SE = 0.011). Additional reliability measures (MS and LCRC) showed reliable unidimensional scale: MS = 0.909, LCRC = 0.949. Also, the Automated Item Selection Procedure (AISP) for the Mokken scale using a genetic algorithm confirmed unidimensionality, (except items THI2 and THI24). The relationships between H_i and IECV measures are plotted in Fig 5.

On the basis of existing sub-scales, model fit, H_i, and IECV values we propose a shortened unidimensional THI scale that consists of only the "best" items. The selection is based on linear ordering (Hellwig method) and the geometric average of H_i and IECV scores. Items THI2, THI8, THI13, THI19, and THI24 were thus removed from the original scale, and the 20 remaining items were selected for unidimensional parametric polytomous IRT models.

Item quality of IRT-based models

IRT analysis results for the three IRT models are summarized in Table 4.

The test information curves of compared models are given in Fig 6.

The Rasch model was rejected and the GPCM and GRM models seemed to be the most appropriate. The GPCM model was chosen for further analysis.

The reliability of all the models was above the threshold of 0.7 and between –2.5 and +2.5 standard deviations from the average level of the standardized latent trait. The GPCM model included 93.05% of respondents who fitted the model and it was selected for more detailed analysis of items and individual person's reliability.

The Yen's Q_3 statistic was used to test the assumption of local independence. The mean value was –0.025 and Q_3 ranged between –0.107 and 0.160. The mean Q_3 value was less than

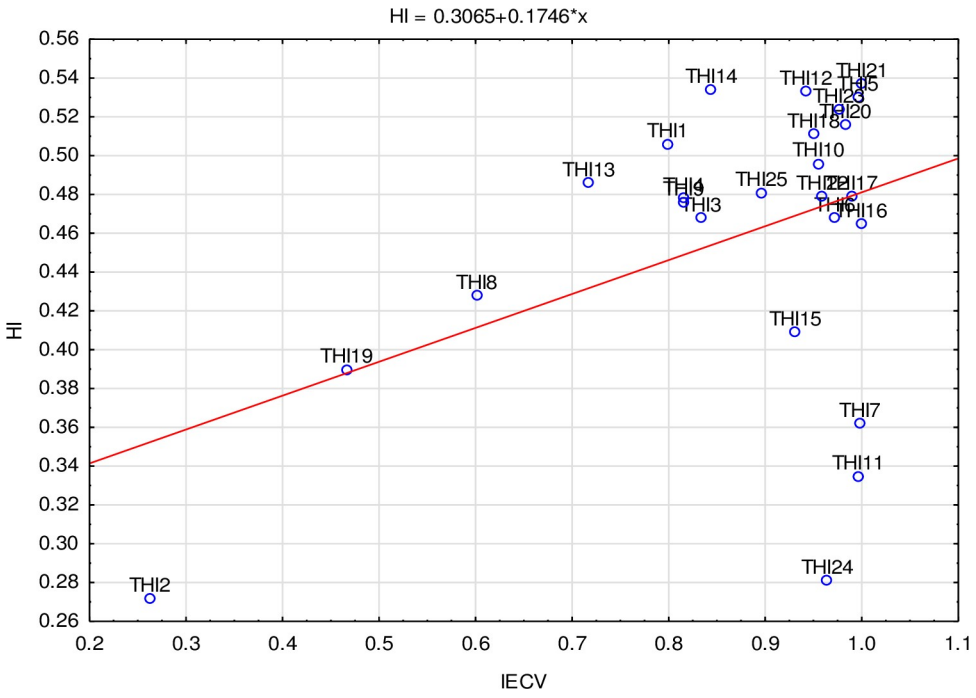


Fig 5. Relationship between H_i and IECV of THI items.

https://doi.org/10.1371/journal.pone.0237778.g005

the threshold value of 0.1 and indicated that the local independence assumption was valid. Additionally, correlations between standardized residuals correlations were calculated and they are gathered in Table 5. The mean value for residual correlations was -0.007 and they ranged between -0.5 to 0.14, and for only one pair of items it was rather high (-0.5).

The parameters of the GPCM model are given in Table 6. Item locations (difficulties) were calculated as an average of threshold parameters for item response categories (for three item categories, two thresholds exist).

Item difficulties ranged between –0.656 (THI6) and 0.798 (THI17), item discrimination between 0.703 (THI11) to 2.440 (THI21), and item information between 1.40 (THI11) and 4.88 (THI21). For those item information values between –2 and 2 standardized values of Θ (latent trait continuum), where the THI scale has the highest precision, the item information values were between 0.940 (THI11) and 4.74 (THI 21), which are shown in Fig 7.

Discussion

Despite widespread use of THI, there are still doubts about its psychometric quality. The first doubt has to do with its unclear factor structure, which means it is not certain whether THI correctly gauges aspects of tinnitus severity. Originally, it was postulated that THI measures three domains of tinnitus severity: functional, emotional, and catastrophic. They were intended to be distinct, although strongly correlated [26].

Our findings do not support these assumptions. Our findings show that, for the clinical population, the original three-factor structure is not the best measure of tinnitus severity. Omega hierarchical sub-scale indices showed that the proportion of the total variance accounted for by the three subscales was, after controlling for the influence of general tinnitus severity, very small. Other indices (AVE, ECV, PUC, PRV, ARPB) showed that the common variance can be regarded as unidimensional, thus supporting one general factor and a

Table 4. Goodness of fit of IRT-based models.

Model level	**Rasch model**	**GPCM**	**GRM**
M2; df (p-level)	1047.694; 169 (<0.001)	926.420;150 (<0.001)	901.572;150 (<0.001)
RMSEA	0.068	0.068	0.067
AIC	37705.55	37197.73	37125.99
Test information function value	63.71	63.71	65.69
Model marginal reliability	0.879	0.929	0.932
Item level	**Chi-square (p-level)**	**Chi-square (p-level)**	**Chi-square (p-level)**
THI1	46.295 (0.792)	43.238 (0.772)	43.054 (0.857)
THI3	41.764 (0.906)	43.262 (0.852)	42.900 (0.901)
THI4	49.804 (0.637)	49.285 (0.725)	51.550 (0.800)
THI5	75.261* (0.024)	62.993 (0.086)	55.121 (0.254)
THI6	72.026* (0.022)	69.026 (0.057)	64.774 (0.172)
THI7	164.566* (<0.001)	59.596 (0.598)	61.102 (0.614)
THI9	65.717 (0.227)	59.063 (0.436)	60.253 (0.467)
THI10	59.197 (0.360)	54.009 (0.360)	57.335 (0.284)
THI11	219.252* (0.000)	52.031 (0.911)	52.481 (0.918)
THI12	83.330* (0.016)	43.736 (0.686)	59.375 (0.197)
THI14	97.678* (<0.001)	75.718* (0.007)	78.870* (0.003)
THI15	116.421* (<0.001)	64.191 (0.400)	69.604 (0.325)
THI16	47.831 (0.773)	47.204 (0.763)	49.085 (0.732)
THI17	38.489 (0.933)	50.166 (0.727)	44.273 (0.923)
THI18	77.759* (0.019)	68.544* (0.042)	69.173* (0.046)
THI20	54.012 (0.512)	44.856 (0.679)	50.947 (0.593)
THI21	90.779* (0.002)	44.792 (0.523)	45.004 (0.556)
THI22	55.634 (0.413)	55.135 (0.470)	57.788 (0.557)
THI23	80.512* (0.027)	45.652 (0.610)	47.467 (0.574)
THI25	61.698 (0.312)	60.652 (0.380)	61.531 (0.421)

* significant at $p<0.05$

https://doi.org/10.1371/journal.pone.0237778.t004

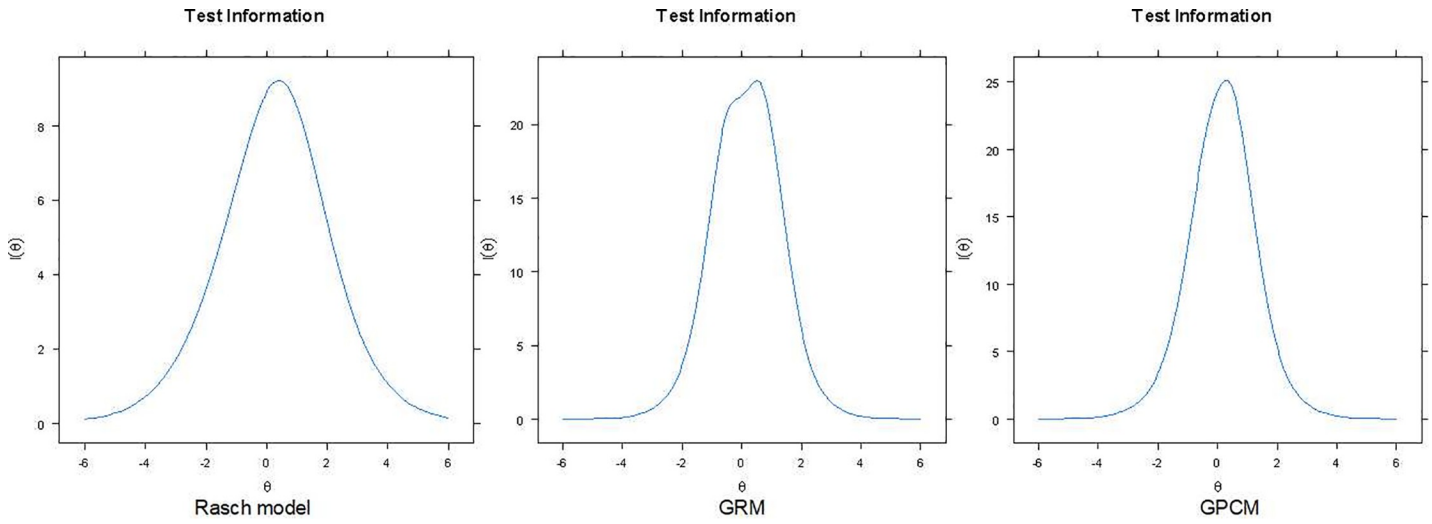


Fig 6. Test information curves of IRT models of THI (20 items).

https://doi.org/10.1371/journal.pone.0237778.g006

Table 5. Correlations between standardized residuals (GPCM Model).

	THI1	THI3	THI4	THI5	THI6	THI7	THI9	THI10	THI11	THI12	THI14	THI15	THI16	THI17	THI18	THI20	THI21	THI22	THI23
THI1																			
THI3	-0.029																		
THI4	0.109	0.044																	
THI5	-0.065	0.089	0.120																
THI6	-0.104	0.082	-0.100	0.063															
THI7	0.075	0.075	-0.054	-0.054	0.086														
THI9	0.060	-0.047	0.077	-0.060	-0.085	-0.080													
THI10	-0.050	0.081	-0.100	0.135	-0.099	-0.028	0.087												
THI11	-0.039	-0.025	0.056	0.088	0.066	-0.015	-0.036	-0.054											
THI12	0.041	-0.076	-0.075	0.079	-0.105	-0.074	0.128	-0.121	-0.025										
THI14	-0.04	0.143	-0.029	-0.109	0.141	0.050	-0.069	0.133	-0.058	-0.097									
THI15	0.083	-0.043	0.072	0.067	-0.032	0.105	-0.079	-0.070	-0.039	-0.021	-0.041								
THI16	-0.095	0.034	-0.064	-0.079	0.124	-0.054	-0.067	-0.068	0.139	-0.031	-0.074	-0.055							
THI17	0.101	-0.053	0.071	-0.096	-0.089	-0.084	0.122	-0.082	-0.044	0.064	0.109	0.062	-0.026						
THI18	0.089	-0.049	-0.063	0.076	0.113	0.064	-0.040	-0.071	0.060	-0.075	-0.096	0.101	-0.060	0.069					
THI20	0.074	-0.087	-0.052	-0.060	-0.108	0.060	0.060	0.055	-0.033	-0.073	0.098	0.046	-0.072	-0.109	-0.072				
THI21	-0.082	-0.061	-0.071	-0.083	0.110	-0.066	-0.045	-0.129	0.050	0.080	-0.075	-0.062	0.116	0.041	-0.088	0.039			
THI22	-0.042	-0.065	0.060	0.109	-0.068	0.072	-0.078	-0.064	0.089	-0.089	-0.50	-0.082	0.078	-0.056	-0.081	-0.062	0.069		
THI23	-0.073	-0.082	-0.093	0.069	0.053	0.059	-0.081	-0.081	0.060	-0.094	-0.054	0.061	0.084	-0.062	0.079	0.085	0.086	0.091	
THI25	0.075	-0.053	0.114	-0.124	-0.042	-0.059	0.072	-0.094	0.069	0.056	-0.052	-0.054	-0.047	0.071	-0.061	-0.056	-0.059	0.088	-0.054

https://doi.org/10.1371/journal.pone.0237778.t005

Table 6. Item parameters of GPCM model for THI (20 items).

	Model parameters				
Items	**Threshold 1**	**Threshold 2**	**Item location**	**Discrimination**	**Info**
THI1	-0.947	0.516	-0.216	1.671	3.34
THI3	-0.577	0.601	0.018	1.570	3.14
THI4	0.211	1.108	0.659	1.321	2.64
THI5	0.165	1.092	0.628	1.948	3.90
THI6	-1.482	0.169	-0.656	1.307	2.61
THI7	-0.408	0.219	-0.090	0.826	1.65
THI9	0.114	0.545	0.329	1.324	2.65
THI10	-0.437	0.715	0.139	1.852	3.71
THI11	0.144	0.869	0.506	0.703	1.40
THI12	-0.430	0.554	0.062	2.194	4.39
THI14	-0.782	0.600	-0.091	2.285	4.57
THI15	-0.102	0.565	0.231	1.006	2.01
THI16	-0.668	0.379	-0.144	1.479	2.96
THI17	0.559	1.037	0.798	1.334	2.67
THI18	-0.280	1.165	0.442	1.865	3.73
THI20	-0.843	0.392	-0.225	1.770	3.54
THI21	-0.556	0.543	-0.006	2.440	4.88
THI22	0.372	0.837	0.604	1.438	2.88
THI23	-0.265	0.894	0.314	2.128	4.26
THI25	0.051	0.459	0.255	1.394	2.79

https://doi.org/10.1371/journal.pone.0237778.t006

unidimensional solution. These results are in line with our previous research [38] and they are also consistent with those obtained by others [32, 33, 36, 37]. This contrasts with the earlier German study of 373 tinnitus patients [35], which confirmed the three-factor structure of THI.

However, it should be noted that the German study compared only a general factor model and a first-order three-factor model. They did not consider a second-order three-factor model or a bifactor model. It is known that a bifactor model is useful for evaluating the validity of multi-item questionnaires which measure both the overall construct and its specific dimensions [47]. In our case, however, the results of bifactor modelling clearly demonstrated that there was a one factor solution. Our results demonstrate that THI should be considered a unidimensional scale, and that the Functional, Emotional, and Catastrophic subscales do not represent separate substantive latent traits. Instead, we believe these subscale share a large portion of overall general negative affectivity associated with tinnitus.

THI is generally considered to be a reliable tool. The claim about high reliability of THI subscales and overall score, demonstrated by several validating studies, is founded on the use of Cronbach's alpha coefficient. But it is worth emphasizing, that reliability depends on a particular study population, while IRT offers in its place test information function, which shows the degree of precision at different values of the latent trait. Fig 7 clearly shows that the standard error of measurement (SEM) is the smallest in the middle of the scale and increases with higher and lower scores. So, the precision of measurement is the highest for the subjects with moderate tinnitus severity. When Cronbach's alpha is embedded in CTT theory, it is assumed that SEM is constant along the scale, and this is, as we can see, an unfounded assumption. Other drawbacks of this index can be found elsewhere [41–43]. Our findings demonstrate that THI is in fact reliable as a unidimensional scale (with no subscales) in our large sample tinnitus

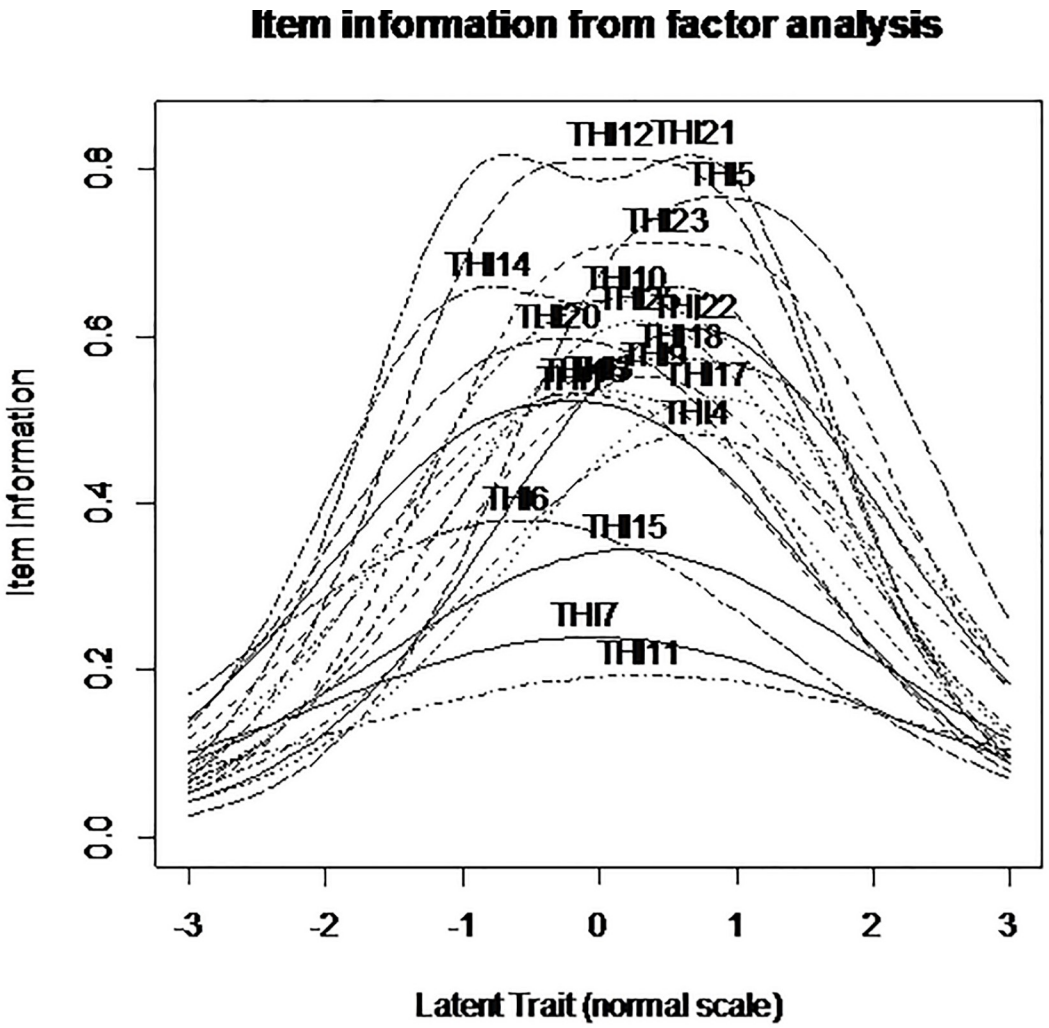


Fig 7. Item information functions.

https://doi.org/10.1371/journal.pone.0237778.g007

sufferers, and its precision of measurement is the highest for subjects with moderate complaints.

Mokken analysis confirmed the unidimensionality of THI and allows us to treat it as a reliable cumulative scale. On the basis on several combined criteria, we propose that five items (THI2, THI8, THI13, THI19, THI24) should be removed in order to refine the scale. Three of these excess items belong to the original Functional subscale, while two belong to the Catastrophic subscale. Of the remaining 20 items, the majority cover the emotional aspect of tinnitus. This allows the whole scale to be more consistent, but it does narrow the range of tinnitus which THI measures. Kennedy and colleagues [60] noted that THI, compared to other tinnitus-related questionnaires, contains a disproportionately large number of items related to psychological/emotional aspects of tinnitus. The results of our study also suggest that tinnitus severity as measured by THI captures mainly the emotional aspects of tinnitus. This may be

either a disadvantage or an advantage, depending on whether THI is used in a clinical or research setting and the underlying goal.

We must admit, that application IRT models to the THI posed some difficulties. Model fit statistic (M2) was significant for all tested models. It needs some comment [61], just like significant χ^2 test values in previous analyses. First of all, CTT and IRT models represent an accept-support approach to model testing, where many "near perfect" models tend to be falsely "rejected". Secondly, the χ^2 statistic is generally susceptible to sample size therefore RMSEA, incremental fit indices and inspection of residuals and residuals correlations were developed and used to support model fit. Thirdly, the IRT models are predominantly psychometric not pure statistical/econometric models, therefore are focused on quality of data (given IRT model) rather than quality of model itself and model improving through its far-reaching respecification. Additionally, the problem of local independence should be also addressed. We used Yen's Q_3 statistic, however as it was shown by Christensen et al. [62] a singular critical value for Q_3 is not fully appropriate and local dependence should be rather considered relative to the average observed residual correlation.

A great advantage and practical application of IRT is in-depth analysis of individual items, which may be used in selecting items during development or refinement of a questionnaire. Item location (level of difficulty) reflects where along the scale the item functions best. Items displaying a low level of item location (e.g. THI6 –*complaining a great deal about tinnitus*) are the 'easiest' items, indicating endorsement of mild tinnitus severity, while items with high item location (e.g. THI17 –*bad social relationship*) are the 'hardest' and they target a higher level of tinnitus severity. Informative items and discrimination were highest for THI21 (*depression*), THI14 (*irritation*), THI12 (*difficulty to enjoy life*), THI23 (*can no longer cope with tinnitus*); while the lowest were for THI11 (*having a terrible disease*) and THI7 (*trouble with sleep*). IRT parameters indicate which items should be selected to optimize measurement precision and achieve the desired goal of the tool. Items providing more information on lower-level traits are suitable for gauging mild tinnitus severity, while items targeting higher-level traits should be selected to optimize measurement of high tinnitus severity, e.g. in monitoring change over time following treatment. Item information function of THI displayed in Fig 7 clearly shows that THI in its present form is good in assessing individuals in the range $\Theta = -1$ to 1, i.e. those with a moderate level of tinnitus severity.

Our findings have important clinical and research implications. The unidimensional factor structure of THI allows clinicians to use the tool without unnecessary additional calculations for subscales, thus saving time. Clinicians or researchers should rely only on the global score, because validity of the three subscales (Functional, Emotional, Catastrophic) is questionable, as they appear to provide little information beyond the general factor (overall tinnitus severity). We conclude that the quality of THI in its current form (25 items) is not satisfactory. Newman and colleagues proposed a short version of THI consisting of only 10 items [31], but they were selected on the basis of just three criteria: a high item–total correlation, representativeness of the three content domains, and face validity. We find such criteria insufficient and propose refining the THI instrument by removing just those items with some degree of misfit. We think that short form questionnaires are essential in busy clinical practice and with extensive research protocols, and we recommend taking into account both the CTT and IRT approaches in constructing a short form of THI.

The strength of our THI study is the large sample of tinnitus patients–the largest assembled so far. Patients came from all over Poland to our tinnitus clinic, so the sample can be considered representative of individuals seeking help for tinnitus. However, it is true that a more heterogeneous sample (e.g. in terms of geographic origin) would reduce the potential selection bias that our data might have.

We admit that not all aspects of IRT analysis have been exhausted in this study. Differential Item Functioning (DIF) analysis was omitted due to constraints on the length of this paper. Therefore, we still are unable to say how to interpret between-group comparisons shown with THI (e.g. difference in tinnitus severity between women and men) as true difference or measurement artifact. Further research is needed to establish measurement invariance in various demographic settings and cross-cultural comparisons.

To conclude, the growth of patient-centered care requires high-quality data from Patient Reported Outcome Measures. Application of IRT theory enables more precise assessment of the THI measurement properties, so that clinicians and researchers can have more confidence about their diagnoses and the results of trials based on THI.

We hope our findings might encourage researchers to use the IRT approach to explore the psychometric properties of other tinnitus-related questionnaires. Done well, we expect it will improve the quality of measures based on patients' perception of their ailment.

Supporting information

S1 Raw data.
(XLSX)

Acknowledgments

We acknowledge the practice staff and patients who participated in the study. We would also like to thank PhD Andrew Bell for proofreading the manuscript.

Author Contributions

Conceptualization: Elżbieta Gos, Adam Sagan, Piotr H. Skarzynski, Henryk Skarzynski.

Data curation: Elżbieta Gos.

Formal analysis: Elżbieta Gos, Adam Sagan.

Funding acquisition: Piotr H. Skarzynski, Henryk Skarzynski.

Methodology: Elżbieta Gos, Adam Sagan.

Project administration: Henryk Skarzynski.

Resources: Piotr H. Skarzynski, Henryk Skarzynski.

Software: Adam Sagan.

Supervision: Henryk Skarzynski.

Writing – original draft: Elżbieta Gos, Adam Sagan, Piotr H. Skarzynski.

Writing – review & editing: Piotr H. Skarzynski, Henryk Skarzynski.

References

1. Food and Drug Administration. Guidance for Industry. FDA; 2009.
2. Cappelleri J, Zou K, Bushmakin A, Alvir J, Alemayehu D, Symonds T. Patient-Reported Outcomes: Measurement, Implementation and Interpretation. Chapman and Hall/CRC; 2014.
3. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res. 2013; 13:211. https://doi.org/10.1186/1472-6963-13-211 PMID: 23758898

4. Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. Patient Relat Outcome Meas. 2018; 9:353–67. https://doi.org/10.2147/PROM.S156279 PMID: 30464666

5. Garratt AM, Bjaertnes ØA, Krogstad U, Gulbrandsen P. The OutPatient Experiences Questionnaire (OPEQ): data quality, reliability, and validity in patients attending 52 Norwegian hospitals. Qual Saf Health Care. 2005; 14(6):433–7. https://doi.org/10.1136/qshc.2005.014423 PMID: 16326790

6. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ. 2015; 350:g7818. https://doi.org/10.1136/bmj.g7818 PMID: 25670183

7. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. Med Care. 2013; 51(8 Suppl 3):S73–79.

8. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002; 11(3):193–205. https://doi.org/10.1023/a:1015291021312 PMID: 12074258

9. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60 (1):34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012 PMID: 17161752

10. Valderas JM, Ferrer M, Mendívil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health. 2008; 11 (4):700–8. https://doi.org/10.1111/j.1524-4733.2007.00309.x PMID: 18194398

11. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010; 19(4):539–49. https://doi.org/10.1007/s11136-010-9606-8 PMID: 20169472

12. Møller AR. Introduction. In: Møller AR, Langguth B, DeRidder D, Kleinjung T, editors. Textbook of Tinnitus. New York: Springer-Verlag; 2011. pp. 3–7.

13. Tunkel DE, Bauer CA, Sun GH, Rosenfeld RM, Chandrasekhar SS, Cunningham ER, et al. Clinical practice guideline: tinnitus. Otolaryngol Head Neck Surg. 2014; 151(2 Suppl):S1–40.

14. Jastreboff PJ. 25 years of tinnitus retraining therapy. HNO. 2015; 63(4):307–11. https://doi.org/10.1007/s00106-014-2979-1 PMID: 25862626

15. Møller AR. Epidemiology of tinnitus in adults. In: Møller AR, Langguth B, DeRidder D, Kleinjung T, editors. Textbook of Tinnitus. New York: Springer-Verlag; 2011. pp. 29–37.

16. Savastano M. Characteristics of tinnitus in childhood. Eur J Pediatr. 2007; 166(8):797–801. https://doi.org/10.1007/s00431-006-0320-z PMID: 17109163

17. Skarzynski PH, Kochanek K, Skarzynski H, Senderski A, Szkielkowska A, Bartnik G, et al. Hearing Screening Program in School-Age Children in Western Poland. Int Adv Otol. 2011; 7(2):194–200.

18. Piotrowska A, Raj-Koziak D, Lorens A, Skarżyński H. Tinnitus reported by children aged 7 and 12 years. Int J Pediat Otorhinolaryngolog. 2015; 79(8):1346–50.

19. Langguth B. A review of tinnitus symptoms beyond „ringing in the ears": a call to action. Curr Med Res Opin. 2011; 27(8):1635–43. https://doi.org/10.1185/03007995.2011.595781 PMID: 21699365

20. Zeman F, Koller M, Langguth B, Landgrebe M. Which tinnitus-related aspects are relevant for quality of life and depression: results from a large international multicentre sample. Health Qual Life Outcomes. 2014; 12:7. https://doi.org/10.1186/1477-7525-12-7 PMID: 24422941

21. Langguth B, Goodey R, Azevedo A, Bjorne A, Cacace A, Crocetti A, et al. Consensus for tinnitus patient assessment and treatment outcome measurement: Tinnitus Research Initiative meeting, Regensburg, July 2006. Prog Brain Res. 2007; 166:525–36. https://doi.org/10.1016/S0079-6123(07)66050-6 PMID: 17956816

22. Meikle MB, Stewart BJ, Griest SE, Henry JA. Tinnitus Outcomes Assessment. Trends Amplif. 2008; 12 (3):223–35. https://doi.org/10.1177/1084713808319943 PMID: 18599500

23. Fackrell K, Hall D, Barry J, Hoare D. Tools for tinnitus measurement: development and validity of questionnaires to assess handicap and treatment effects. IN: Signorelli F, Turjman F, editors. Tinnitus: causes, treatment and short and long-term health effects. New York: Nova Biomedical; 2014. pp. 13–60.

24. Cima DRFF, Mazurek PDB, Haider H, Kikidis D, Lapira A, Noreña AJ, et al. A multidisciplinary European guideline for tinnitus: diagnostics, assessment, and treatment. HNO. 2019; 67:10–42. https://doi.org/10.1007/s00106-019-0633-7 PMID: 30847513

25. Skarżyński Piotr H., Rajchel Joanna J., Gos Elżbieta, Dziendziel Beata, Kutyba Justyna, Świerniak Weronika, et al. A revised grading system for the Tinnitus Handicap inventory based on a large clinical population. Int J Audiol. 2020; 59(1):61–67. https://doi.org/10.1080/14992027.2019.1664778 PMID: 31608728

26. Newman CW, Jacobson GP, Spitzer JB. Development of the Tinnitus Handicap Inventory. Arch Otolaryngol Head Neck Surg. 1996; 122(2):143–8. https://doi.org/10.1001/archotol.1996.01890140029007 PMID: 8630207

27. Salviati M, Macrì F, Terlizzi S, Melcore C, Provenzano A, Capparelli E, et al. The Tinnitus Handicap Inventory as a screening test for psychiatric comorbidity in patients with tinnitus. Psychosomatics. 2013; 54(3):248–56. https://doi.org/10.1016/j.psym.2012.05.007 PMID: 23219227

28. McCombe A, Baguley D, Coles R, McKenna L, McKinney C, Windle-Taylor P, et al. Guidelines for the grading of tinnitus severity: the results of a working group commissioned by the British Association of Otolaryngologists, Head and Neck Surgeons, 1999. Clin Otolaryngol Allied Sci. 2001; 26(5):388–93. https://doi.org/10.1046/j.1365-2273.2001.00490.x PMID: 11678946

29. Gudex C, Skellgaard PH, West T, Sørensen J. Effectiveness of a tinnitus management programme: a 2-year follow-up study. BMC Ear Nose Throat Disord. 2009; 9:6. https://doi.org/10.1186/1472-6815-9-6 PMID: 19558680

30. Zeman F, Koller M, Figueiredo R, Aazevedo A, Rates M, Coelho C, et al. Tinnitus handicap inventory for evaluating treatment effects: which changes are clinically relevant? Otolaryngol Head Neck Surg. 2011; 145(2):282–7. https://doi.org/10.1177/0194599811403882 PMID: 21493265

31. Newman CW, Sandridge SA, Bolek L. Development and psychometric adequacy of the screening version of the tinnitus handicap inventory. Otol Neurotol. 2008; 29(3):276–81. https://doi.org/10.1097/MAO.0b013e31816569c4 PMID: 18277308

32. Zachariae R, Mirz F, Johansen LV, Andersen SE, Bjerring P, Pedersen CB. Reliability and validity of a Danish adaptation of the Tinnitus Handicap Inventory. Scand Audiol. 2000; 29(1):37–43. https://doi.org/10.1080/010503900424589 PMID: 10718675

33. Baguley DM, Andersson G. Factor analysis of the Tinnitus Handicap Inventory. Am J Audiol. 2003; 12 (1):31–4. https://doi.org/10.1044/1059-0889(2003/007) PMID: 12894865

34. Oron Y, Sergeeva NV, Kazlak M, Barbalat I, Spevak S, Lopatin AS, et al. A Russian adaptation of the tinnitus handicap inventory. International Journal of Audiology. 2015; 54(7):485–9. https://doi.org/10.3109/14992027.2014.996823 PMID: 25620408

35. Kleinstäuber M, Frank I, Weise C. A confirmatory factor analytic validation of the Tinnitus Handicap Inventory. J Psychosom Res. 2015; 78(3):277–84. https://doi.org/10.1016/j.jpsychores.2014.12.001 PMID: 25582803

36. Meng Z, Zheng Y, Liu S, Wang K, Kong X, Tao Y, et al. Reliability and validity of the chinese (mandarin) tinnitus handicap inventory. Clin Exp Otorhinolaryngol. 2012; 5(1):10–6. https://doi.org/10.3342/ceo.2012.5.1.10 PMID: 22468196

37. Bolduc D, Désilets F, Tardif M, Leroux T. Validation of a French (Québec) version of the Tinnitus Handicap Inventory. Int J Audiol. 2014; 53(12):903–9. https://doi.org/10.3109/14992027.2014.935495 PMID: 25140601

38. Skarzynski PH, Raj-Koziak D, J Rajchel J, Pilka A, Wlodarczyk AW, Skarzynski H. Adaptation of the Tinnitus Handicap Inventory into Polish and its testing on a clinical population of tinnitus sufferers. Int J Audiol. 2017; 56(10):711–5. https://doi.org/10.1080/14992027.2017.1319080 PMID: 28537137

39. Aqeel M, Ahmed A. Translation, Adaptation and Cross Language Validation of Tinnitus Handicap Inventory in Urdu. J Audiol Otol. 2017; 22(1):13–9. https://doi.org/10.7874/jao.2017.00108 PMID: 29325390

40. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951; 16(3):297–334.

41. Schmitt N. Uses and abuses of coefficient alpha. Psychological Assessment. 1996; 8(4):350–3.

42. Sijtsma K. On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. Psychometrika. 2009; 74(1):107–20. https://doi.org/10.1007/s11336-008-9101-0 PMID: 20037639

43. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol. 2014; 105(3):399–412. https://doi.org/10.1111/bjop.12046 PMID: 24844115

44. Hooper D, Coughlan J, Mullen M. Structural Equation Modelling: Guidelines for Determining Model Fit. Articles [Internet]. 1 styczeń 2008; Available from://arrow.dit.ie/buschmanart/2

45. Sagan A. Analiza rzetelności skal w wielopoziomowych modelach pomiaru (Reliability Analysis in Multilevel Measurement Models). In: Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu (Research Papers of Wrocław University of Economics). Wrocław: Uniwersytet Ekonomiczny we Wrocławiu; 2014. pp. 49–59.

46. Rodriguez A, Reise SP, Haviland MG. Evaluating bifactor models: Calculating and interpreting statistical indices. Psychol Methods. 2016; 21(2):137–50. https://doi.org/10.1037/met0000045 PMID: 26523435

47. Rodríguez AC, Reise SP, Haviland MG. Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. J Pers Assess. 2016; 98(3):223–37. https://doi.org/10.1080/00223891.2015.1089249 PMID: 26514921

48. Reise SP, Bonifay WE, Haviland MG. Scoring and modeling psychological measures in the presence of multidimensionality. J Pers Assess. 2013; 95(2):129–40. https://doi.org/10.1080/00223891.2012.725437 PMID: 23030794

49. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. J Market Res. 1981; 18(1):39–50.

50. Bentler P. Covariance Structure Models for Maximal Reliability of Unit-Weighted Composites. In: Lee S, editor. Handbook of Latent Variable and Related Models. New York: Elsevier; 2007. pp. 1–19.

51. Hancock G, Mueller R. Rethinking Construct Reliability within Latent Variable Systems. In: Cudeck R, du Toit S, Sorbom D, editors. Structural Equation Modeling: Present und Future—A Festschrift in Honor of Karl Joreskog. Lincolnwood, IL: Scientific Software International; 2001. pp. 195–216.

52. Reise SP. The rediscovery of bifactor measurement models. Multivar Behav Res. 2012; 47(5):667–96.

53. Stucky BD, Thissen D, Edelen MO. Using Logistic Approximations of Marginal Trace Lines to Develop Short Assessments. Appl Psychol Meas. 2013; 37(1):41–57.

54. Muthén B, Kaplan D, Hollis M. On structural equation modeling with data that are not missing completely at random. Psychometrika. 1987; 52(3):431–62.

55. Mokken RJ. A Theory and Procedure of Scale Analysis, With Applications in Political Research [Internet]. Reprint 2011. Berlin, Boston: De Gruyter Mouton; 2011. Available on: https://www.degruyter.com/view/product/46584

56. Muraki E. A Generalized Partial Credit Model: Application of an Em Algorithm. ETS Research Report Series. 1992;(1):i–30.

57. Samejima F. Estimation of Latent Ability Using a Response Pattern of Graded Scores1. ETS Research Bulletin Series. 1968; 1968(1):i–169.

58. Maydeu-Olivares A, Joe H. Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2n Contingency Tables. Journal of the American Statistical Association. 2005; 100(471):1009–20.

59. Yen WM. Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. Appl Psychol Meas. 1984; 8(2):125–45.

60. Kennedy V, Wilson C, Stephens D. Quality of life and tinnitus. Audiol Med 2009; 2:29–40.

61. Barrett P. Structural equation modelling: Adjudging model fit. Pers Individ Dif. 2007; 42(5):815–24.

62. Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. Appl Psychol Meas. 2017; 41(3):178–94. https://doi.org/10.1177/0146621616677520 PMID: 29881087

Wakabayashi et al. Head & Face Medicine (2020) 16:4
https://doi.org/10.1186/s13005-020-00217-3

Head & Face Medicine

RESEARCH Open Access

Factor analysis and evaluation of each item of the tinnitus handicap inventory

Satoko Wakabayashi[1,2], Naoki Oishi[1*], Seiichi Shinden[1,3] and Kaoru Ogawa[1]

Abstract

Purpose: This study aims to examine the availability of subscales in the Tinnitus Handicap Inventory (THI) originally proposed by Newman and the possibility of other useful subscales. We also examine whether each item of the THI could be used to better understand the status of patients with tinnitus.

Methods: This study included 1332 patients who answered the THI on their first visit. Confirmatory factor analysis was conducted to the 25 items of the THI to confirm the usefulness of the emotional, functional, and catastrophic subscales. Exploratory factor analysis was performed to discover the availability of other suitable subscales in addition to the proposed subscales. The proportion of patients who chose "yes" in each item of the THI was also examined to understand the status of patients with tinnitus.

Results: In the confirmatory factor analysis, the emotional, functional, and catastrophic subscales did not fit the model. In the exploratory factor analysis, data were extremely biased to one factor. Examination of each item of the THI showed the tendency of worsening of comorbid symptoms when tinnitus handicap became worse.

Conclusions: As a result of the factor analysis, only the total score, not any subscale, would be clinically useful in the THI. Examination of each item of the THI was helpful to understand the status of patients with tinnitus and comorbid symptoms of tinnitus. It is necessary to consider treatment by taking these comorbid symptoms into account.

Keywords: Tinnitus, THI (tinnitus handicap inventory), Subscales, Total score, Comorbid symptoms

Background

Tinnitus Handicap Inventory (THI) is a reliable and valid questionnaire to evaluate tinnitus-related disability in patients with tinnitus [1]. Tinnitus Research Initiative (TRI), an international academic organization founded in 2006, recommended the use of THI for evaluation of tinnitus handicap and therapeutic effect [2].

The THI has been translated into many languages and used internationally. The translations include Italian [3], Brazilian Portuguese [4], Chinese (Chinese-Mandarin [5] and Chinese-Cantonese [6]), etc. Reliability and validity have been demonstrated for these translations. In Japan, the Japanese version of THI validated by Shinden et al. is used [7]. The used questionnaires to assess the construct validity differ among languages. In Italian version, the MOS (Medical Outcomes Study) 36-Item Short Form Health Survey (SF-36) [8] and the Hospital Anxiety and Depression Scale (HADS) [9] were used, while the Beck Depression Inventory (BDI) [10] was used in Brazilian Portuguese version. In Japanese version, the Self-Rating Depression Scale (SDS) [11] and the State-Trait Anxiety Inventory (STAI) [12] were used [13]. The TRI stated that differences in the culture, language, and health care system have a significant impact on tinnitus evaluation [2].

In daily medical practice, the total score of THI is mainly used, and the severity is divided based on the total score. There is a report concluded that the THI could only be discussed by the total score [14], whereas a recent report concluded that, by using factor analysis, the THI can be discussed by subscales originally indicated by Newman [15]. In Baguley's report, data were collected from 80 patients with tinnitus and 116 patients with vestibular schwannomas awaiting surgery. In Kleinstauber's report, data were collected from 373 patients

* Correspondence: oishin@keio.jp
[1]Department of Otorhinolaryngology-Head and Neck Surgery, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
Full list of author information is available at the end of the article

© The Author(s). 2020 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

with tinnitus who were recruited through the internet for different research.

In the present study, more than 1000 patients presented to a hospital with tinnitus were recruited, and this large amount of data were considered to be meaningful to evaluate the contradictory results. This study aims to examine the appropriateness of using the THI subscales indicated by Newman for discussion and to determine the possibility of other useful subscale apart from the total score. We also examine whether each item of the THI could be used to better understand the status of patients with tinnitus.

Methods

The present study included a total of 1332 patients with tinnitus, 696 men (52.2%) and 636 females (47.7%), who visited the department of otolaryngology in Keio University Hospital between 1 January 2004 and 31 December 2011 and answered the THI on their first visit. The average age of the subjects was 58.7 ± 13.8 years. We used the Japanese version of THI [7]. Patients who could not complete the Japanese version of THI due to their age or language were excluded. The THI consists of 25 items. Each question item of the Japanese version of THI is listed in Table 1. The original version of THI is listed in Table 2.

In the THI, scores of 0, 2, or 4 are assigned to each answer, and thus the total score ranges from 0 to 100. Higher THI scores indicate a greater handicap from tinnitus, and five categories are used: (i) no handicap (0–16), (ii) mild handicap (18–36), (iii) moderate handicap (38–56), (iv) severe handicap (58–76), and (v) catastrophic handicap (78–100) [16]. The three subscales indicated by Newman: functional, emotional, and catastrophic were included. The distribution of each question to subscales is described in Tables 1 and 2.

Confirmatory factor analysis was conducted to the 25 items of the THI to confirm the usefulness of the three proposed subscales. Subsequently, exploratory factor

Table 1 The Japanese version of Tinnitus Handicap Inventory

	Subscale[a]		よくある	たまにある	ない
1	F	耳鳴のために物事に集中できない。	4	2	0
2	F	耳鳴の音が大きくて人の話が聞き取れない。	4	2	0
3	E	耳鳴に対して腹が立つ。	4	2	0
4	F	耳鳴のために混乱してしまう。	4	2	0
5	C	耳鳴のために絶望的な気持ちになる。	4	2	0
6	E	耳鳴について多くの不満を訴えてしまう。	4	2	0
7	F	夜眠るときに耳鳴が妨げになる。	4	2	0
8	C	耳鳴から逃れられないかのように感じる。	4	2	0
9	F	あなたの社会的活動が耳鳴により妨げられている。(例えば、外食をする、映画を観るなど)	4	2	0
10	E	耳鳴のために挫折を感じる。	4	2	0
11	C	耳鳴のために自分がひどい病気であるように感じる。	4	2	0
12	F	耳鳴があるために日々の生活を楽しめない。	4	2	0
13	F	耳鳴が職場や家庭での仕事の妨げになる。	4	2	0
14	E	耳鳴のためにいらいらする。	4	2	0
15	F	耳鳴のために読書ができない。	4	2	0
16	E	耳鳴のために気が動転する。	4	2	0
17	E	耳鳴のために家族や友人との関係にストレスを感じる。	4	2	0
18	F	耳鳴から意識をそらすのは難しいと感じる。	4	2	0
19	C	自分一人で耳鳴を管理していくのは難しいと感じる。	4	2	0
20	F	耳鳴のために疲れを感じる。	4	2	0
21	E	耳鳴のために落ち込んでしまう。	4	2	0
22	E	耳鳴のために体のことが心配になる。	4	2	0
23	C	耳鳴とこれ以上つき合っていけないと感じる。	4	2	0
24	F	ストレスがあると耳鳴がひどくなる。	4	2	0
25	E	耳鳴のために不安な気持ちになる。	4	2	0

[a]F represents an item contained on the functional subscale; E, an item contained on the emotional subscale; and C, an item contained on the catastrophic subscale

Table 2 The original version of Tinnitus Handicap Inventory

	Subscale[a]		Yes	Sometimes	No
1	F	Because of your tinnitus, is it difficult for you to concentrate?	4	2	0
2	F	Does the loudness of your tinnitus make it difficult for you to hear people?	4	2	0
3	E	Does your tinnitus make you angry?	4	2	0
4	F	Does your tinnitus make you feel confused?	4	2	0
5	C	Because of your tinnitus, do you feel desperate?	4	2	0
6	E	Do you complain a great deal about your tinnitus?	4	2	0
7	F	Because of your tinnitus, do you have trouble falling to sleep at night?	4	2	0
8	C	Do you feel as though you cannot escape your tinnitus?	4	2	0
9	F	Does your tinnitus interfere with your ability to enjoy your social activities (such as going out to dinner, to the movies)?	4	2	0
10	E	Because of your tinnitus, do you feel frustrated?	4	2	0
11	C	Because of your tinnitus, do you feel that you have a terrible disease?	4	2	0
12	F	Does your tinnitus make it difficult for you to enjoy life?	4	2	0
13	F	Does your tinnitus interfere with your job or household responsibilities?	4	2	0
14	E	Because of your tinnitus, do you find that you are often irritable?	4	2	0
15	F	Because of your tinnitus, is it difficult for you to read?	4	2	0
16	E	Does your tinnitus make you upset?	4	2	0
17	E	Do you feel that your tinnitus problem has placed stress on your relationships with members of your family and friends?	4	2	0
18	F	Do you find it difficult to focus your attention away from your tinnitus and on other things?	4	2	0
19	C	Do you feel that you have no control over your tinnitus?	4	2	0
20	F	Because of your tinnitus, do you often feel tired?	4	2	0
21	E	Because of your tinnitus, do you feel depressed?	4	2	0
22	E	Does your tinnitus make you feel anxious?	4	2	0
23	C	Do you feel that you can no longer cope with your tinnitus?	4	2	0
24	F	Does your tinnitus get worse when you are under stress?	4	2	0
25	E	Does your tinnitus make you feel insecure?	4	2	0

[a]F represents an item contained on the functional subscale; E, an item contained on the emotional subscale; and C, an item contained on the catastrophic subscale

analysis was performed to discover the availability of other suitable subscales in addition to the proposed subscales. Besides, the proportion of patients who chose "yes" (score 4) in each THI item was examined in each severity. We picked up items if the proportion of the patients who chose "yes" exceeded 10%, and examined the tendency of which items increased when tinnitus became severe.

For statistical analysis, SPSS Statistics 24 (IBM, New York, NY, USA) was used for calculation of average value and exploratory factor analysis, and SPSS Amos 24 (IBM, New York, NY, USA) for confirmatory factor analysis. Statistical values were considered significant at 5% level.

Results

The average THI total score on the first visit was 54.2 ± 26.0. The tinnitus severity distributions of the subjects were 8.9% with no handicap ($n = 119$), 19.9% with mild handicap ($n = 265$), 24.1% with moderate handicap ($n = 321$), 22.7% with severe handicap ($n = 303$), and 24.3% with catastrophic handicap ($n = 324$).

Confirmatory factor analysis

A three-factor model was set using the three subscales and confirmatory factor analysis was conducted. Results are shown in Fig. 1. Next, Good of Fit Index (GFI), Adjusted GFI (AGFI), Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI) were used as indicators of model fitness. GFI, AGFI, and CFI should be 0.95 or higher and RMSEA 0.06 or lower [17, 18]. As a result, in the three-factor model of THI, none of the criteria were met, i.e., GFI = 0.844, AGFI = 0.814, RMSEA = 0.080, CFI = 0.895. Besides, the correlation between the three factors was very high. These results suggested that the three subscales did not fit the model.

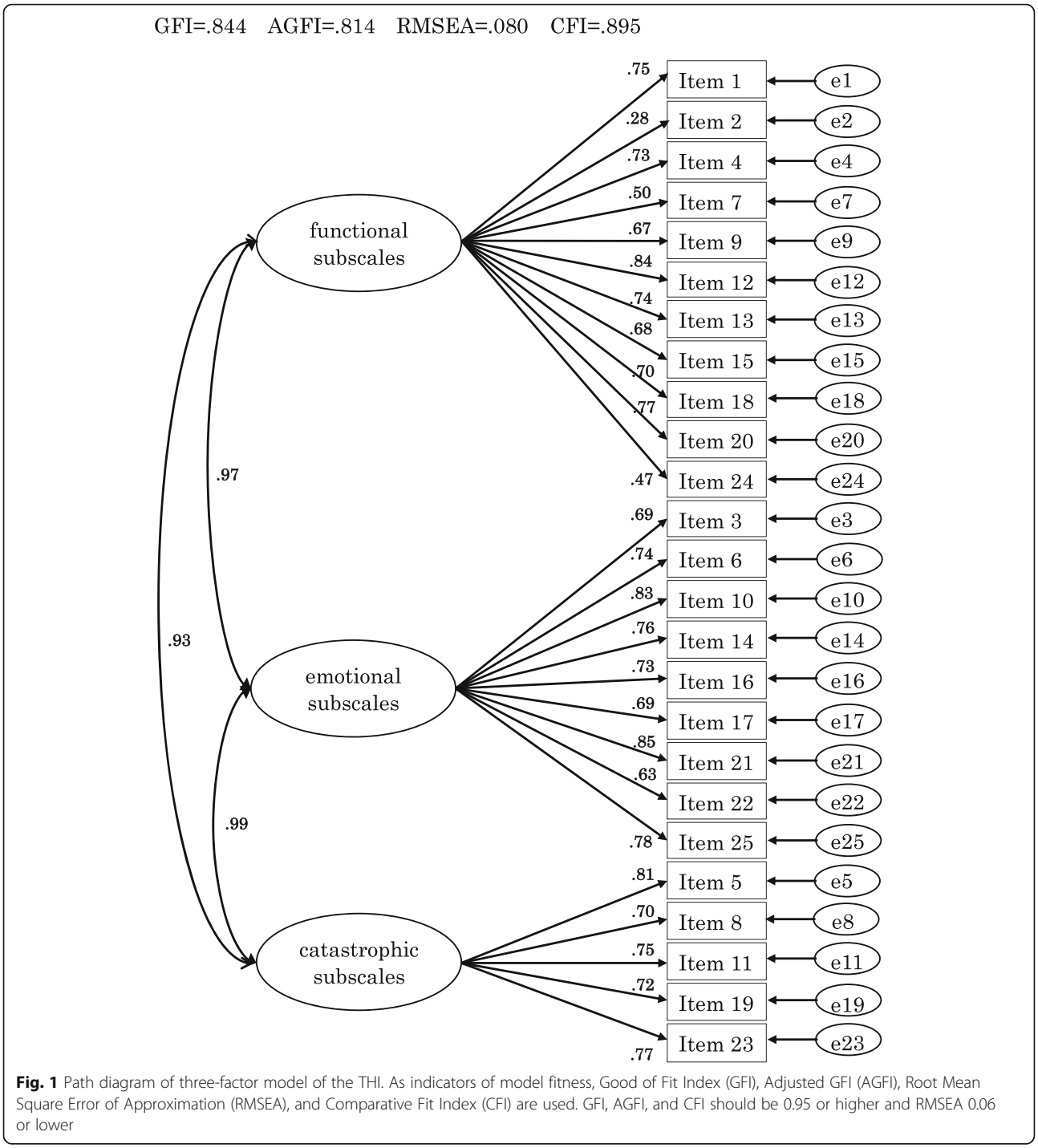


Fig. 1 Path diagram of three-factor model of the THI. As indicators of model fitness, Good of Fit Index (GFI), Adjusted GFI (AGFI), Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI) are used. GFI, AGFI, and CFI should be 0.95 or higher and RMSEA 0.06 or lower

Exploratory factor analysis

For the 25 items of THI, we used maximum likelihood method and promax rotation method to perform exploratory factor analysis. The scree plot is shown in Fig. 2, and the factor matrix is shown in Table 3. The data were extremely biased to one factor, and there were only two factors of which eigenvalue exceeded one. The eigenvalues for the factors were 12.88 (Factor 1) and 1.45 (Factor 2), explaining 57.3% of the variance. Adding more factors contributed little variance, with Factor 3 adding 3.6% (eigenvalue = 0.91) and Factor 4 adding 3.4% (eigenvalue = 0.86). The correlation between Factor 1 and Factor 2 was very high (0.77), and in the goodness of fit test, the result suggested that the factors did not fit the model ($X^2 = 1660.89$, df = 251, $p = 0.000$).

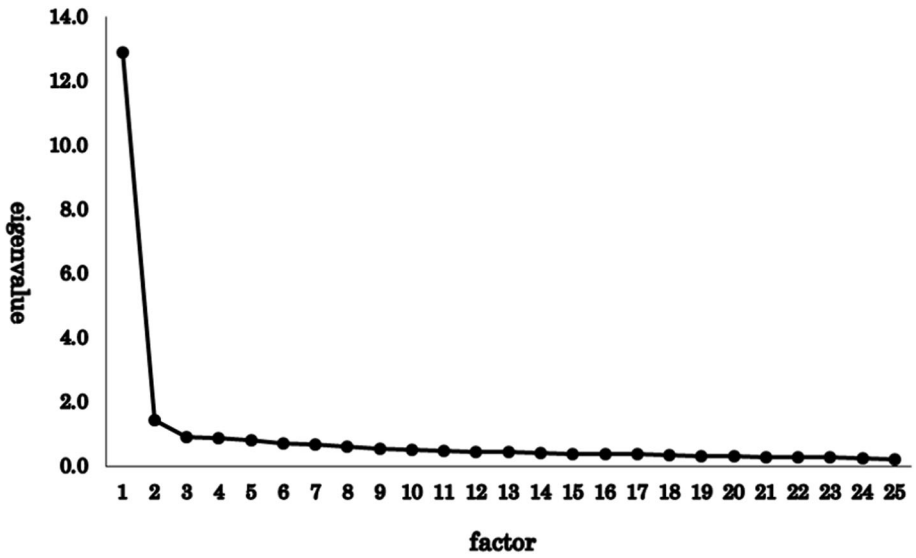


Fig. 2 Scree plot of the exploratory factor analysis for 25 items of THI. The eigenvalues for the factors were 12.88 (factor 1) and 1.45 (factor 2), explaining 57.3% of the variance

Table 3 Factor loadings of exploratory factor analysis (loadings above 0.40 are presented)

Item	Factor 1	Factor 2
1	0.72	
2		0.39
3	0.68	
4	0.74	
5	0.8	
6	0.74	
7	0.5	
8	0.69	
9	0.65	
10	0.82	
11	0.73	
12	0.83	
13	0.72	
14	0.76	
15	0.66	
16	0.73	
17	0.71	
18	0.7	
19	0.72	
20	0.76	
21	0.85	
22	0.62	
23	0.75	
24	0.47	
25	0.77	

These results indicated that only the total score, not any subscale, would be clinically useful.

Each item of the THI

Items being selected "yes" by more than 10% of the subjects were chosen for plotting a graph for each severity category of the THI (Fig. 3A, B, C, D). In the group of no handicap, there was no item whose proportion of being selected "yes" exceeded 10%. The item that was selected "yes" by the largest proportion in the group of no handicap was item 24: "Does your tinnitus get worse when you are under stress?," and the proportion was 9.2%.

In the group of mild handicap, 5 items were selected "yes" by more than 10% of the subjects. They were as follows:

(i) item 2: "Does the loudness of your tinnitus make it difficult for you to hear people?";
(ii) item 7: "Because of your tinnitus, do you have trouble falling to sleep at night?";
(iii) item 8: "Do you feel as though you cannot escape your tinnitus?";
(iv) item 19: "Do you feel that you have no control over your tinnitus?"; and
(v) item 24: "Does your tinnitus get worse when you are under stress?,"

These indicate that these symptoms are observed in relatively mild cases. Figure 3A shows the change of the proportion of being selected "yes" in each tinnitus severity for these five items. Difficulty in listening, as suggested from item 2, was relatively high in mild handicap

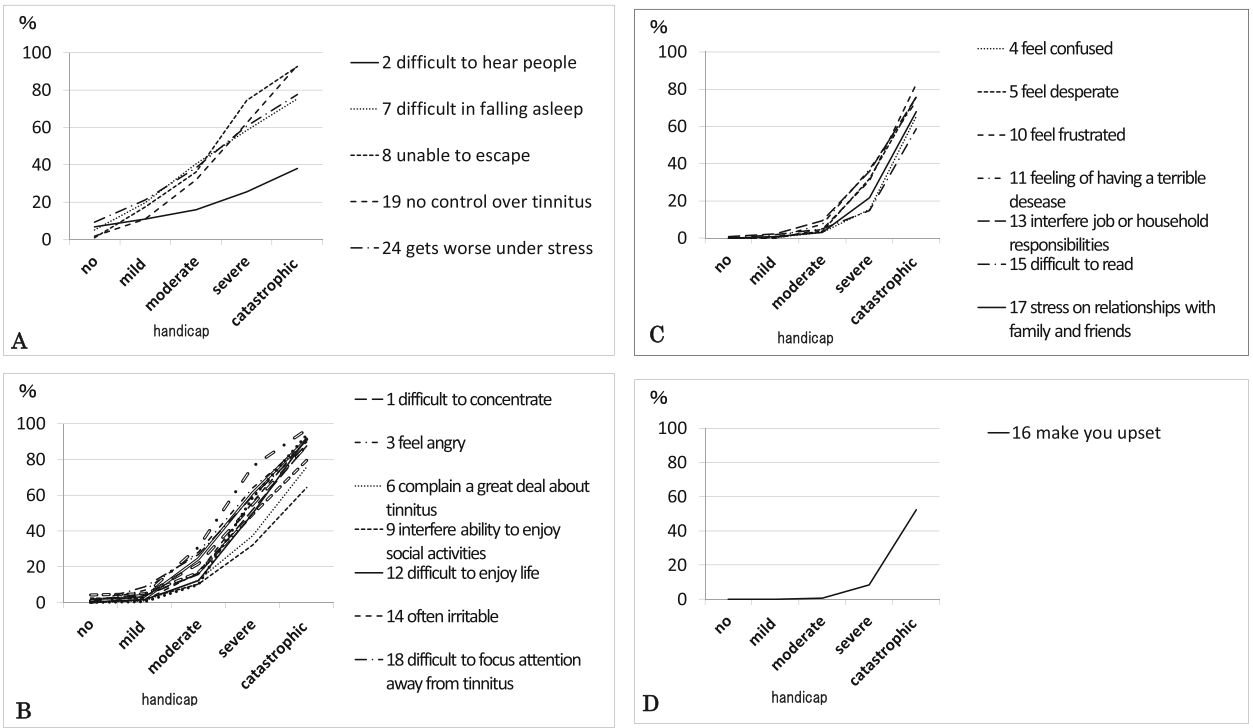


Fig. 3 Changes in the proportion of the patients who chose "yes" in each item. The graphs are divided according to the tinnitus handicap severity of which more than 10% of patients chose "yes". **A**: Five items were chosen as 'yes' by more than 10% of the 'mild handicap' patients. **B**: Twelve items were chosen as 'yes' by more than 10% of the 'moderate handicap' patients. **C**: Seven items were chosen as 'yes' by more than 10% of the 'severe handicap' patients. **D**: One item was chosen as 'yes' by more than 10% of the 'catastrophic handicap' patients

patients, whereas this symptom was relatively low in the catastrophic handicap patients.

In the group of moderate handicap, other 12 items were selected "yes" by more than 10% of the subjects. The items were as follows:

(i) item 1: "Because of your tinnitus, is it difficult for you to concentrate?";
(ii) item 3: "Does your tinnitus make you angry?";
(iii) item 6: "Do you complain a great deal about your tinnitus?";
(iv) item 9: "Does your tinnitus interfere with your ability to enjoy your social activities?";
(v) item 12: "Does your tinnitus make it difficult for you to enjoy life?";
(vi) item 14: "Because of your tinnitus, do you find that you are often irritable?";
(vii) item 18: "Do you find it difficult to focus your attention away from your tinnitus and on other things?";
(viii) item 20: "Because of your tinnitus, do you feel tired?";
(ix) item 21: "Because of your tinnitus, do you feel depressed?";
(x) item 22: "Does your tinnitus make you feel anxious?";
(xi) item 23: "Do you feel that you can no longer cope with your tinnitus?"; and
(xii) item 25: "Does your tinnitus make you feel insecure?."

Figure 3-B shows the change of the proportion of being selected "yes" in each tinnitus severity for these 12 items. These results suggested that in moderate handicap patients, the influence of tinnitus on social activities, daily life, and work has increased. It is also suggested that the proportion of admitting psychological discontent, irritation, fatigue, depression, and anxiety has increased.

In the group of severe handicap, almost all of the remaining items were selected "yes" by more than 10% of the subjects. They were as follows:

(i) item 4: "Does your tinnitus make you feel confused?";
(ii) item 5: "Because of your tinnitus, do you feel desperate?";
(iii) item 10: "Because of your tinnitus, do you feel frustrated?";

(iv) item 11: "Because of your tinnitus, do you feel that you have a terrible disease?";
(v) item 13: "Does your tinnitus interfere with your job or household responsibilities?";
(vi) item 15: "Because of your tinnitus, is it difficult for you to read?"; and
(vii) item 17: "Do you feel that your tinnitus problem has placed stress on your relationships with members of your family and friends?."

Figure 3-C shows the change of the proportion of being selected "yes" in each tinnitus severity for these 7 items. In the group of severe handicap, the proportion of "yes" exceeded 50% in 12 items such as not being able to concentrate (item 1), getting angry (item 3), sleep disorder (item 7), and depression (item 21). It is suggested that in severe handicap patients, many patients feel frustrated and feel that they have serious illness and human relationship failure in their daily life.

In the group of catastrophic handicap, the proportion of patients who chose "yes" newly exceeded 10% in only one item. That was as follows:

(i) item 16: "Does your tinnitus make you upset?".

There was no item whose proportion of being selected "yes" was 10% or less in the group of catastrophic handicap. The proportion of patients who select "yes" in item 16 rise drastically in the group of catastrophic handicap (52.2%) compared with in the group of severe handicap (8.3%). Figure 3-D shows the change in the proportion of being selected "yes" in each tinnitus severity for item 16. In the group of catastrophic handicap, the proportion of "yes" exceeded 50% in nearly all item. Only item 2: "Does the loudness of your tinnitus make it difficult for you to hear people?" showed a relatively low proportion, 38%.

Discussion

As a result of factor analysis in this study, it was considered that evaluation based on the total score was the most appropriate for the THI, as indicated by Baguley [14]. By examining individual items of the THI, it was suggested that the tendency of how tinnitus became severe could be understood.

We examined the proportion of patients with tinnitus who selected "yes" in each item of THI on their first visit to our hospital and found the tendency of symptoms which patients experienced as their tinnitus became severe. The symptoms of sleep disorders and difficulty in hearing were often seen in relatively mild cases. Their work, social lives, and daily lives were disturbed by tinnitus in moderate cases. In a psychological aspect, the patients were not able to enjoy their lives and tended to feel fatigue, depression, and anxiety. As tinnitus became severe, many patients felt despair, frustration, and severe illness, and human relations also interfered.

Sleep disorders, depression, anxiety have been reported to be related to tinnitus [13, 19, 20]. In this study, it became clear that patients with tinnitus are often bothered with sleep disorders, even if the tinnitus handicap is relatively mild. This suggests that management of sleep problems is important at the early stages of tinnitus. In our previous study, sleep disorder was observed in 70% of all patients with tinnitus [21]. In the previous study, there was no clear relationship between the presence or absence of sleep disorder and severity of tinnitus, which is consistent with the results obtained from each THI item in this study.

As tinnitus handicap becomes moderate, more patients suffer from depressive symptoms and anxiety symptoms. Although the THI can easily evaluate severity in a convenient manner, using only THI is insufficient to grasp details of comorbid symptoms such as depressive and anxiety symptoms. In the TRI, for evaluation of the status of patients and the therapeutic effect, some questionnaires are mentioned such as the BDI (Beck Depression Inventory [10]), the STAI (State-Trait Anxiety Inventory [12]), the PSQI (Pittsburgh Sleep Quality Index [22]) [2]. The recommendation degree of these questionnaires is C: might be of interest, but considering that the proportion of patients with depression and anxiety symptoms is higher in patients with tinnitus of moderate or severe handicap, these questionnaires should be useful.

In patients with severe and catastrophic handicap, there are many patients who feel despair, frustration, and severe illness, requiring both physical and mental care. These results suggest that in the treatments of patients with severe and catastrophic tinnitus handicap who show symptoms of despair, frustration, and severe illness, it is important to cooperate with psychiatrists or psychologists proactively.

Also, the percentage of patients who selected "yes" to item 2: "Does the loudness of your tinnitus make it difficult for you to hear people?" was more than 10% in mild handicap patients, whereas it did not exceed 40% in catastrophic handicap patients. In the catastrophic handicap patients, the proportion of patients who selected "yes" in this item 2 was overwhelmingly low compared with the other question items. As a treatment of tinnitus, the use of a hearing aid is recommended when hearing loss is accompanied, and its effectiveness has been reported [23, 24]. However, based on our results, the use of hearing aids alone might not be efficient to treat patients with catastrophic tinnitus, and hence psychiatric treatment should be considered.

Our results show that evaluation based on the total score is appropriate for THI, which differs from the report by Kleinstauber [15]. In Kleinstauber's study, patients with tinnitus were part of those participating in

their cognitive behavior therapy recruited through the Internet. The average total score of THI was lower than our research (41.3 in the Kleinstauber's study; and 54.2 in this study). In our study, the participants were the patients who visited us for treatment as chief complaints of tinnitus. As shown in the path diagram (Fig. 1), the correlation between factors is very strong, which is higher than that reported by Kleinstauber. By item, the path coefficients of item 8 (Do you feel as though you cannot escape your tinnitus?) and item 19 (Do you feel that you have no control over your tinnitus?) was significantly different from Kleinstauber's report. These items are included in the catastrophic subscale proposed by Newman, but in our study, these items were selected "yes" by relatively high proportion of patients with mild tinnitus handicap. Differences in these results may be related to different backgrounds including cultural difference.

In this study, we collected data from as many as 1332 patients. We found that symptoms such as sleep disorder and difficulty in hearing are relatively frequently seen in mild tinnitus handicap patients. When tinnitus handicap becomes moderate, the proportion of patients who feel tired, depression, anxiety, difficulty in enjoying life, and so on, increased. As tinnitus handicap becomes severe, many patients feel desperate, frustrated, severe illness, and human relations are disturbed. These changes are important in grasping the patient's condition and choosing treatment.

Conclusion

As a result of the factor analysis, the THI can be used as a unifactorial measurement of tinnitus handicap, and it is most appropriate to evaluate based on the total score.

By examining each item of the THI, we could observe which comorbid symptoms will frequently appear when tinnitus handicap becomes severe. It is necessary to choose treatment by considering that these comorbid symptoms emerge when tinnitus handicap becomes severe.

Abbreviations

AGFI: Adjusted Good of Fit Index; BDI: Beck Depression Inventory; CFI: Comparative Fit Index; GFI: Good of Fit Index; HADS: Hospital Anxiety and Depression Scale; MOS: Medical Outcomes Study; PSQI: Pittsburgh Sleep Quality Index; RMSEA: Root Mean Square Error of Approximation; SDS: Self-Rating Depression Scale; SF-36: 36-Item Short Form Health Survey; STAI: State-Trait Anxiety Inventory; THI: Tinnitus Handicap Inventory; TRI: Tinnitus Research Initiative

Acknowledgements

The authors would like to thank Enago (www.enago.jp) for the English language review.

Authors' contributions

All authors are involved in patient data collection. SW and NO analyzed and interpreted patient data. All authors have read and approved the manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was approved by the ethics committee of the Keio University School of Medicine (JPRN-UMIN000008901) and conducted in accordance with the Declaration of Helsinki. No animals are involved.

Consent for publication

Detail of this clinical research was displayed at a consultation room and oral consent was collected from all participants. We notified the research subjects, or made public information concerning the research including the purpose of collection and utilization of research information, and that there is an opportunity to refuse participation or remove their data from the study after commencement. This information was also documented in each patient's chart. For all participants who were under 20 years of age, we received consent from the parents or legal gardians. All patients consented to use of their data for future studies. Data were anonymized at the time of collection.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Otorhinolaryngology-Head and Neck Surgery, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan. [2]Department of Otorhinolaryngology, Tokyo Metropolitan Hospital Ohtsuka, 2-8-1 Minamiohtsuka, Toshima-ku, Tokyo 170-8476, Japan. [3]Department of Otorhinolaryngology, Saiseikai Utsunomiya Hospital, 911-1 Takebayashicho, Utsunomiya-shi, Tochigi 321-0974, Japan.

Received: 13 August 2019 Accepted: 12 February 2020
Published online: 07 March 2020

References

1. Newman CW, Jacobson GP, Spitzer JB. Development of the tinnitus handicap inventory. Arch Otolaryngol. 1996;122:143–8. https://doi.org/10.1001/archotol.1996.01890140029007.
2. Langguth B, Goodey R, Azevedo A, Cacace A, Crocetti A, et al. Consensus for tinnitus patient assessment and treatment outcome measurement: tinnitus research initiative meeting, Regensburg, July 2006. Prog Brain Res. 2007;166:525–36 https://doi.org/10.1016/S0079-6123(07)66050-6.
3. Monzani D, Genovese E, Marrara A, Gherpelli C, Pingani L, et al. Validity of the Italian adaptation of the tinnitus handicap inventory: focus on quality of life and psychological distress in tinnitus-sufferers. Acta Otorhinolaryngol Ital. 2008;28(3):126–34. 18646574.
4. Schmidt LP, Teixeira VN, Dall'Igna C, Dallagnol D, Smith MM. Brazilian Portuguese language version of the "tinnitus handicap inventory": validity and reproducibility. Braz J Otorhinolaryngol. 2006;72(6):808–10 https://doi.org/10.1016/S1808-8694(15)31048-X.
5. Meng Z, Zheng Y, Liu S, Wang K, Kong X, et al. Reliability and validity of the Chinese (mandarin) tinnitus handicap inventory. Clin Exp Otorhinolaryngol. 2012;5(1):10–6 https://doi.org/10.3342/ceo.2012.5.1.10.
6. Kam AC, Cheung AP, Chan PY, Leung EK, Wong TK, et al. Psychometric properties of the Chinese (Cantonese) tinnitus handicap inventory. Clin Otolaryngol. 2009;34(4):309–15 https://doi.org/10.1111/j.1749-4486.2009.01946.x.
7. Shinden S, Ogawa K, Inoue Y, Tazoe M, Asano K. Tinnitus annoyance and difficulty in activities of daily life evaluated by the tinnitus handicap inventory (THI). Audiology Japan. 2002;45:685–91 https://doi.org/10.4295/audiology.45.685.
8. McHorney CA, Ware JE, Raczek AE. The MOS 36-item short form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care. 1993;31:247–63 https://www.jstor.org/stable/3765819.
9. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67:361–70 https://doi.org/10.1111/j.1600-0447.1983.tb09716.x.

10. Beck AT, Steer RA. International consistencies of the original and revised beck depression inventory. J Clin Psychol. 1984;40:1365–7 https://doi.org/10.1002/1097-4679(198411)40:63.0.CO;2-D.
11. Zung WW. A self-rating depression scale. Arch Gen Psychiatry. 1965;112:63–70 https://doi.org/10.1001/archpsyc.1965.01720310065008.
12. Spielberger CD. State-trait anxiety inventory: a comprehensive bibliography. Palo Alto: Consulting Psychologists Press; 1984.
13. Oishi N, Shinden S, Kanzaki S, Saito H, Inoue Y, et al. Influence of depressive symptoms, state anxiety, and pure-tone thresholds on the tinnitus handicap inventory in Japan. Int J Audiol. 2011;50:491–5 https://doi.org/10.3109/14992027.2011.560904.
14. Baguley DM, Andersson G. Factor analysis of the tinnitus handicap inventory. Am J Audiol. 2003;12:31–4 https://doi.org/10.1044/1059-0889(2003/007).
15. Kleinstauber M, Frank I, Weise C. A confirmatory factor analytic validation of the tinnitus handicap inventory. J Psychosom Res. 2015;78:277–84 https://doi.org/10.1016/j.jpsychores.2014.12.001.
16. McCombe A, Baguley D, Coles R, McKenna L, McKinney C, et al. Guidelines for the grading of tinnitus severity: the results of a working group commissioned by the British Association of Otolaryngologists, head and neck surgeons, 1999. Clin Otolaryngol Allied Sci. 2001;26:388–93 https://doi.org/10.1046/j.1365-2273.2001.00490.x.
17. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model. 1999;6:1–55 https://doi.org/10.1080/10705519909540118.
18. Haebara T. On the relevance of goodness of model fit to the purpose of research – should we keep the number of indicator variables small? Japanese J Behaviormetrics. 2002;29:160–6 https://doi.org/10.2333/jbhmk.29.160.
19. Cronlein T, Langguth B, Geisler P, Hajak G. Tinnitus and insomnia. Prog Brain Res. 2007;166:227–33 https://doi.org/10.1016/S0079-6123(07)66021-X.
20. Zöger S, Svedlund J, Holgers KM. Relationship between tinnitus severity and psychiatric disorders. Psychosomatics. 2006;47:282–8 https://doi.org/10.1176/appi.psy.47.4.282.
21. Wakabayashi S, Saito H, Oishi N, et al. Effects of tinnitus treatments on sleep disorders in patients with tinnitus. Int J Audiol. 2018;57:110–4 https://doi.org/10.1080/14992027.2017.1374565.
22. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28:193–213 https://doi.org/10.1016/0165-1781(89)90047-4.
23. Tunkel DE, Bauer CA, Sun GH, et al. Clinical practice guideline: Tinnitus. Otolaryngol Head Neck Surg. 2014;151:S1–S40 https://doi.org/10.1177/0194599814545325.
24. Shekhawat GS, Searchfield GD, Stinear CM. Role of hearing AIDS in tinnitus intervention: a scoping review. J Am Acad Audiol. 2013;24:747–62 https://doi.org/10.3766/jaaa.24.8.11.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

Comments Submitted by Interested Parties on a Pending AMA Guides® Editorial Change Proposal

Instructions: By submitting comments on this form regarding an Editorial Change Proposal, I attest that I have read AMA Guides® Editorial Change Proposal and Submission Requirements and will use them as the primary points of consideration when submitting the Comment Form. As an interested party, I understand that my comments are limited to the original editorial change proposal.

Name or Topic of Proposal: Chapter 11 Ear, Nose, Throat, and Related Structures

Individual or Organization Submitting Comments: American Speech-Language Hearing Association

Date: 7/15/22

I. General Criteria for Guides Editorial Changes

- The proposed change is carefully drafted and conforms to the prevailing style of the *AMA Guides 6th Edition*;
- The terminology and the analytical frameworks used in the proposal are consistent with the World Health Organization's International Classification of Functioning, Disability, and Health (ICF);
- The structure and content of the proposed editorial change ensures that impairment ratings are transparent, clearly stated, and reproducible, to insure physician interrater reliability;
- The clinical soundness of the proposed editorial change is demonstrated with the best available evidence except in the case of minor editorial changes.

1. **Does the requested procedure meet the AMA Guides® Editorial Change Proposal and Submission Requirements?**

   ☒Yes

   ☐No

   If No, please explain. (1500 character limit)

2. **Does the submitted literature adequately support the Editorial Change Proposal?**

   ☒Yes

   ☐No

   ☐N/A

   If No, please explain. (1500 character limit)

3. **Are you aware of contradictory literature related to the Editorial Change Proposal?**

   ☒Yes

   ☐No

   ☐N/A

   If Yes, please include a maximum of five (5) articles when submitting this form. Articles in full text or PDF formats are required. Citations only will not be considered.

The following articles support the change from "no contraindications" to "few contraindications."

Jacobson, G. P., E. G. Piker, C. Do, D. J. McCaslin, and L. Hood. 2012. "Suppression of the Vestibulo-Ocular Reflex Using Visual and Nonvisual Stimuli." American Journal of Audiology 21 (2): 226–231. doi:10.1044/1059-0889(2012/12-0021).

Doettl SM, Easterday MK, Plyler PN, Behn LL, Poget AS. Mental tasking and rotary Chair-Induced vestibular nystagmus utilizing Video-Oculography. Int J Audiol. 2020 May;59(5):360-366. doi: 10.1080/14992027.2019.1706768. Epub 2019 Dec 26. PMID: 31876202.

Makowiec K, Smith K, Deeb A, Bennett E, Sis J. Influence of Tasking During Vestibular Testing. Am J Audiol. 2021 Sep 10;30(3):755-760. doi: 10.1044/2021_AJA-20-00227. Epub 2021 Aug 20. PMID: 34415794.

McNerney KM, Coad ML, Burkard R. The Influence of Caffeine on Rotary Chair and Oculomotor Testing. J Am Acad Audiol. 2018 Jul/Aug;29(7):587-595. doi: 10.3766/jaaa.16118. PMID: 29988007.

Discussion on efficient use/when to incorporate testing: Zuniga SA, Adams ME. Efficient Use of Vestibular Testing. Otolaryngol Clin North Am. 2021 Oct;54(5):875-891. doi: 10.1016/j.otc.2021.05.011. Epub 2021 Jul 20. PMID: 34294436; PMCID: PMC8453116.

4. **Do you support this Editorial Change Proposal?**

   ☒Yes
   ☐No

   If No, please provide the rationale for lack of support, citing the specific criteria not met shown at the top of this form. (1500 character limit)

5. **Does the Editorial Change Proposal have any impact on other *AMA Guides* content that may not have been recognized or considered, or conflict with other precedents in the AMA Guides that might affect usage?**

   ☐Yes
   ☒No

   If Yes, please explain. (1500 character limit)

Please provide additional commentary related to the editorial change proposal.

The American Speech-Language-Hearing Association (ASHA) is the national professional, scientific, and credentialing association for 223,000 members and affiliates who are audiologists; speech-language pathologists; speech, language, and hearing scientists; audiology and speech-language pathology support personnel; and students. ASHA generally supports the revisions made by AAO-HNS and offers the following feedback and suggestions to the proposed changes.

ASHA recommends referring to food and liquid textures in accordance with the International Dysphagia Diet Standardization Initiative (IDDSI). The framework is designed to avoid the confusion created by variable terminology and definitions to describe modified diets. The effort is expected to improve the safety and care for all individuals with dysphagia, across all cultures. In the future, this framework may be beneficial to refine levels of impairment related to dietary restrictions in table 11-7. Please see citation and attached article for additional information.

Cichero, J.A.Y., Lam, P., Steele, C.M. et al. Development of International Terminology and Definitions for Texture-Modified Foods and Thickened Fluids Used in Dysphagia Management: The IDDSI Framework. Dysphagia 32, 293–314 (2017). https://doi.org/10.1007/s00455-016-9758-y.

ASHA also recommends amending the sentence “Rotary chair testing has no contraindications such as neck trauma, which is not the case with ENG or VNG” to read “Rotary chair testing has few contraindications, which is not the case with ENG or VNG.” See citations above and articles attached to support this wording change.

Additionally, ASHA has collected comments and suggestions from an audiologist subject matter expert with an expertise in vestibular issues. ASHA submits the following information for panel consideration when refining AMA guides pertaining to vestibular assessment and treatment.

Equilibrium:

- VNG and Rotational Chair be in a subcategory entitled Vestibular Diagnostic Testing (VNG is a large part of this) and vestibular-evoked myogenic potentials should also be included in this section.
- Very few contemporary clinics employ ENG anymore—the vast majority use VNG; however, this section focuses heavily on ENG. Pediatric centers rely on ENG more than adult centers. It is important to note that the noise floor in VNG is much lower than ENG. That is, VNG is more sensitive to eye movement disorders and facilitates the recording on video for formal documentation in the medical record.
- ENG is not only a technique for recording spontaneous and induced nystagmus; it also records any ocular motor (peripheral or central) abnormality.
- Consider revising the statement on avoidance of drugs. There is almost no published double-blinded evidence that indicates that drugs have an impact on VNG findings. In fact, many contemporary clinics are allowing for patients to take all of their medications before testing. Medications that are prescribed for sleep can cause issues with vigilance but very little published data exists on the direct effect.
- VNG is not a test done in phases, it is a technique. Tests of ocular motor function, positioning testing, and peripheral vestibular assessment (caloric) are all tests that use this technique. The sentence that starts with “These include calibration…” should clarify that calibration does not assess cerebellar function.
- Dix-Hallpike testing should be replaced with “positioning testing”. Positioning testing includes the Dix-Hallpike technique (posterior canal), roll test (lateral canal), and deep supine head hanging (anterior canal). It is different than positional testing that is listed.
- Consider changing “Interpretation is complex and difficult” to “The ability to accurately interpret VNG data requires a thorough working knowledge of vestibular and ocular motor system physiology and disorders.”

Rotational Chair:

- Consider rewriting the first sentence to read: Rotational testing evaluates the horizontal canal-ocular reflex (VOR) and the central vestibular system (i.e. vestibular nuclei, cerebellum/neural integrator).
- Consider replacing “more sophisticated information” with “suspected and a full accounting of the vestibular responses across the frequency range or VOR timing characteristics is needed.”
- Consider changing “Unlike caloric testing, rotary chair testing stimulates both ears” to “Rotary chair assesses the integrity of both the right and left horizontal semicircular canals simultaneously.”
- The frequency of the caloric is thought to be 0.003Hz.

Thank you for the opportunity to offer suggestions on the proposed changes to the AMA Guides. If you or your staff have any questions, please contact Rebecca Bowen M.A., CCC-SLP, ASHA’s director for health care policy, value, and innovation, at rbowen@asha.org.

Dysphagia (2017) 32:293–314
DOI 10.1007/s00455-016-9758-y

ORIGINAL ARTICLE

Development of International Terminology and Definitions for Texture-Modified Foods and Thickened Fluids Used in Dysphagia Management: The IDDSI Framework

Julie A. Y. Cichero[1,2] · Peter Lam[1,3,4] · Catriona M. Steele[1,5,6] · Ben Hanson[1,7] · Jianshe Chen[1,8] · Roberto O. Dantas[1,9] · Janice Duivestein[1,10,11] · Jun Kayashita[1,12] · Caroline Lecko[1,13] · Joseph Murray[1,14] · Mershen Pillay[1,15,16] · Luis Riquelme[1,17,18] · Soenke Stanschus[1,19]

Received: 2 August 2016 / Accepted: 8 November 2016 / Published online: 2 December 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Dysphagia is estimated to affect ~8% of the world's population (~590 million people). Texture-modified foods and thickened drinks are commonly used to reduce the risks of choking and aspiration. The International Dysphagia Diet Standardisation Initiative (IDDSI) was founded with the goal of developing globally standardized terminology and definitions for texture-modified foods and liquids applicable to individuals with dysphagia of all ages, in all care settings, and all cultures. A multiprofessional volunteer committee developed a dysphagia diet framework through systematic review and stakeholder consultation. First, a survey of existing national terminologies and current practice was conducted, receiving 2050 responses from 33 countries. Respondents included individuals with dysphagia; their caregivers; organizations supporting individuals with dysphagia; healthcare professionals; food service providers; researchers; and industry. The results revealed common use of 3–4 levels of food texture (54 different names) and ≥3 levels of liquid thickness (27 different names). Substantial support was expressed for international standardization. Next, a systematic review regarding the impact of food texture and

Electronic supplementary material The online version of this article (doi:10.1007/s00455-016-9758-y) contains supplementary material, which is available to authorized users.

✉ Julie A. Y. Cichero
j.cichero@uq.edu.au

Peter Lam
peter.lam@iddsi.org

Catriona M. Steele
catriona.steele@iddsi.org

Ben Hanson
ben.hanson@ucl.ac.uk

Jianshe Chen
jschen@zjgsu.edu.cn

Roberto O. Dantas
rodantas@fmrp.usp.br

Janice Duivestein
jduivestein@cw.bc.ca

Jun Kayashita
kayashita@pu-hiroshima.ac.jp

Caroline Lecko
caroline.lecko@nhs.net

Joseph Murray
joe.murray@va.gov

Mershen Pillay
pillaym1@uksn.ac.za

Luis Riquelme
luisslp@aol.com

Soenke Stanschus
soenke.stanschus@gmx.de

1 International Dysphagia Diet Standardisation Initiative (IDDSI) Working Committee, Brisbane, QLD, Australia

2 School of Pharmacy, Pharmacy Australia Centre of Excellence (PACE), The University of Queensland, 20 Cornwall St, Brisbane, QLD 4102, Australia

3 Faculty of Land and Food Systems, University of British Columbia, Vancouver, BC, Canada

4 Peter Lam Consulting, Vancouver, BC, Canada

5 Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

6 Rehabilitation Sciences Institute, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

liquid consistency on swallowing was completed. A meeting was then convened to review data from previous phases, and develop a draft framework. A further international stakeholder survey sought feedback to guide framework refinement; 3190 responses were received from 57 countries. The IDDSI Framework (released in November, 2015) involves a continuum of 8 levels (0–7) identified by numbers, text labels, color codes, definitions, and measurement methods. The IDDSI Framework is recommended for implementation throughout the world.

Keywords Deglutition · Deglutition disorders · Swallowing · Dysphagia diet · Texture-modified food · Thickened fluid · Food and fluid standards

Introduction

Standardized terminology exists to reduce misunderstanding and ambiguity and to improve communication efficiency [1]. The field of dysphagia has benefited from standardized scales in outcome measurement that allow clinicians to reliably document change in status during management. Examples of dysphagia-specific standardized scales include the Penetration–Aspiration Scale [2]; the Swal-QOL and Swal-CARE [3]; the Dysphagia Outcome Severity Scale [4]; and the Functional Oral Intake Scale [5]. However, despite the fact that texture modification is one of the most common intervention approaches for dysphagia [6], the descriptions of thickened drinks and texture-modified foods vary throughout the world, including within countries, and even across hospitals located with close geographic proximity [7]. We hypothesize that a standardized framework for dysphagia diets could offer benefits including but not limited to improved patient safety; improved communication within and between health professionals, healthcare providers and patients; increased visibility of professional interventions; and greater opportunities to collect and evaluate treatment outcomes [7–11]). Of these, the two most compelling reasons to pursue standardization of dysphagia diets are to promote patient safety and to facilitate evolution of the field to deliver better treatment outcomes.

Much like dose-driven medication prescriptions for different severities of medical conditions, individuals with dysphagia are assessed and prescribed graded food textures and drink thicknesses that are commensurate with their physical and cognitive abilities. Also similar to medication adverse events, inconsistencies and errors in labeling of texture-modified foods have unfortunately resulted in deaths attributed to the delivery of inappropriate food textures to patients with dysphagia [7, 12–14]. In recent years, a number of countries have worked hard to create standards for texture-modified foods and thickened drinks with the goal of improving patient safety and care [7, 15–20]. However, with an increase in mobility of the global community and access to information via the internet, the plethora of dysphagia diet terminology, labels, numbers, and levels of food texture and thickened drinks has only led to greater opportunities for confusion. Furthermore, a proliferation of companies producing thickeners and ready-to-use products means that patients and their caregivers cannot assume similarity in thickness across brands. This scenario is in contrast to expectations of bioequivalence in medicine between name brand medications and generic versions, which must have the same active ingredient, strength, dosage form, and route of administration as the brand name product [21]. There is no such regulation of dysphagia products to ensure 'like-for-like' in terms of commercially prepared fluid thickness levels or food texture modification. To be fair, manufacturers cannot be held accountable for producing products that conform to standards until such standards have been established, and this requires stakeholder agreement, clarity regarding labels, detailed definitions, and testing methods to demonstrate conformance to desired properties.

A lack of standardized nomenclature regarding food texture and drink thickness is a major barrier to research in the dysphagia field. Without clear definitions, we cannot presume that the outcomes of research conducted on the efficacy of prescribing 'nectar-thick' drinks for patients

[7] Department of Mechanical Engineering, University College London, London, UK

[8] Zhejiang Gongshang University, Hangzhou, China

[9] Medical School of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil

[10] Access Community Therapists, Vancouver, BC, Canada

[11] Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[12] Department of Health Sciences, Prefectural University of Hiroshima, Hiroshima, Japan

[13] National Health Service Improvement, London, UK

[14] Ann Arbor Veterans Affairs, Ann Arbor, MI, USA

[15] Speech Pathology, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban, South Africa

[16] Manchester Metropolitan University, Manchester, UK

[17] Department of Speech-Language Pathology, New York Medical College, Valhalla, NY, USA

[18] Barrique Speech-Language Pathology at Center for Swallowing & Speech-Language Pathology, New York Methodist Hospital, Brooklyn, NY, USA

[19] Swallowing and Speech Pathology, Hospital zum Heiligen Geist, Kempen, Germany

with Parkinson's disease in one country, for example, can be generalized around the world. Use of a term such as 'nectar-thick' in research conducted in the USA (e.g., [22, 23]) may not translate to products or liquid consistencies used in other countries, such as the United Kingdom, Japan, or Australia, regardless of the fact that each has a set of National Descriptors. Without agreement on a single standardized terminology, clinical research and development of therapies is impeded.

The International Dysphagia Diet Standardisation Initiative Inc. (IDDSI) was founded in 2012 by a multi-professional international group of volunteers. IDDSI is an independent, not-for profit entity (Incorporation Number IA40577). The ultimate objective of the initiative is to pursue a patient-safety-oriented innovation in practice, based on consideration of research evidence, current practice, and stakeholder feedback. There was no plan for the initiative to address the nutritional adequacy or the patient acceptability of texture-modified foods or thickened fluids.

The aims of the initiative that are discussed in this manuscript were to

1. determine the number of food texture and drink thickness levels for international standardized use (adult and pediatric);
2. develop culturally sensitive standard names/identifiers for each food and drink level;
3. develop detailed definitions for each level of food texture and drink thickness;
4. develop user-friendly, inexpensive, easily accessible measurement methods for determining classification of food textures and thickened drinks;
5. seek input and consensus from international stakeholders; and
6. publish and widely communicate the international standards.

The process used to develop the framework follows the key elements of evidence-based practice guideline development including those recommended by the National Health and Medical Research Council of Australia (NHMRC), the National Institute of Health and Clinical Excellence in the UK (NICE), the New Zealand Guidelines Group (NZGG), the Scottish Intercollegiate Guideline Network (SIGN), the Council of Europe, and the World Health Organization (WHO) [24]. These key elements include: establishment of a multidisciplinary guideline development group; involvement of consumers; clear identification of clinical issues; systematic review and appraisal of quality literature; a process for drafting the recommendation of the multidisciplinary group; and consultation with others beyond the multidisciplinary framework development group [24, 25].

Methods

Strategy and Incorporation

An inaugural multi-professional expert panel meeting was held in Toronto, Canada (2012) to discuss international standardization of terminology and definitions for texture-modified foods and drinks. A snowball sampling methodology was used to populate the expert panel following initial recruitment of two members with experience in the development of national terminologies [7, 19] and one who had commenced but not completed national terminology development in Canada (PL). Remaining members were invited to join the panel based on their previous work with national guideline development [16], their representation of key stakeholder groups, and their ability to contribute international perspectives. In 2013, IDDSI was incorporated as an independent, not-for-profit association operating under the regulatory guidelines of its registration in Australia. Two people volunteered to be co-chairs, with ten others agreeing to serve as members of the IDDSI Foundation Committee. All positions were voluntary. The committee, representing ten countries, was composed of experts from the fields of Nutrition and Dietetics, Food Service and Catering, Speech Pathology, Occupational Therapy, Physiotherapy, Gastroenterology, Nursing, Mechanical Engineering, and Food Science. The group counted among its members published scientists, journal editors, representatives from international organizations such as the Patient Safety - Nursing Directorate, National Health Service (NHS) England, and internationally recognized dysphagia clinicians and researchers. The committee met by teleconference on a monthly basis, with two in-person meetings over the project period (2013–2015). Sponsors were approached for financial support to cover costs associated with administration, research, and data analysis (e.g., research assistant support for the systematic review). At no time have sponsors been involved with the design or development of the IDDSI framework; rather, IDDSI sponsors have been briefed about IDDSI progress using time-zone sensitive teleconferences at key milestones over the course of the IDDSI project. Professional associations and organizations were also contacted to alert them to the IDDSI project and invite their participation and support.

A dedicated website was developed to provide an internationally accessible repository for information and a way for interested individuals or groups to contact IDDSI (www.iddsi.org). The IDDSI project plan, committee member profiles, lists of supporting organizations, and sponsors can be found on the website. A multi-phased work plan was approved by the committee with the goal of

bringing forth a framework between 2013 and 2015. Each of these phases is summarized in Fig. 1.

Ethical Considerations

The IDDSI committee considered ethical issues associated with the collection of survey data in different phases of the project. It was agreed that participation in IDDSI surveys involved minimal risk and was entirely voluntary. The committee agreed that the purpose of each survey and the overall project would be communicated at the time of invitations to participate. The introductory text of IDDSI surveys stated clearly that information gathered from individuals or organizations would remain non-identifying in all reports arising from the project. Participants were also free to withhold responses to survey questions at any stage without penalty. Several key stakeholder groups were identified and attempts were made to disseminate invitations regarding survey opportunities to all of these groups. To avoid commercial conflicts of interest, it was agreed that industry sponsors would not be involved with any aspects of IDDSI sub-project design, conduct, writing, or interpretation of results.

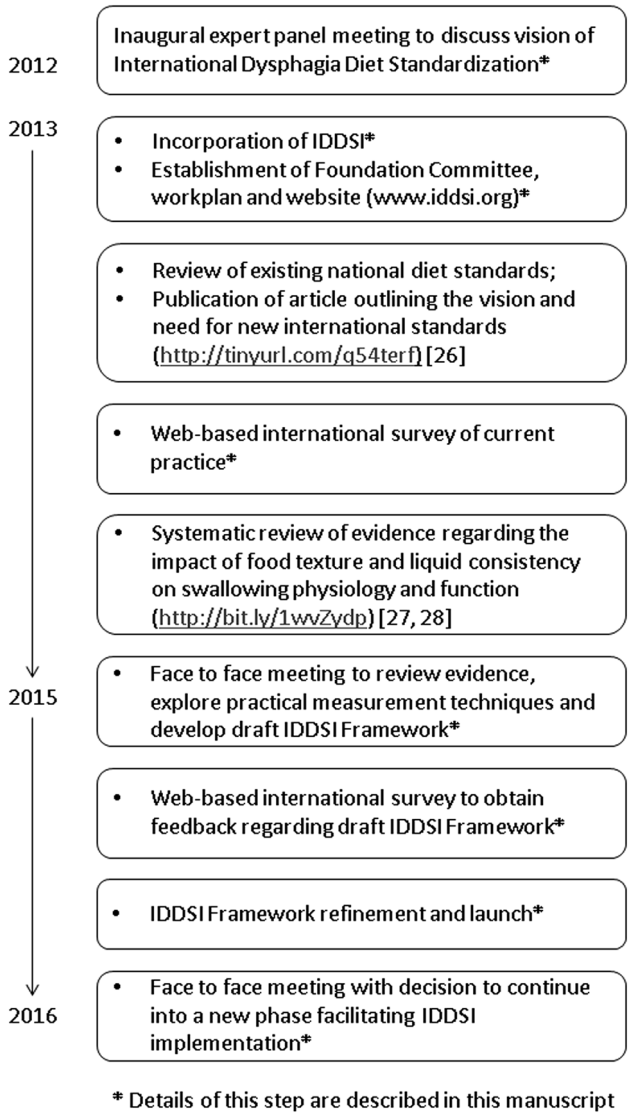


Fig. 1 Timeline of the international dysphagia diet standardisation initiative

Review of Existing National Terminologies

A review of existing national terminologies was conducted and published in 2013. Further details can be found in the open access journal publication [26].

Survey 1 (International Current Practice)

In 2013, a set of five stakeholder-specific surveys was developed to gather information regarding the current use of standardized dysphagia diet terminology or other terms used, any testing done prior to serving to ensure correct consistency/thickness and appropriate texture, use of schemes to differentiate levels (e.g., colors, shapes), and comments or recommendations for the development of an international standardized framework. Each survey was tailored to one of five stakeholder groups: (a) individuals with dysphagia, their caregivers, or organizations providing support to people with dysphagia; (b) healthcare professionals and food service professionals; (c) dysphagia research scholars; d) industry representatives from companies manufacturing texture-modified foods; and (e) industry representatives from companies manufacturing thickeners or thickened drinks for people with dysphagia. English language style and complexity of the surveys was adjusted to be appropriate for each group. There were commonalities in some survey questions, while other elements were specific to stakeholder experiences (see Table 1).

Surveys comprised forced choice and free-text response formats. An information sheet about the survey and invitation to participate were translated by native speaker volunteers into 10 languages other than English. Surveys were launched via the IDDSI website to individuals who had signed up to receive information about the initiative. In addition, 45 national healthcare professional associations and three dysphagia-specific associations were emailed information about the survey and asked to forward notices to their membership with embedded web links to facilitate ease of survey access. Invitations to complete the survey were also announced at international conferences. Survey responses were collected from October 2013 to November 2014 using Surveymonkey™. Upon closure of the surveys, the response data were transferred to an independent research group for analysis (Australian Survey Research

Table 1 Questions included in the current international practice survey by stakeholder group

Question	Response options	Patients and their carers or support organizations	Healthcare professionals and food service staff	Dysphagia research scholars	Industry representatives
What country are you located in?	Free text	X	X	X	X
How would you describe yourself?	An adult with dysphagia	X			
	I care for an adult with dysphagia	X			
	I care for a child with dysphagia	X			
	I am a member of an organization that specifically supports people with dysphagia	X			
	I am a member of an organization that supports children with dysphagia	X			
	Healthcare professional (specify)		X		
	Food service or catering professional		X		
	Representative of company manufacturing thickeners, thickened liquids, or barium				X
	Representative of company manufacturing texture-modified foods				X
What is your work setting?	Hospital/acute care		X	X	
	Rehab hospital		X		
	Long-term care/Home for the aged		X		
	Community care		X		
	Outpatient clinic		X		
	University			X	
	Government or private laboratory			X	
Who are your patients/clients?	Adults		X		
	Children		X		
	Mixed age group		X		
What type(s) of dysphagia research do you conduct?	Clinical research into oropharyngeal dysphagia			X	
	Clinical research into esophageal dysphagia			X	
	Fundamental swallowing research			X	
	Product development			X	
	Other (specify)				
Do you/the person you care for/your patients use	Texture-modified foods	X	X		
	Commercially prepared texture-modified foods	X	X		
	Thickened liquids	X	X		
	Commercially prepared thickened liquids	X	X		
	A combination of the above	X	X		
Do you use standard terminology and guidelines for texture-modified foods and liquid thicknesses?	Yes/No/I don't know with comments option	X	X		

Table 1 continued

Question	Response options	Patients and their carers or support organizations	Healthcare professionals and food service staff	Dysphagia research scholars	Industry representatives
Do you produce thickened liquids or products that will thicken liquids, for people with swallowing difficulties or those who are nutritionally compromised?	Ready-to-use drinks				X
	Powder thickeners				X
	Liquid thickeners				X
	Other (specify)				X
Do you produce thickened liquids, thickening agents, or barium products specifically for people with dysphagia?	Yes/No with comments option				X
Do you produce foods for people with chewing and swallowing difficulties or those who are nutritionally compromised?	Ready-to-eat frozen meals				X
	Ready-to-eat packaged foods (tins, pouches)				X
	Foods that need reconstituting				X
	Other (specify)				X
Do you produce texture-modified foods specifically for people with dysphagia?	Yes/No with comments option				X
Which terms do you use for texture-modified foods and liquid thicknesses	Terms for texture-modified foods (from least to most modified, separated by a comma)	X	X	X	X
	Terms for thickened liquids (from least to most modified, separated by a comma)	X	X	X	X
Where did you source the terms you use for thickened liquids and texture-modified foods?	National standards			X	X
	From the literature			X	X
	Self-developed			X	X
	Hospital or work facility			X	X
	Other (specify)			X	X
Do you test the consistency of foods or liquids before eating or serving?	Yes/No with comments option	X	X		
How do you determine whether liquids are of the correct thickness?	Free text			X	
How do you determine if foods are of the correct texture?	Free text			X	
Do you formally test thickened liquids or texture-modified foods as part of your research?	Liquids: always/usually/rarely/never with comments option			X	
	Food: always/usually/rarely/never with comments option			X	
	Barium: always/usually/rarely/never with comments option			X	
For thickened liquids (including barium), do you record information about:	Viscosity: always/usually/rarely/never with comments option			X	X

Table 1 continued

Question	Response options	Patients and their carers or support organizations	Healthcare professionals and food service staff	Dysphagia research scholars	Industry representatives
	Density: always/usually/rarely/never with comments option			X	X
	Yield stress: always/usually/rarely/never with comments option			X	X
If you formally evaluate thickened liquids or texture-modified foods, what measurement device(s) do you use? (For industry respondents, target measurement details were requested for each method)	Visual inspection				X
	Line spread test			X	X
	Bostwick consistometer			X	X
	Brookfield viscometer			X	X
	Cone and plate or parallel plate rheometer			X	X
	Other rheometer			X	X
	Food texture analyzer			X	X
	Sieve				X
	Image analysis				X
	Other (specify)			X	X
For texture-modified foods, do you record information about	Particle size: always/usually/rarely/never with comments option			X	X
	Cohesiveness: always/usually/rarely/never with comments option			X	X
	Adhesiveness: always/usually/rarely/never with comments option			X	X
	Firmness: always/usually/rarely/never with comments option			X	X
	Springiness: always/usually/rarely/never with comments option			X	X
	Brittleness or fracturability: always/usually/rarely/never with comments option			X	X
	Hardness: always/usually/rarely/never with comments option			X	X
	Yield stress: always/usually/rarely/never with comments option			X	X
	Other (specify)			X	X
Do you use a scheme (colors, numbers, shapes, etc.) to differentiate/communicate the different food textures and liquid thicknesses?	Yes/No with comments option	X	X		X
What problems, if any, are there with the terminology or definitions you currently use for thickened liquids or texture-modified foods?				X	X

Table 1 continued

Question	Response options	Patients and their carers or support organizations	Healthcare professionals and food service staff	Dysphagia research scholars	Industry representatives
Where do you distribute your products?	Locally/regionally/nationally/internationally with product type identified				X
Are there any other comments or recommendations you would like to make for the International Dysphagia Diet Standardisation Initiative?	Free text	X	X	X	X
Additional optional information	Name and address for future contact	X	X	X	X

Group) in order that the IDDSI committee did not have any opportunity to inadvertently bias the results.

Systematic Review of the Literature

A systematic review of the literature regarding the influence of food texture and liquid consistency on swallowing physiology was conducted in 2014, with the results published in 2015. The key findings from the systematic review showed that there is evidence that thicker liquids not only reduce the risk of penetration–aspiration, but also increase the risk of post-swallow residue in the pharynx. Further, the existing literature is insufficient to support the delineation of specific viscosity boundaries or other quantifiable material properties related to clinical outcomes. With regards to food texture used in dysphagia management, the systematic review determined that the best available evidence for selecting optimal food consistency comes from careful exploration of tolerance for different foods as part of a comprehensive swallowing assessment. The systematic review also demonstrated evidence that solid food and thick consistencies require greater effort in oral processing and swallowing. Note that terms related to choking, airway obstruction, or asphyxiation were not included in the search strategy for the systematic review. Further details can be found in the open access journal publication [27, 28].

Draft Framework Development

With information gathered from (a) existing national dysphagia diet terminology from around the world [26]; (b) the current practice international stakeholder surveys; and (c) the systematic review [27, 28], the IDDSI committee gathered in Vancouver, Canada in January, 2015 for a 2 ½ day in-person expert panel meeting to develop a draft international framework. Committee members from Australia, Canada, Germany, Japan, and the United Kingdom were able to attend in person, covering areas of expertise in nutrition and dietetics, food service and catering, speech pathology, occupational therapy, physiotherapy, food science, mechanical engineering, research, and both adult and pediatric clinical dysphagia services. Input from absent committee members on key questions was obtained via e-mail and telephone both during the meeting and over the following months.

The objectives of the expert panel meeting were to determine

(a) the number of levels of texture-modified foods for inclusion in a new standardized international dysphagia diet framework;
(b) the number of levels for thin and thickened drinks for inclusion in a new standardized international dysphagia diet framework;
(c) English language labels for texture-modified foods at each level;
(d) English language labels for thickened drinks at each level;
(e) a numbering system;
(f) whether to use a color scheme;
(g) graphical representations to capture the framework;
(h) detailed definitions and descriptions of the texture or flow characteristics of food and drink items included at each level; and
(i) reproducible testing methods to enable end users to assign foods and drinks to the different levels.

A group nomination process was used to achieve decisions for objectives a) to g). After discussion of the available evidence (both from the scientific literature and collected through the current international practice survey), motions were put forward and committee members indicated their agreement or dissent through a blinded ballot process. Unanimous voting resulted in adoption of that

particular motion. Less than unanimous voting resulted in further rounds of discussion and further blinded voting until unanimous consensus was reached. There were only two occasions where a second round of voting was required.

Classification of levels and exploration of measurement methods

Based on consideration of the scientific and survey evidence, the IDDSI Committee achieved consensus that a new framework should include 5 levels of drink thickness (thin plus 4 levels of thickness) and 5 levels of food texture (regular plus 4 levels of modification). The next aim was to define and describe the specific texture/flow characteristics for each level.

Liquids

Thirteen powder, gel, or liquid-thickening agents and four brands of commercially pre-prepared thickened liquids (produced by manufacturers from Australia, Canada, Japan, the United Kingdom and the United States of America) were either donated or purchased prior to the meeting. Thickening agents included starch, gums, or combinations of starch and gums. Participants at the face-to-face meeting worked in pairs to prepare samples of graded thickness according to the manufacturer's instructions. The various thickening agents were mixed with cranberry juice (Ocean Spray). Previous national guidelines have identified a need for dysphagia diet frameworks to include a level for thickened infant milk, which is thinner than the first level of thickness commonly used for adults, but will still flow through a nipple/teat [7]. Therefore, in cases where the manufacturer's instructions on products typically used for adults specified only three gradations of thickening, an additional level between thin and the first level of thickening was prepared to produce a thickness akin to thickened infant milk. Half of the manufacturer recommended the amount of thickener for the first level of thickened drinks was used to prepare this new thickness level. Samples of human milk and infant formulas, including specialty anti-regurgitation, semi-elemental, and elemental formulas were also prepared and thickened with a view to ensuring that the framework would address needs across the pediatric-to-adult continuum. The resulting array of samples comprised four columns by 17 rows of liquid, such that the far left column was the thickest item and the far right column was the thinnest for that product according to the manufacturer's instructions. This array, which is illustrated by the schematic diagram in Fig. 2, allowed comparison of consistency across the nominally similar items in each row. Participants then continued to work in pairs to evaluate, measure, and describe the flow characteristics of all items in the array.

As well as assessing the consistency of products in each level, participants concurrently assessed the efficacy and reliability of several available methods for subjective and objective measurement, including: visual inspection; stirring; pouring from a spoon or cup; oral sampling and tasting; the line spread test [29]; and gravity flow tests using drinking straws and syringes of various dimensions. Rheological data were not obtained at the time, however, some members of the committee had familiarity with existing rheological measures for some of the products present. Based on these extensive tests, clusters of similarly behaving liquids were created. Further testing and discussion enabled the committee to confirm the similarity of liquids in each of five clusters corresponding to five levels of drink thickness (including thin) and to develop descriptions of the flow characteristics of each level. The syringe-based flow test was confirmed to be the preferred testing method for quantifying liquid consistency and it was agreed that four members of the committee with food science expertise would do further testing of this method upon return to their cities of origin to confirm construct validity in comparison to laboratory rheology and to establish boundary points between levels for liquids with different flow characteristics. This subsequent verification testing led to finalization of the IDDSI Syringe Flow Test (see Results section, below). The syringe-based flow test also affords the ability to evaluate liquids that are not typically considered 'drinks' such as condiments (e.g., sauces), liquid foods (soups), and nutritional supplements or liquid medication. The text term 'thickened liquids' is intended to include all of these items in addition to thickened drinks.

Foods

In order to develop a system for categorizing food texture, labels and descriptors for five different levels were proposed through a group nomination process. A hotel chef (naïve to dysphagia and texture-modified food used for this population) was then asked to prepare foods from the hotel menu in consistencies matching the draft labels. These samples, together with samples of ready-to-use texture-modified foods donated by industry, were assessed by the committee using spoons and forks (dropping and pressure tests) and oral appraisal, providing the opportunity to consider mouthfeel and the behavior of the sample in the mouth.

Through debate to the point of group consensus, the committee developed definitions of thickened liquids and texture-modified foods, together with the physiologic rationale for each level in the draft framework. Descriptors

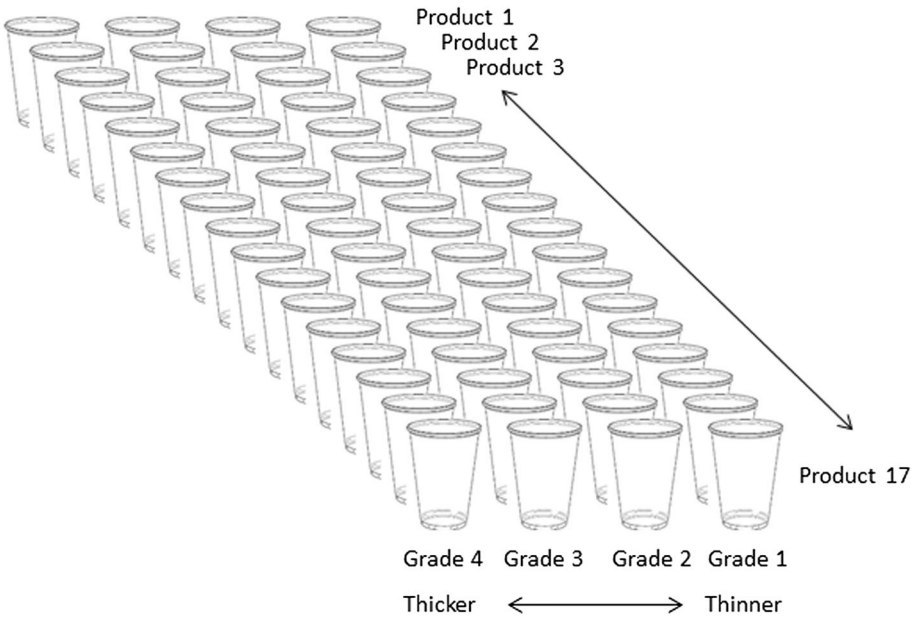


Fig. 2 Set up of thickened liquids for comparison and evaluation

and physiologic rationale for each level were based on the shared experience of experimenting with a very broad range of currently available dysphagia products, combined with each expert's relevant experience, and drawing from descriptions in all available national standard documents. Proposed labels were assessed via readability scores (Flesch–Kincaid Reading Ease score [30]) to confirm ease and understanding of terms in English. In addition, translation to languages other than English was achieved with assistance from personal contacts and volunteers, so that provisional translations of the terms were developed over the next month in Afrikaans, Arabic, Dutch, Farsi, French, German, Greek, Hebrew, Italian, Japanese, Korean, Mandarin, Portuguese, Spanish, Swedish, Turkish, and Vietnamese. Preferred methods of objective testing for liquids and foods were discussed with plans to review and finalize these following laboratory assessments upon return to cities of origin and a planned second international stakeholder survey of the draft framework. These steps led to a consensus-based and evidence-informed draft framework for public consultation, described in the Results section below.

Survey 2: Feedback on Draft IDDSI Framework

A second international stakeholder online survey was designed with assistance from the Australian Survey Research group (ASR) to gather feedback on the draft framework. ASR administered, analyzed, and reported on the survey. The survey was announced and disseminated in the same way as the preceding current practice survey.

Respondents were specifically asked to use Likert-scale responses with additional free-text comment boxes to

- provide demographic information (e.g., the stakeholder group they identified with; country they lived/worked in);
- provide feedback regarding the draft framework:
  - colors (ability to distinguish; ease of implementation; ease of reproducibility);
  - number of levels (too few/too many, about right; ease of implementation);
  - pyramid diagram (ease of understanding and implementation);
  - names or labels of each level (ease of understanding and implementation);
  - specific questions about the terms "slightly thick" liquids, "minced and moist" food, and the label "Level 7 minus";
  - detailed definitions of each level (ease of understanding; usefulness; relevance); and
  - the Syringe Test (ease of understanding; likelihood of implementing the test).
- Comment about the overall framework; and
  - what works well with the proposed framework; likelihood of implementation; and factors that would assist or impede implementation.

The online survey was open from May 1 to 1 June 1, 2015. The results of the stakeholder survey informed the final framework. Robust committee discussion followed via email and teleconferences between July–November 2015. The final framework comprises: (a) a diagram of the framework, including labels and colors; (b) detailed definitions and testing methods for liquids; and (c) detailed

definitions and testing methods for foods (see Appendix in supplementary material).

Results

Survey 1 (Current International Practice)

The current practice survey yielded responses from 2050 participants representing 33 countries. The majority of responses came from Canada, the USA, Australia, New Zealand, and the United Kingdom. Figure 3 illustrates the distribution of respondents by stakeholder group. Eighty percent of health professionals who responded saw adults with dysphagia, 8.2% saw children with dysphagia, and 16.5% saw a mixed caseload. Health professionals predominantly saw individuals in hospital settings (>60%) and approximately one quarter saw individuals in the community or aged care settings.

Healthcare professional respondents reported use of both site-prepared and commercial ready-to-use modified products. This was particularly true with respect to the preparation of texture-modified foods, for which fewer than 1% reported exclusive use either of commercial or on-site preparation methods. For drinks, exclusive use of commercially pre-thickened drinks was reported by 17% of respondents who had pediatric caseloads, and 30% of respondents whose caseloads included adults or a mix of adults and children. Exclusive in-house preparation of thickened drinks was more common for those working with pediatric caseloads (46%) compared to 30% for those working with adults or mixed caseloads.

Between 85 and 90% of health professionals reported using standardized terminology to describe thickened drinks and texture-modified food. However, considerable variation in terminology was observed from the responses obtained both within and between countries around the world. There were 27 different labels reported to be in use to refer to ≤ 5 levels of drink thickness. Most commonly, drink options were reported to include regular thin liquids plus three or more levels of thickened drinks (see Table 2). Of particular note, survey responses confirmed use with pediatric and palliative care clients of slightly thickened drinks that are thicker than water but thinner than the thickened drinks commonly used for adults [7, 11, 17, 31–34].

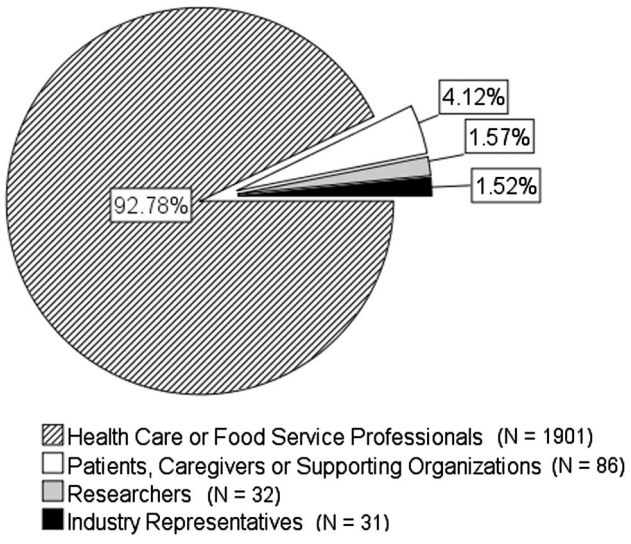


Fig. 3 Distribution of survey respondents by stakeholder group

For texture-modified foods, a total of 54 labels were reported to be in use to refer to ≤5 levels. Food options were reported to commonly include regular, non-modified foods plus four to five levels of texture modification (see Table 3). Responses from all stakeholder groups indicated support for international standardization.

The survey responses showed that some terms were not commonly used or familiar in all countries. For example, the terms "pudding," "minced," and "nectar" while understood by respondents from western cultures were not understood by respondents from Asia. Some currently used terms were considered to be problematic for certain populations. For example, it was noted that in the pediatric population and specifically children under 12 months of age, 'honey from bees' is contraindicated due to botulism risk. Thus, use of the term 'honey thick' was not felt to be an appropriate label for liquids served to pediatric populations. In addition, comments suggested that perceptions of honey differ considerably, as honey comes in crystalline, thick, and thin runny forms. Color coding was reported to be the most commonly used schema (53%) for differentiating different levels of thickened drinks or texture-modified foods; however, there was no congruency in colors chosen.

Of the respondents to the healthcare and food service professional stakeholder survey, 41 and 43%, respectively, reported that they test the consistency of foods and drinks to confirm suitability prior to serving. Consistency testing was more common among patients and caregivers, of whom, 57 and 60% reported testing foods and drinks, respectively. Visual inspection or observation was the most commonly used method of testing, regardless of stakeholder type. Patients, caregivers, and health professionals also reported using a spoon drop test or a utensil such as a fork for testing foods and liquids. Industry respondents, however, were most likely to assess liquids using a viscometer, Bostwick consistometer, or rheometer, in conjunction with visual inspection. For foods, industry respondents reported use of a texture analyzer, sieve, Bostwick consistometer, and visual inspection.

Table 2 Thickened drink names and number of levels by world region

Region	Names (least to most modified)
Africa	Normal/regular, nectar, syrup, pudding, thick
Australia + New Zealand	Thin, mildly thick/level 150, moderately thick/level 400, extremely thick/level 900
Asia	Thin, slightly thick, mildly thick, medium thick, extra thick
Canada	Thin, nectar, honey, pudding
Europe	Normal, syrup/slightly thick, nectar, honey, pudding
Ireland	Regular/normal, Gr 1, Gr 2, Gr 3, Gr 4
Middle East	Thin, mildly thick, moderately thick, other thick
South America	Liquid, slightly thick, nectar, honey, pudding
United Kingdom	Normal, stage 1, syrup, custard, pudding/stage 3
United States of America	Thin, nectar, honey, pudding

Note 27 different labels were identified internationally for ≥ 5 liquid thickness levels

Table 3 Texture-modified food names and number of levels by world region

Region	Names (least to most modified)
Africa	Normal, Soft, chopped, puree/mashed, liquid/blender
Australia + New Zealand	Full/normal, soft, minced + moist, puree/smooth puree
Asia	Regular, soft, minced/shredded, congee/puree, liquidized/blenderized
Canada	Regular, soft, minced, puree
Europe	Normal, soft/tender/cut up, ground/puree, liquid
Ireland	Regular, soft, minced + moist, puree/smooth puree, liquidized
Middle East	Solid, soft, minced + mashed, other puree
South America	Solid, soft, mashed, thick puree, liquidized
United Kingdom	Normal, fork mashable/soft, pre-mashed/texture D, puree, thin puree
United States of America	Regular, advanced/stage 3, mechanical soft/chopped/stage 2, ground, puree/stage 1

Note 54 different labels were identified internationally for ≥5 food texture modification levels

Draft IDDSI Framework

The draft framework resulting from the 2015 face-to-face meeting was represented as a continuum of 8 levels with foods and liquids displayed on a single scale using a twin-pyramid design showing foods in the top, inverted pyramid and liquids in the bottom, standing pyramid (see Fig. 4). The decision to use the pyramid image was partly influenced by the fact that a pyramid was already in use nationally in Japan for dysphagia diets. In addition to making decisions about the pyramid graphic, the number of levels, and the numbering scheme, the committee chose a draft color scheme with the aim of making each color as distinguishable as possible. It was decided that the color red should be avoided, given that red is frequently used as a color to denote alarm and danger in medical contexts and may also have other symbolism in some cultures.

A novel feature of the draft framework was the decision to recognize that certain food textures shared flow properties with thickened liquids creating an overlap zone in the middle of the framework. Using the same numbers to refer to both food and drink items at these levels, recognized the shared flow properties of these textures. Specifically, Level 3 was used both for Liquidized foods and Moderately Thick fluids, while Level 4 was used both for Pureed food and Extremely Thick fluids. All other levels had distinct flow or texture properties.

Draft Definitions

The committee developed detailed definitions for each level of the draft Framework, based on (a) the measurement activities conducted at the 2015 face-to-face meeting; (b) drawing from descriptors in all available national standards documents; and (c) the literature describing properties that increase risk for choking [35–44]. The draft definitions included a warning after Level 6 to clarify that the physiological skills of being able to both bite and chew

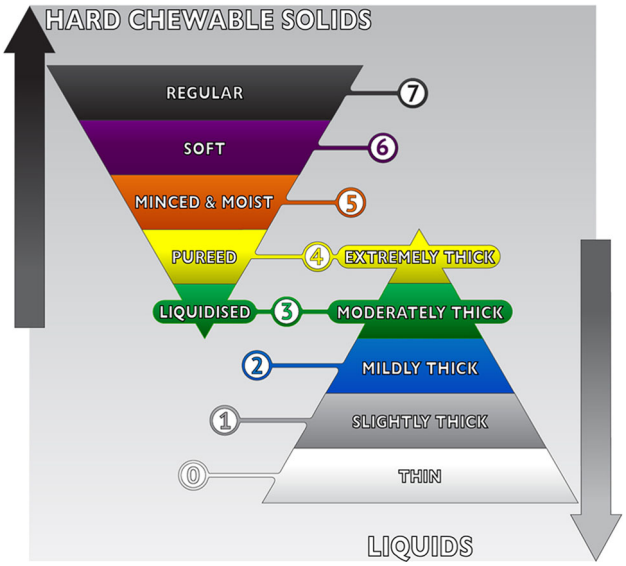


Fig. 4 Draft pyramid image, highlighting the overlap zone

food were required to safely transition to Level 7 Regular foods. Physiological contraindications for advancing to Level 7 were listed, such as xerostomia, requirement for dentures, difficulty managing mixed textures, impulsive behavior, cognitive impairment, delayed oral skills (dentition, chewing development), and fatigue (impaired strength or stamina). A level that was tentatively titled 'Level 7 Minus' was included to capture food textures that are hard in their original state but break down quickly with moisture or temperature change and can then be manipulated with minimal chewing or just with tongue pressure.

Draft Measurement Guidelines

Liquids

The 2015 face-to-face meeting included evaluation and discussion of the available testing methods for liquids: viscosity measurement was rejected due to being inaccessible in most situations and not necessarily capturing the important textural properties for swallowing (see "Discussion" section below). The draft IDDSI framework included a description of a gravity flow testing method for liquids using a syringe aiming to provide physiologically relevant flow conditions in a convenient, accessible, inexpensive test (see "Discussion"). An explanation of the gravity flow test was included in the survey to gauge acceptance of the method prior to final development. Stakeholder feedback indicated that the test was easy to understand and to implement. Detailed information about the gravity flow test is shown in "Final IDDSI Framework" section (see "Results" below).

Foods

Formal assessment of food texture commonly requires complex and expensive machinery, such as Food Texture Analyzers. This type of assessment was rejected as a practical measurement option given the lack of access to food texture analyzers and expertise or interpretation. The draft framework did not include quantitative methods for testing food texture, although the committee agreed that a method to distinguish food into the various categories was highly desirable. Subsequent to stakeholder feedback on the draft framework the committee developed practical quantitative methods for testing food size and texture (see "Final IDDSI Framework" in "Results" for more details).

Survey 2: International Feedback on Draft IDDSI Framework

The draft framework was submitted to international stakeholder consultation with a total of 3190 respondents residing in 57 different countries. The majority of respondents (87%) were health professionals working with dysphagia, although responses were also collected from caterers providing food to people with dysphagia, researchers/academics, industry that provides products to people with dysphagia, professional associations, government/regulatory bodies, caregivers to persons with dysphagia, and persons with dysphagia. Ninety percent of respondents came from English-speaking backgrounds and predominantly northern hemisphere countries. Fifty-three percent of respondents indicated that they or their organization were likely to implement the framework, with 28% neutral (see Fig. 5). Fewer than 19% of respondents indicated that implementation of the framework was unlikely.

Feedback regarding the colors representing the different food textures and thickened drink levels showed that they were considered easy to distinguish from each other and easy to implement. The number of levels was considered by more than two-thirds of respondents to be 'about right' and response to the twin-pyramid design was positive. Eighty percent of respondents rated the relevance and amount of information in the detailed definitions as 'excellent' or 'good.' Seventy-three percent of respondents indicated that the description of the syringe test was easy to understand. Clinicians who treated pediatric populations and people with developmental disability confirmed the need to have a category that included 'meltable' or 'dissolvable' solid foods. Forty percent of respondents agreed with the inclusion of Level 7 Minus with the same number neutral regarding its inclusion. The survey consultants (ASR) recommended that IDDSI review the framework based on the feedback received and make adjustments. This process of review and discussion occurred between June and November 2015.

Final IDDSI Framework

The final framework is shown in Fig. 6. Notable changes from the draft to the final framework included delineation of the 'transitional foods' side-bar category to replace 'Level 7 minus,' changes to the color scheme and the inclusion of specific testing methods for foods.

The label 'Level 7 Minus' was deleted from the framework and replaced with the term 'Transitional foods,' running alongside Levels 5–7 on the inverted food pyramid. This location reflects the fact that transitional foods are regular foods (Level 7) with special textural properties such that with the application of moisture (e.g., saliva) or a change in temperature, they rapidly change their texture, crossing boundaries between levels. The colors were reviewed in detail and assessed for suitability for people with color blindness (e.g., protanopia, deuteranopia, tritanopia and monochromatism) to distinguish the framework colors. Based on the review, certain colors were changed to maximize the difference in color between neighboring levels. The final scheme has six colors plus black and white that are individually distinguishable across all the different types of color blindness tested and particularly for red blind and green blind, which is the most common variant [45]. Specifically, Level 0 is white; Level 1 is gray; Level 2 is pink; Level 3 is yellow; Level 4 is green; Level 5 is orange; Level 6 is blue; and Level 7 is black.

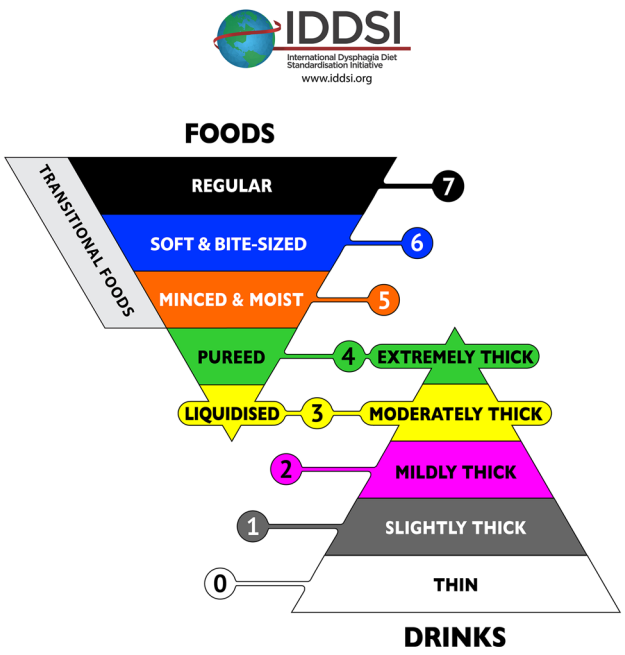


Fig. 6 The final IDDSI framework graphic

Liquid Specifications and Measurement

The draft framework introduced the concept of the gravity flow test. The gravity flow test uses a 10-mL slip tip hypodermic syringe. Although 10-mL syringes were initially thought to be identical throughout the world based on

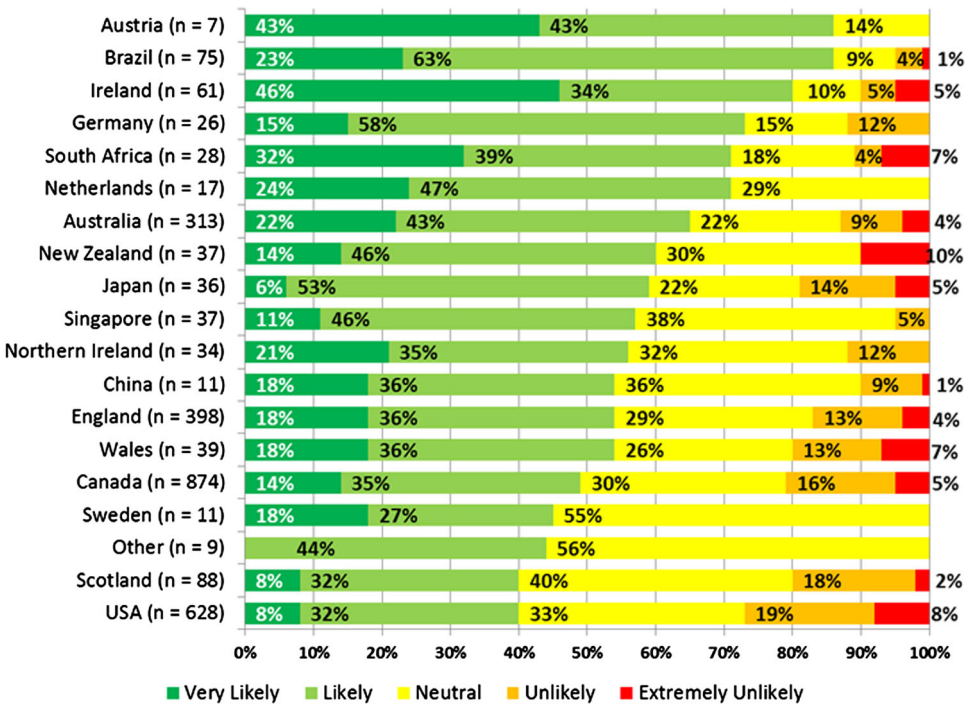


Fig. 5 Stakeholder survey 2—International indications of likelihood of implementation. *Note* Caution should be used in interpreting small sample sizes from specific countries

reference to an ISO standard (ISO 7886-1) [46], it has subsequently been determined that the ISO document refers only to the nozzle of the syringe and that variability in barrel length and dimensions may exist between brands. As illustrated in Fig. 7, a syringe with a measured length of 61.5 mm from the zero line to the 10 mL line was used as the reference syringe (BD^{TM} syringes were used for the development of the tests). To conduct the flow measurement, 10 mL of liquid is placed into an empty syringe and a stopper or finger is placed at the nozzle to impede flow until ready. When ready, the stopper or finger is removed from the syringe nozzle with flow allowed for 10 s. At 10 s, the nozzle is again blocked so that the volume of liquid remaining in the syringe can be recorded. The IDDSI Flow Test instructions and interpretations are included in the Appendix in supplementary material. During developmental testing by the committee, the IDDSI Flow Test was found to be suitable for thin liquids, naturally thick liquids and liquids thickened with a range of thickening agents (gums and starches) as well as items such as gravy, sauce, condiments, smooth soup, nutritional supplements, and liquid medication. Although the equipment is simple, the test has been found to categorize a wide range of liquids reliably in agreement with currently existing laboratory tests and expert judgment. It has been found to be sensitive enough to demonstrate small changes in thickness associated with change in serving temperature. The test requires that liquids are able to flow under their own weight, which corresponds to the threshold between level 3 and 4. While the test can be used to confirm whether a material is above the threshold for level 4 (no flow will occur), it is more convenient to simply use a spoon to determine whether the material is able to hold its shape or not. A number of countries use Fork Drip Tests to describe flow of thickened drinks or pureed food in their national terminologies [7, 17, 19]. Fork Drip Test criteria were developed for IDDSI Levels 3–5.

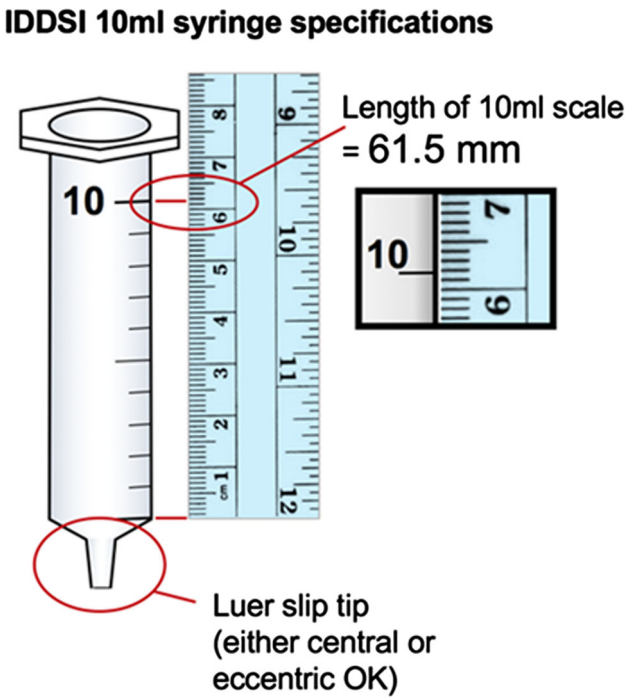


Fig. 7 Example of a slip tip syringe that complies with IDDSI measurement requirements

Food Texture Specifications and Measurement

The systematic review demonstrated that the properties of hardness, cohesiveness, and slipperiness were important factors for consideration [27, 28]. In addition, as noted in the initial publication documenting the need for a new international framework, the size and shape of food samples have been identified as relevant factors for choking risk [26]. In view of this information, the IDDSI committee agreed that measurement of foods needed to capture *both* the mechanical properties (e.g., hardness, cohesiveness, adhesiveness, etc.) and the geometrical, size, or shape attributes of the food. Prior to release of the final framework, the committee worked to develop specifications based on the best available practical tools: the surveys had reported that utensils such as forks and spoons were commonly used for assessment of texture-modified food and thickened liquids. Assessments using chopsticks and finger tests have also been incorporated in recognition that these may be the most accessible methods in some countries.

Food Particle Size

Assessment of foods requires a combination of evaluation for particle size and food hardness, cohesiveness, and adhesiveness. With regard to particle size, 2–4 mm represents the size of chewed particles that healthy adult individuals naturally masticate and reduce hard foods to for swallowing [47]. For Level 5 Minced & Moist, the recommended particle size for food served to adults is 4 mm. In recognition of the smaller anatomy and in lieu of pediatric research, for infants, the recommended particle size for Level 5—Minced & Moist food is 2 mm. The slots/-gaps between the tines/prongs of a standard metal fork typically measure 4 mm, which provides a useful compliance measure for particle size of Minced & Moist foods served to adults.

For hard and soft solid foods served to adults, a maximum food sample size of $\sim 1.5 \times 1.5$ cm is recommended, which is the approximate size of the adult human thumb nail [48] and the approximate width (from left to right] of the tip a standard metal fork. These dimensions

represent the food texture industry standard 'bite sample' [47, 49], but most importantly are small enough to pass completely into the average adult trachea rather than obstruct it at the laryngeal inlet if accidentally inhaled [50, 51]. Tracheal size for adult males is 22 mm (range 15–27 mm) and for adult females is 17 mm (13–25 mm) [50]. Furthermore, food particle size of these dimensions has been identified as reducing asphyxiation risk [51].

Particle sizes for soft and hard food served to children younger than 5-year old are recommended to be no larger than 0.8 cm, which again relates to tracheal size and reduction of asphyxiation and choking risk [52]. Tracheal size of infants obviously changes as children grow. At age 20 months, the infant's anteroposterior dimensions of the region just below the vocal cords, at the entrance to the trachea are approximately 3.8 mm × 6.5 mm. At 3 years 4 months (40 months), the dimensions are 7 mm x 3.9 mm and at 5 years of age the dimensions are approximately 8 mm × 4 mm [53]. It is for this reason that the Level 6—Soft & Bite-Sized specifies a particle size of 0.8 cm or less for children (i.e., 8 mm) and Level 5—Minced & Moist specifies a pediatric particle size of 0.2 cm (2 mm). Note also that food samples that are smaller than the maximum width of the child's fifth fingernail (littlest finger) are unlikely to represent a choking risk, as this measurement is used to predict the internal diameter of an endotracheal tube in the pediatric population [54].

Food Hardness, Cohesiveness, and Adhesiveness

Chewing results in the breaking down of food, determined by a number of factors including: toughness, moisture content of the food, ability to adsorb or absorb saliva, and the fibrous nature of the food [47, 55]. The level of moisture content in food has been particularly singled out as an important variable for determining food readiness for swallowing [55]. Salivation moistens the food bolus and assists with softening, disintegration, and dilution, thus reduced salivation will hinder even fully dentate individuals from adequately preparing a bolus for swallowing. During particle size reduction while chewing, the normal bolus is not 'lump-free,' however, it is moist and cohesive. For assessment of cohesiveness and adhesiveness a spoon tilt test is recommended. In each case the sample should (a) hold its shape on the spoon; and (b) fall easily from the spoon when tilted or turned sideways. There should be little residue left on the spoon. These characteristics provide a bolus that is moist and cohesive, but not sticky or adhesive.

Quantification of food hardness is technically challenging because the mechanical structure of foods is generally complex. In industrial and scientific laboratories, a food texture analyzer is used to crush a sample of the food under controlled pressure and motion, but that requires motors and sensors. A practical test using a fork or spoon was previously recommended as part of the United Kingdom dysphagia diet standards [19] for assessing foods that would fall into IDDSI Levels 5–7 and transitional foods. The test involves applying a fork to the food sample to observe its behavior when pressure is applied, however, this varies with the level of force applied by the individual. In order to provide some standardization of the pressure applied, the IDDSI fork pressure test recommends that the fork be pressed onto the food sample by placing the thumb onto the bowl of the fork (just below the prongs), and pressing just hard enough to cause blanching of the thumbnail, Fig. 8a. Blanching occurs when the pressure overcomes mean arterial blood pressure and has been quantified at approximately 17 kPa, Fig. 8b. This pressure corresponds closely to a typical tongue pressure used during swallowing [56, 57]. In places where forks are not used, descriptions and testing methods have been developed for chopsticks and finger pressure testing.

To meet the requirements for Level 6—Soft & Bite-sized, a food sample should squash with the application of pressure and *not* return to its original shape when pressure is released. Transitional foods can also be identified using the Fork Pressure Test. For transitional foods, a sample 1.5 × 1.5 cm is placed in a container with 1 mL of water. Testing occurs after 1 min of food soaking has occurred. The sample qualifies as transitional food texture if the sample squashes and disintegrates and no longer resembles its original shape, or if it has melted significantly so that it no longer looks like its original shape.

Consistent with existing national terminologies and evidence from autopsy data, tables showing 'texture requirements' and 'texture restrictions' for each level were generated (see Appendix in supplementary material). Foods that have been identified in multiple autopsy reviews to increase choking risk were specifically addressed in a Frequently Asked Questions (FAQs) section (www.iddsi.org).

Release of the Final IDDSI Framework

The final framework was released by staggered roll out. The framework design including the twin-inverted pyramid design was launched at the Japanese Society of Dysphagia Rehabilitation Conference September 2015. The detailed descriptors for drinks were released online and via poster at the European Society of Swallowing Disorders Conference in September 2015; and the detailed descriptors for foods were released online and at the Food for the Elderly Conference, Hangzhou, China, in November 2015. Further to the release of the framework and detailed descriptors, and following consultation with a representative from the Australian Government Open Access and Licensing

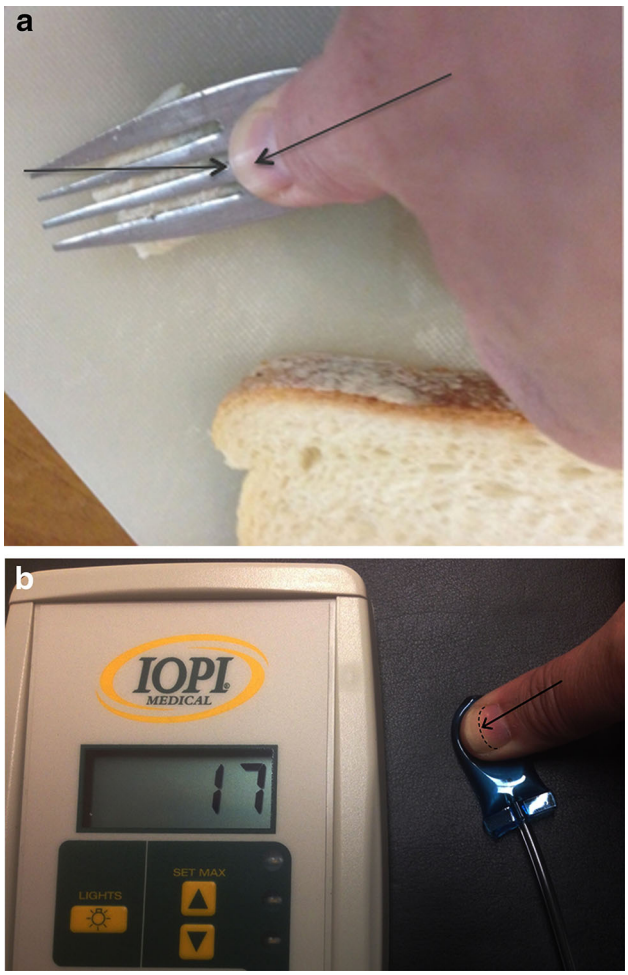


Fig. 8 **a** Illustration of the thumb nail blanching to *white* (shown by *arrow*) during Fork Pressure Test. **b** Amount of pressure required (in kPa) to blanch the thumb nail to *white*. Image used with permission from IOPI Medical (www.iopimedical.com)

Framework (AusGOAL) and Creative Commons Australia, the IDDSI Framework and detailed descriptors were licensed under the CreativeCommons Attribution Share-alike 4.0 Licence https://creativecommons.org/licenses/by-sa/4.0/legalcode to facilitate language translation.

To use the IDDSI framework and detailed descriptors, we request the following attribution:

> © The International Dysphagia Diet Standardisation Initiative 2016 @http://iddsi.org/framework/. Attribution is NOT PERMITTED for derivative works incorporating any alterations to the IDDSI Framework that extend beyond language translation.
> Supplementary Notice: Modification of the diagrams or descriptors within the IDDSI Framework is DISCOURAGED and NOT RECOMMENDED. Alterations to elements of the IDDSI framework may lead to confusion and errors in diet texture or drink selection for patients with dysphagia. Such errors have previously been associated with adverse events including choking and death.

Discussion

The International Dysphagia Diet Standardisation Initiative utilized an evidence-based method of guideline development [24, 25] to produce new global standardized terminology and definitions to describe texture-modified foods and thickened liquids used for individuals with dysphagia of all ages, in all care settings and all cultures. The final framework was developed with reference to existing national terminologies, empirical data from multiple international stakeholder consultations of people from 57 countries, systematic review of the research literature and collaborative, feedback-driven refinement. Feedback about the framework was collated and analyzed by a research group independent to IDDSI further strengthening confidence in the refinement of the final framework. The final framework consists of eight levels (Levels 0–7) that are identified by numbers, text labels, and color codes. Text labels have been scrutinized for ease of translation and color codes have been developed to be sensitive to color blindness. Descriptors are supported by simple, accessible yet objective measurement methods that can be used by people with dysphagia and their caregivers, clinicians, food service professionals, researchers, and industry to confirm the level of attribution of a food or liquid. The IDDSI framework provides a solid platform for the development of future research in the dysphagia field.

The IDDSI framework provides categorization of liquid thickness levels applicable to neonates, infants, children, and adults with dysphagia. The IDDSI systematic review found evidence confirming that thickening liquids reduces the likelihood of aspiration, however, it was not able to pinpoint specific viscosities that represent minimally effective thickening to reduce aspiration. The review did, however, find evidence to suggest that some extremely thick liquids may promote the accumulation of pharyngeal residue [23, 27, 28, 58]. This finding has been further corroborated by Newman and colleagues [59], who conducted an independent systematic review of the literature on the efficacy of thickened liquids for the management of dysphagia. Recognition that some liquids may promote residue by being ‘too thick’ is an important development for the dysphagia field. Given the paucity of research regarding therapeutic thickness levels for thickened drinks, the IDDSI framework is based on an understanding that increasing thickness has a demonstrated therapeutic benefit for reducing the likelihood of penetration/aspiration. The number of levels of drink thickness included in the framework and recommended for best practice is based on

the synthesis of international stakeholder consensus on current clinical practice. The systematic review points to an urgent need to conduct quality research to determine specific thickness levels that provide therapeutic benefit by reducing risk for penetration/aspiration and/or improving swallowing function. The IDDSI framework provides a reference point for this research and with future developments it is anticipated that the IDDSI levels will be refined to reflect new evidence regarding therapeutic thickness levels.

International Food Textures for Dysphagia Management

The IDDSI framework provides categorization of food textures applicable to babies, infants, children, and adults with dysphagia. Children younger than three years of age, adults over 65 years of age, individuals with poor dentition, and those with neurological conditions are at high risk of death from asphyxiation on food [35, 60]. In healthy people, regardless of the initial state of the food, after oral processing and at the point of swallow initiation, the bolus is a cohesive mass. Texture modification mechanically alters the food prior to ingestion to the level that is required to promote safe swallowing of the bolus. The paucity of research into the therapeutic use of food texture modification for dysphagia management means that the recommendations in this document regarding food texture are based on an understanding that altering food texture modification has demonstrated a therapeutic benefit for reducing the risk of choking. Empirical evidence gathered from the current practice survey indicated that foods are commonly altered in both size (chopped, diced) and texture (soft, puree) to reduce choking risk. This practice is consistent with evidence in the literature specific to choking and asphyxiation risk, which reveals that food textures that pose the most risk are categorized according to texture, shape, and size. Specifically, foods that are described as hard or dry; chewy or sticky; crunch or crumbly; floppy; fibrous or 'tough'; have husks; are stringy; round or long in dimension or consist of multiple or 'dual' textures are high choking risks [35–44]. Additional discussion regarding choking risk can be found in the IDDSI Foundation manuscript [26].

The IDDSI framework promotes strict adherence to both particle size and food texture requirements. For Level 6—Soft & Bite-Sized, international feedback has requested justification for the food particle size on this diet. To reduce choking risk, pre-cut food to 1.5 × 1.5 cm has been recommended. For easy reference, it has been determined that the width of a standard dinner fork (from left to right running perpendicular to the prongs) corresponds approximately to this 1.5 cm dimension. It is not possible to guarantee that a person with dysphagia will be able to cut food to this size, or that care staff or family will be available to pre-cut the food. Individuals with cognitive impairment are at increased risk of choking with poor ability to self-monitor food size and rate of ingestion [51, 61]. Some elderly people without a formal dysphagia diagnosis, but with fewer than 20 teeth, or with dentures may benefit from soft food for ease of mastication. These individuals do not strictly require the stringent particle size requirement described in Level 6—Soft & Bite-Sized, but perhaps they also do not strictly require a dysphagia diet. In these cases, it is suggested that facilities consider specifying soft options from a regular diet. This option should not be considered as part of the dysphagia diet. The testing methods outlined in the IDDSI Framework are generalizable to testing the softness of food texture in such circumstances. It should be noted that the loss of occlusal units affects bite force. Individuals with greater than 20 teeth (10 paired occlusal units) are reported to have normal bite force values of ~555 N. An exponential decline in bite force is observed with a reduction in the number of teeth, for example, 383 N for 10–19 teeth remaining; 180 N for 1–9 teeth remaining and 155 N for edentulous individuals [62]. Regardless of a formal dysphagia diagnosis, reduced bite force and poor masticatory efficiency increases choking risk [63].

Accessible, Objective Testing Methods for Texture-Modified Liquids and Foods

To date, the measurement of fluid thickness in most national terminologies has been based on subjective methods such as flow through the prongs of a fork which has inherent variability [64]. Objective quantification was highly desirable but was challenging. The only national standard to recommend categorization of liquids according to quantified viscosity ranges is the National Dysphagia Diet (NDD), developed in the USA in 2002 [15]. However, the IDDSI committee considered that there were major practical and scientific limitations to viscosity measurement as follows:

- The lack of access to testing equipment and the expertise required to perform and interpret rheological testing.
- Viscosity is only one of a number of relevant parameters affecting liquid flow; others include density, yield stress, sample temperature, elasticity, and propulsion pressure [65–69].
- Drinks thickened with different thickening agents—or naturally thick—may have the same measurement of apparent viscosity at the specified test shear rate (e.g.,

NDD: 50 s^{-1}) and yet may have very different flow characteristics in practice [27, 28, 70–73].
- The non-Newtonian nature of thickened drinks makes them impossible to characterize fully with only one viscosity measurement [70, 74, 75].
- In addition to variations in flow associated with drink characteristics, flow rates during swallowing are expected to differ depending on a person's age and level of impairment of swallowing function [59].

For these reasons, a measurement of viscosity has not been included in the IDDSI descriptors. Instead, an objective and practical measurement has been selected by IDDSI to classify liquids based on their rate of flow under the action of gravity down a narrow tube with an orifice at the bottom. Such tests have a history in the dairy industry for studying oral perceptions of milk, cream, and yogurt (e.g., [76, 77]). The controlled dimensions selected are broadly representative of drinking through a straw or beaker and the regime of top-down flow through a narrow tube with exit through a small orifice has physiological parallels with bolus flow through the pharynx, with exit via the upper esophageal sphincter. This type of extensional flow (as opposed to shear) has been hypothesized to be more relevant to perception and to dysphagia [78–80]. Rather than specify a proprietary instrument, we have specified a common 10-ml syringe due to its increased availability and affordability globally. The syringe may be disposed after each use or washed and re-used; the 10 ml sample fluid would be discarded. The nature of the test means it is possible to measure drinks at the point of service. Although we would not expect this to be performed routinely, it does provide a standardized objective measure for training, auditing, and research. The ability to measure liquid thickness also provides opportunity to accurately audit thickness of sauces, condiments, soups, nutritional supplements, and liquid medications at the time of preparation and at the point of serving.

It is desirable to develop similar practical tools for quantifying food textures, however, food texture assessment provides more variables for assessment than drinks as both texture and size requirements are needed. Additionally, a degree of added pressure is required in order to deform the materials (which will not flow under gravity), and that is difficult to control in an inexpensive and globally standardized manner. The IDDSI Fork pressure test provides guidelines with greater quantification than would be achievable by text alone. The dimensions of common forks have been found to be fairly consistent internationally, which provides some ability to specify particle size, and a version of the test has been produced for chopsticks. However, it is acknowledged that further work in the area of food texture assessment for texture-modified foods is warranted.

Limitations of the Current Study

The IDDSI process utilized online stakeholder surveys to gather empirical evidence. It is acknowledged that despite the range of stakeholder groups engaged that the sample size of the stakeholder groups was uneven. The largest group of respondents in both surveys was healthcare professionals. People with dysphagia and their caregivers made up the smallest stakeholder group. Often communication impairment accompanies dysphagia, and it is possible that communication impairment may have limited respondents' ability to participate in the surveys. The surveys were heavily influenced by responses from English-speaking countries, with the top 10 countries of origin responses coming in order from Canada, USA, England, Australia, Scotland, Brazil, Ireland, New Zealand, Singapore, and Japan. The stakeholder groups consisted of motivated responders. It is noted that in the development of the Australian standardized terminology for texture-modified foods and fluids that differences were seen in the responses from motivated vs. targeted respondents [7]. Due to the scale of the IDDSI initiative, it was not possible to solicit responses from individuals or organizations that had not already volunteered to take part in the surveys. It could be argued that those motivated to respond will be more active in change management.

Future Directions

The IDDSI framework has been well received by the international community. The IDDSI Board is in the process of developing materials and resources to assist interested parties to transition to the IDDSI framework (www.iddsi.org). The IDDSI web site aims to provide a large and up-to-date resource for the international community to share and discuss ideas and experiences relating to texture modification and adoption of the IDDSI framework. We hope this will include practical tips and guidance for local regions. The web site is a channel for stakeholders to feed back evidence of the success or limitations of the IDDSI framework across settings internationally.

The IDDSI framework is considered a living document such that it will be formally reviewed at specified intervals with new editions noted by updated version numbers and year of review. As research is conducted and technology continues to expand, it is anticipated that further refinements to the framework and detailed definitions will occur. The framework and detailed definitions will be formally reviewed in 2020 to ensure that the evidence base supporting the IDDSI framework remains current.

Acknowledgements The IDDSI committee would like to acknowledge the interest and participation of the global community including

patients, caregivers, health professionals, food service professionals, industry, professional associations, government and regulatory bodies, and researchers. The International Dysphagia Diet Standardisation Initiative Inc (IDDSI) is an independent not-for-profit entity. IDDSI is grateful to a large number of agencies, organizations, and industry partners for financial and other support. Sponsors have not been involved with the design or development of the IDDSI framework. The assistance and expertise of the Australian Survey Research Group is also gratefully acknowledged for their help with survey design, web-hosting, and data interpretation (http://aussurveys.com/). The assistance and advice of Barrister Baden Appleyard, National Program Director- AusGOAL is gratefully acknowledged for advice and assistance to determine wording and type of Creative Commons license. The following industry sponsors are gratefully acknowledged for their financial support of the IDDSI initiative: Nestlé Nutrition Institute; Hormel Thick & Easy; Nutricia Advanced Medical Nutrition; Campbell's Food Service; Trisco Foods; apetito; Food Care (Japan); Flavour Creations; Simply Thick; Lyons.

Compliance with Ethical Standards

Conflict of interest IDDSI and the authors have no conflicts of interest to disclose

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Castellvi MTC. Theories of terminology. Terminology. 2003;9:163–99.
2. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. Dysphagia. 1996;11:93–8.
3. McHorney CA, Robbins J, Lomax K, Rosenbek JC, Chignell K, Kramer AE, Bricker DE. The SWAL-QOL and SWAL-CARE outcomes tool for oropharyngeal dysphagia in adults: III. Documentation of reliability and validity. Dysphagia. 2002;17:97–114.
4. O'Neil KH, Purdy M, Falk J, Gallo L. The dysphagia outcome and severity scale. Dysphagia. 1999;14:139–45.
5. Crary MA, Mann GD, Groher ME. Initial psychometric assessment of a functional oral intake scale for dysphagia in stroke patients. Arch Phys Med Rehabil. 2005;86:1516–20.
6. Robbins J, Nicosia MA, Hind JA, Gill GD, Blanco R, Logemann JA. Defining physical properties of fluids for dysphagia evaluation and treatment: perspectives on swallowing and swallowing disorders. Dysphagia. 2002;11(2):16–9.
7. Atherton M, Bellis-Smith N, Cichero JAY, Suter M. Texture-modified foods and thickened fluids as used for individuals with dysphagia: Australian standardised labels and definitions. Nutr Diet. 2007;64:S53–76.
8. Rutherford MA. Standardized nursing language: what does it mean for nursing practice? Online J Issues Nurs. 2008;13:1–9.
9. Norton K, Norton L, Sadgrove D. Position statement on physical activity and exercise intensity terminology. J Sci Med Sport. 2010;13:496–502.
10. Creswell KM, Bates DW, Phansalkar S, Sheikh A. Opportunities and challenges in creating an international centralized knowledge data base for clinical decision support systems in ePrescribing. BMJ Qual Saf. 2011;20(7):625–30.
11. Jukes S, Cichero JA, Haines T, Wilson C, Paul K, O'Rourke M. Evaluation of the uptake of the Australian standardized terminology and definitions for texture modified foods and fluids. Int J Speech Lang Pathol. 2012;14:214–25.
12. South Australia: Coronial Inquest, Dimitra Damianou. 1997. http://www.courts.sa.gov.au/CoronersFindings/Lists/Coroners%20Findings/Attachments/284/DAMIANOU%20Dimitra.pdf. Access date: June 21, 2016.
13. Australian Commission on Safety and Quality in Health Care: Literature review: Medication safety in Australia. Sydney, 2013. http://www.safetyandquality.gov.au/wp-content/uploads/2014/02/Literature-Review-Medication-Safety-in-Australia-2013.pdf. Accessed 21 June 2016.
14. Licea M, Klein M: Patient chokes to death after hospital feeds her sandwich she can't eat: New York Post: News Corp, 2016. http://nypost.com/2016/06/19/patient-chokes-to-death-after-hospital-feeds-her-sandwich-she-cant-eat/. Accessed 21 June 2016.
15. American Dietetic Association. National dysphagia diet: standardization for optimal care. Chicago, IL: American Dietetic Association; 2002.
16. Ministry of Health Labour and Welfare, Japan: Food Safety Information: Food with Health Claims, 2009. http://www.mhlw.go.jp/english/topics/foodsafety/fhc/index.html Access date: June 21, 2016.
17. Irish Association of Speech-Language Therapists and Irish Nutrition and Dietetic Institute: Irish consistency descriptors for modified fluids and foods. 2009. http://www.iaslt.ie/newFront/Documents/PublicInformation/Dysphagia/Irish%20consistency%20descriptors%20for%20modified%20fluidsandfood.pdf. Access date: June 21, 2016.
18. Wendin K, Ekman S, Bulow M, Ekberg O, Johansson D, Rothenberg E, Stading M. Objective and quantitative definitions of modified food textures based on sensory and rheological methodology. Food Nutr Res. 2010;54:5135.
19. National Patient Safety Agency, Royal College of Speech and Language Therapists, British Dietetic Association, National Nurses' Nutrition Group, Hospital Caterers' Association: Dysphagia diet food texture descriptions. 2011. http://www.thenacc.co.uk/assets/downloads/170/Food%20Descriptors%20for%20Industry%20Final%20-%20USE.pdf. Accessed 21 June 2016.
20. Anderson UT, Beck AM, Kjaersgaard A, Hanson T, Poulsen I. Systematic review and evidence based recommendations on texture modified foods and thickened fluids for adults (≥18 years) with oropharyngeal dysphagia. e-SPEN J. 2013;8:e127–34.
21. United States of America Food and Drug Administration: Facts about generic drugs. 2015. http://www.fda.gov/Drugs/ResourcesForYou/Consumers/BuyingUsingMedicineSafely/UnderstandingGenericDrugs/ucm167991.htm. Accessed 21 June 2016.
22. Logemann JA, Gensler G, Robbins J, Lindblad AS, Brandt D, Hind JA, Kosek S, Dikeman K, Kazandjian M, Gramigna GD, Lundy D, McGarvey-Toler S, Miller Gardner PJ. A randomized study of three interventions for aspiration of thin liquids in patients with dementia or Parkinson's disease. J Speech Lang Hear Res. 2008;51:173–83.
23. Robbins J, Gensler G, Hind J, Logemann JA, Lindblad AS, Brandt D, Baum H, Lilienfeld D, Kosek S, Lundy D, Dikeman K, Kazandjian M, Gramigna GD, McGarvey-Toler S, Miller Gardner PJ. Comparison of 2 interventions for liquid aspiration on pneumonia incidence: a randomized trial. Ann Intern Med. 2008;148:509–18.
24. Turner T, Misso M, Harris C, Green S. Development of evidence-based clinical practice guidelines (CPGs): comparing approaches. Implement Sci. 2008;3:45.

25. Qaseem A, Forland F, Macbeth F, Ollenschlager G, Phillips S, van der Wees P. Board of Trustees of the Guidelines International N: guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med. 2012;156:525–31.
26. Cichero JA, Steele C, Duivestein J, Clave P, Chen J, Kayashita J, Dantas R, Lecko C, Speyer R, Lam P, Murray J. The need for international terminology and definitions for texture-modified foods and thickened liquids used in dysphagia management: foundations of a global initiative. Curr Phys Med Rehabil Rep. 2013;1:280–91.
27. Steele CM, Alsanei WA, Ayanikalath S, Barbon CE, Chen J, Cichero JA, Coutts K, Dantas RO, Duivestein J, Giosa L, Hanson B, Lam P, Lecko C, Leigh C, Nagy A, Namasivayam AM, Nascimento WV, Odendaal I, Smith CH, Wang H. Erratum to: The influence of food texture and liquid consistency modification on swallowing physiology and function: a systematic review. Dysphagia. 2015;30:272–3.
28. Steele CM, Alsanei WA, Ayanikalath S, Barbon CE, Chen J, Cichero JA, Coutts K, Dantas RO, Duivestein J, Giosa L, Hanson B, Lam P, Lecko C, Leigh C, Nagy A, Namasivayam AM, Nascimento WV, Odendaal I, Smith CH, Wang H. The influence of food texture and liquid consistency modification on swallowing physiology and function: a systematic review. Dysphagia. 2015;30:2–26.
29. Mann LL, Wong K. Development of an objective method for assessing viscosity of formulated foods and beverages for the dysphagic diet. J Am Diet Assoc. 1996;96:585–8.
30. Flesch-Kincaid Reading Ease. https://readability-score.com/text/ Accessed 21 June 2016.
31. Stuart S, Motz JM. Viscosity in infant dysphagia management: comparison of viscosity of thickened liquids used in assessment and thickened liquids used in treatment. Dysphagia. 2009;24:412–22.
32. de Almeida MBM, de Almeida JAG, Moreira MEL, Novak FR. Adequacy of human milk viscosity to respond to infants with dysphagia: Experimental study. J Appl Oral Sci. 2011;19:554–9.
33. Cichero J, Nicholson T, Dodrill P. Liquid barium is not representative of infant formula: characterisation of rheological and material properties. Dysphagia. 2011;26:264–71.
34. September C, Nicholson TM, Cichero JA. Implications of changing the amount of thickener in thickened infant formula for infants with dysphagia. Dysphagia. 2014;29:432–7.
35. Kennedy B, Ibrahim JE, Bugeja L, Ranson D. Causes of death determined in medicolegal investigations in residents of nursing homes: a systematic review. J Am Geriatr Soc. 2014;62:1513–26.
36. Chapin MM, Rochette LM, Annest JL, Haileyesus T, Conner KA, Smith GA. Nonfatal choking on food among children 14 years or younger in the United States, 2001-2009. Pediatrics. 2013;132:275–81.
37. Food Safety Commission of Japan: Risk assessment report: Choking accidents caused by foods. 2010. https://www.fsc.go.jp/english/topics/choking_accidents_caused_by_foods.pdf Accessed 21 June 2016.
38. Morley RE, Ludemann JP, Moxham JP, Kozak FK, Riding KH. Foreign body aspiration in infants and toddlers: recent trends in British Columbia. J Otolaryngol. 2004;33:37–41.
39. Mu L, Ping H, Sun D. Inhalation of foreign bodies in Chinese children: A review of 400 cases. Laryngoscope. 1991;101:657–60.
40. Berzlanovich AM, Muhm M, Sim E, Bauer G. Foreign body asphyxiation—an autopsy study. Am J Med. 1999;107:351–5.
41. Wolach B, Raz A, Weinberg J, Mikulski Y, Ben Ari J, Sadan N. Aspirated foreign bodies in the respiratory tract of children: eleven years experience with 127 patients. Int J Pediatr Otorhinolaryngol. 1994;30:1–10.
42. Center for Disease Control and Prevention: Nonfatal Choking-Related Episodes Among Children - United States, 2001. 2002. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5142a1.htm Accessed 21 June 2016.
43. Rimell FL, Thome A Jr, Stool S, Reilly JS, Rider G, Stool D, Wilson CL. Characteristics of objects that cause choking in children. J Am Med Dir Assoc. 1994;274:1763–6.
44. Seidel JS, Gausche-Hill M. Lychee-flavored gel candies: a potentially lethal snack for infants and children. Arch Pediatr Adolesc Med. 2002;156:1120–2.
45. National Eye Institute, National Institute of Health, USA: Facts about color blindness. 2015. https://nei.nih.gov/health/color_blindness/facts_about. Accessed 21 June 2016.
46. International Standards Organisation: ISO-7886-1: 1993 (E). Sterile hypodermic syringes for single use: Part I: Syringes for manual use. 1993. https://www.iso.org/obp/ui/#iso:std:iso:7886:-1:ed-1:v1:en. Accessed 21 June 2016.
47. Mishellany A, Woda A, Labas R, Peyron MA. The challenge of mastication: preparing a bolus suitable for deglutition. Dysphagia. 2006;21:87–94.
48. Murdan S. Transverse fingernail curvature in adults: a quantitative evaluation and the influence of gender, age, and hand size and dominance. Int J Cosmet Sci. 2011;33:509–13.
49. Kohyama K, Mioche L, Bourdiol P. Influence of age and dental status on chewing behaviour studied by EMG recordings during consumption of various food samples. Gerodontology. 2003;20:15–23.
50. Brodsky JB, Macario A, Mark JB. Tracheal diameter predicts double-lumen tube size: a method for selecting left double-lumen tubes. Anesth Analg. 1996;82:861–4.
51. Samuels R, Chadwick DD. Predictors of asphyxiation risk in adults with intellectual disabilities and dysphagia. J Intellect Disabil Res. 2006;50:362–70.
52. Salgo B, Schmitz A, Henze G, Stutz K, Dullenkopf A, Neff S, Gerber AC, Weiss M. Evaluation of a new recommendation for improved cuffed tracheal tube size selection in infants and small children. Acta Anaesthesiol Scand. 2006;50:557–61.
53. Litman RS, Weissend EE, Shibata D, Westesson PL. Developmental changes of laryngeal dimensions in unparalyzed, sedated children. Anesthesiology. 2003;98:41–5.
54. Turkistani A, Abdullah KM, Delvi B, Al-Mazroua KA. The 'best fit' endotracheal tube in children –comparison of four formulae. Middle East J Anaesthesiol. 2009;20:383–7.
55. Yven C, Bonnet L, Cormier D, Monier S, Mioche L. Impaired mastication modifies the dynamics of bolus formation. Eur J Oral Sci. 2006;114:184–90.
56. Fei T, Polacco RC, Hori SE, Molfenter SM, Peladeau-Pigeon M, Tsang C, Steele CM. Age-related differences in tongue-palate pressures for strength and swallowing tasks. Dysphagia. 2013;28:575–81.
57. Youmans SR, Stierwalt JA. Measures of tongue function related to normal swallowing. Dysphagia. 2006;21:102–11.
58. Hind J, Divyak E, Zielinski J, Taylor A, Hartman M, Gangnon R, Robbins J. Comparison of standardized bariums with varying rheological parameters on swallowing kinematics in males. J Rehabil Res Dev. 2012;49:1399–404.
59. Newman R, Vilardell N, Clave P, Speyer R. Effect of Bolus Viscosity on the Safety and Efficacy of Swallowing and the Kinematics of the Swallow Response in Patients with Oropharyngeal Dysphagia: White Paper by the European Society for Swallowing Disorders (ESSD). Dysphagia. 2016;31:232–49.
60. Kramarow E, Warner M, Chen LH. Food-related choking deaths among the elderly. Inj Prev. 2014;20:200–3.
61. Morad M, Kandel I, Merrick J. Residential care centers for persons with intellectual disability in Israel: trends in mortality from

food choking during 1991-2004. Med Sci Monit. 2009;15: PH75–7.
62. Yamanaka K, Nakagaki H, Morita I, Suzaki H, Hashimoto M, Sakai T. Comparison of the health condition between the 8020 achievers and the 8020 non-achievers. Int Dent J. 2008;58: 146–50.
63. Ansai T, Takata Y, Akifusa S, Sogame A, Shimada N, Yoshida A, Hamasaki T, Awano S, Fukuhara M, Takehara T. Relationship between chewing ability and 4-year mortality in a cohort of 80-year old Japanese people. Oral Disease. 2007;13:214–9.
64. Hadde EK, Cichero JAY, Nicholson TM: Viscosity of thickened fluids that relate to the Australian National Standards. Int J Speech Language Pathol 1–9, 2015.
65. Sopade PA, Halley PJ, Cichero JAY, Ward LC. Rheological characterization of food thickeners marketed in Australia in various media for the management of dysphagia. I: water and cordial. J Food Eng. 2007;79:69–82.
66. Sopade PA, Halley PJ, Cichero JAY, Ward LC, Liu J, Teo KH. Rheological characterization of food thickeners marketed in Australia in various media for the management of dysphagia. II: Milk as a dispersing medium. J Food Eng. 2008;84:552–62.
67. Sopade PA, Halley PJ, Cichero JAY, Ward LC, Liu J, Varliveli S. Rheological characterization of food thickeners marketed in Australia in various media for the management of dysphagia. III: Fruit juice as a dispersing medium. J Food Eng. 2008;86:604–15.
68. Hadde EK, Nicholson TM, Cichero JAY. Rheological characterisation of thickened milk components (protein, lactose and minerals). J Food Eng. 2015;166:263–7.
69. Hadde EK, Nicholson TM, Cichero JAY. Rheological characterisation of thickened fluids under different temperature, pH and fat contents. Nutr Food Sci. 2015;45:270–85.
70. Vickers Z, Damodhar H, Grummer C, Mendenhall H, Banaszynski K, Hartel R, Hind J, Joyce A, Kaufman A, Robbins J. Relationships among rheological, sensory texture, and swallowing pressure measurements of hydrocolloid-thickened fluids. Dysphagia. 2015;30:702–13.
71. Funami T, Ishihara S, Nakauma M, Kohyama K, Nishinari K. Texture design for products using food hydrocolloids. Food Hydrocoll. 2012;26:412–20.
72. Ashida I, Iwamori H, Kawakami SY, Miyaoka M, Murayama A. Analysis of physiological parameters of masseter muscle activity during chewing of agars in healthy young males. J Texture Stud. 2007;38:87–99.
73. Garcia JM, Chambers E, Matta Z, Clark M. Viscosity measurements of nectar- and honey-thick liquids: product, liquid, and time comparisons. Dysphagia. 2005;20:325–35.
74. O'Leary M, Hanson B, Smith C. Viscosity and non-Newtonian features of thickened fluids used for dysphagia therapy. J Food Sci. 2010;75:E330–8.
75. Mackley MR, Tock C, Anthony R, Butler SA, Chapman G, Vadillo DC. The rheology and processing behavior of starch and gum-based dysphagia thickeners. J Rheol. 2013;57(6):1533–53.
76. Skriver A, Holstbord J, Qvist KB. Relation between sensory texture analysis and rheological properties of stirred yogurt. J Dairy Res. 1999;66:609–18.
77. Kutter A, Singh JP, Rauh C, Delgado A. Improvement of the prediction of mouthfeel attributes of liquid foods by a posthumus funnel. Journal of Texture Studies. 2011;42:217–27.
78. Nystrom M, Qazi WM, Bulow M, Ekberg O, Stading M. Effects of rheological factors on perceived ease of swallowing. Applied Rheology. 2015;25:63876.
79. Rosenthal A, Chen J. Food texture and structure. In: Rosenthal A, Chen J, editors. Modifying food texture. Amsterdam: Elsevier; 2015. p. 3–24.
80. Moberg T, Rigdahl M, Stading M, Levenstam Bragd E. Extensional viscosity of microfibrillated cellulose suspensions. Carbohydr Polym. 2014;102:409–12.

Julie A. Y. Cichero Ph.D.

Peter Lam R.D., C.F.E.

Catriona M. Steele Ph.D.

Ben Hanson Ph.D.

Jianshe Chen Ph.D.

Roberto O. Dantas Ph.D.

Janice Duivestein M.R.Sc., OT/PT.

Jun Kayashita Ph.D.

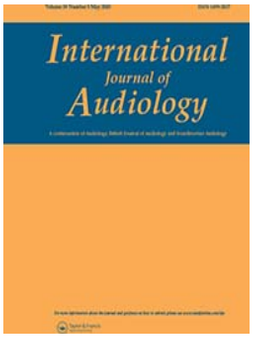
Caroline Lecko R.N.

Joseph Murray Ph.D.

Mershen Pillay D.Ed.

Luis Riquelme Ph.D.

Soenke Stanschus M.A.



International Journal of Audiology

ISSN: 1499-2027 (Print) 1708-8186 (Online) Journal homepage: https://www.tandfonline.com/loi/iija20

Mental tasking and rotary Chair-Induced vestibular nystagmus utilizing Video-Oculography

Steven M. Doettl, Mary K. Easterday, Patrick N. Plyler, Lacey L. Behn & Allison S. Poget

To cite this article: Steven M. Doettl, Mary K. Easterday, Patrick N. Plyler, Lacey L. Behn & Allison S. Poget (2020) Mental tasking and rotary Chair-Induced vestibular nystagmus utilizing Video-Oculography, International Journal of Audiology, 59:5, 360-366, DOI: 10.1080/14992027.2019.1706768

To link to this article: https://doi.org/10.1080/14992027.2019.1706768

Published online: 26 Dec 2019.

Submit your article to this journal

Article views: 235

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=iija20

INTERNATIONAL JOURNAL OF AUDIOLOGY
2020, VOL. 59, NO. 5, 360–366
https://doi.org/10.1080/14992027.2019.1706768

The British Society of Audiology · The International Society of Audiology · NAS Nordic Audiological Society · Taylor & Francis Taylor & Francis Group

ORIGINAL ARTICLE

Check for updates

Mental tasking and rotary Chair-Induced vestibular nystagmus utilizing Video-Oculography

Steven M. Doettl, Mary K. Easterday, Patrick N. Plyler, Lacey L. Behn and Allison S. Poget

Department of Audiology and Speech Pathology, University of Tennessee Health Science Center, Knoxville, TN, USA

ABSTRACT

Objective: To investigate whether the use of mental tasking, when compared to no mental task, affects measurement of nystagmus response with regard to gain, phase & symmetry, and artefact when utilising video-oculography (VOG) as the measurement technique in rotary chair testing (RCT).

Design: A within-subject repeated-measures design was utilised.

Study samples: Seventeen (17) healthy adults were evaluated (age 22–25 years). Each participant underwent slow harmonic acceleration (SHA) testing for 0.01, 0.02, 0.04, 0.08, and 0.16 Hz using RCT at two separate counterbalanced visits. At one visit mental tasking was utilised while the other visit did not utilise mental tasking. The following outcomes were measured for each visit: gain, phase, symmetry, and artefact.

Results: No significant difference between the tasking conditions with regard to gain, phase, symmetry, or artefact. Significant frequency affects were noted, as expected, for gain, phase, and artefact. Analysis of individual subject data did, however, describe significant effects of tasking with regard to gain, phase, symmetry, and artefact.

Conclusion: These results suggest that the use of mental tasking during RCT using VOG had no significant group effect on SHA gain, phase, symmetry, and artefact. However, individual subject effects were observed indicating variability in the effects of mental tasking during RCT.

ARTICLE HISTORY
Received 4 March 2019
Revised 11 December 2019
Accepted 15 December 2019

KEYWORDS
Vestibular; mental tasking; rotary chair; video-oculography; vestibulo-ocular reflex

Introduction

Measurement of the vestibulo-ocular reflex (VOR) is a foundation of several vestibular assessment techniques including caloric irrigation testing, rotary chair testing (RCT), and video head impulse testing (vHIT). Several studies have investigated the effects of mental arousal and/or mental tasking activities indicating that mental tasking is imperative for accurate assessment of the VOR using electro-oculography (EOG) (Mahoney, Harlan, and Bickford 1957; Barber and Wright 1967; Barr, Schultheis, and Robinson 1976; Collins 1962a, 1962b; Collins et al. 1960; Formby et al. 1992; Jacobson, Newman, and Kartush 1997; Kileny, McCabe, and Ryu 1980; Tjernstrom 1973; Yamanobe et al. 1967).

Foundational studies described experimentation with normal, healthy subjects with reverie and non-attentiveness resulting in significantly less recorded nystagmus than when mental alerting tasks were used (Collins et al. 1960; Collins 1962a, 1962b). Kileny, McCabe, and Ryu (1980) and Formby et al. (1992) also suggested that mental tasking resulted in increased VOR gain and reduced variability. These studies as well as several others consistently noted increased vestibular responses, decreased variability within subjects over time and between subjects, and increased overall quality of the nystagmus when mental tasking was instituted (Mahoney, Harlan, and Bickford 1957; Barber and Wright 1967; Barr, Schultheis, and Robinson 1976; Collins 1962b; Jacobson, Newman, and Kartush 1997; Matta and Enticott 2004; Tjernstrom 1973; Yamanobe et al. 1967).

However, recent changes in vestibular testing measurement techniques suggest the need to evaluate the use of mental tasking using the most current evaluation techniques. EOG is a measurement technique that relies on the opposing electrical charges of the cornea and the retina (Wallace and Norris 1966). This method utilises electrodes placed around the eyes and nose on the face in order to compare the movements of the eye to the ground electrode in the experimental setup with tracings representing deviations from the baseline behaviour represented through computer software (Ganança, Caovilla, and Ganança 2010; Wallace and Norris 1966). EOG testing is completed in a darkened room in order to prevent visual suppression of the nystagmus which is most often achieved by closing the eyes (Jacobson, Newman, and Kartush 1997).

Closure of the eyes has been shown to have a significant negative effect on the gain of VOR nystagmus and overall VOR stability (Collins 1962b; Goebel et al. 1983; Mahoney, Harlan, and Bickford 1957; Naito et al. 1963; Tjernstrom 1973; Yamanobe et al. 1967). The physical closing of the eyes during EOG has been shown to result in reduced reliability and even complete inhibition of the VOR with the eyes closed (Mahoney, Harlan, and Bickford 1957). Collins (1962b) specifically described the idea that the combination of EOG and testing with eyes closed was a duplicative suppressive effect on the measurement of the VOR, thus making mental alertness a critical tool. Additionally, Naito et al. (1963), Tjernstrom (1973), and Yamanobe et al. (1967) also agreed that eye closure results in

CONTACT Steven M. Doettl sdoettl@uthsc.edu Department of Audiology and Speech Pathology, University of Tennessee Health Science Center, 355 South Stadium Hall, Knoxville, TN 37996-0740, USA

This article has been republished with minor changes. These changes do not impact the academic content of the article.
© 2019 British Society of Audiology, International Society of Audiology, and Nordic Audiological Society

mechanical suppression of the VOR related to the muscle contraction from the eyelids.

Video-oculography (VOG) is an alternative method to EOG. As opposed to EOG, VOG is completed with the eyes open removing mechanical suppression of the VOR as a factor in measurement. With VOG, instead of using electrodes, infra-red goggles are utilised to fixate the light on the pupil in order to measure and record the eye movements with computer software to generate a tracing (Ganança, Caovilla, and Ganança 2010). When VOG is used to assess VOR function, the eyes must be open to allow for tracking of the movements. Therefore, the goggles are covered to prevent visual suppression (Ganança, Caovilla, and Ganança 2010; Jacobson, Newman, and Kartush 1997).

There is conflicting data regarding the use of mental tasking when evaluating the VOR with the eyes open. Naito et al. (1963) evaluated 30 subjects with vertigo and 10 normal healthy controls using RCT with EOG across 4 conditions (eyes open in complete darkness with and without tasking and eyes closed with and without tasking) and noted no significant tasking effect in the eyes open in complete darkness recordings. Additionally, Torok (1970) evaluated 7 normal, healthy subjects in a light-proof room with eyes open in complete darkness using both caloric irrigation and RCT and found no significant difference on VOR measures between the tasking (mental arithmetic) and no tasking conditions. Collins (1962a and 1962b) and Matta and Enticott (2004) on the other hand evaluated normal subjects for both caloric irrigation and RCT using EOG in the dark with eyes open and reported a significantly increased VOR response with mental tasking.

Jacobson et al. (2012) analysed RCT results of normal young healthy participants using VOG for multiple types of suppression (mental tasking, auditory, somatosensory, visual and imaginary visual). The results indicated no significant difference in VOR gain values between the tasking and no tasking conditions. Easterday, Plyler, and Doettl (2016) evaluated 16 normal young healthy adults (ages 19–31 years) using bithermal caloric irrigation with VOG using separate counterbalanced visits (1 with mental tasking, 1 without mental tasking). The results indicated no significant difference for tasking versus no tasking with peak slow phase velocity (SPV), peak latency, or response duration. However, their results did indicate a significant increase in the amount of eye blink artefact for the tasking condition suggesting a possible deleterious effect.

Further evaluation is needed as additional questions exist which have not been previously addressed in the other studies. Torok (1970) and Jacobson et al. (2012) reported no significant change in VOR gain with RCT, but with no description of these effects on phase, symmetry, or artefact or gain across a large number of stimulation frequencies. Symmetry, which is a measure based on gain values, also can be representative of variability of the response and is an important and separate measure from gain for clinical evaluation. Phase is also a separate measure from gain, which describes the timing relationship of the VOR response as opposed to the strength which inherently may be impacted differently by mental tasking. Additionally, evaluating the effect of mental tasking during RCT using a wider range and number of stimulation frequencies will also provide an increase in the knowledge base of these effects with regard to frequency as well as guidance for clinic measurement. Finally, previous results from our lab (Easterday, Plyler, and Doettl, 2016) reported a significant increase in artefact when using mental tasking during caloric testing. However, no studies have investigated the instances of artefact during RCT using VOG under tasking and no tasking conditions. Based on the previous works and the need for further investigation to document tasking effects in RCT using VOG, the current study aimed to assess the effect of mental tasking on RCT, specifically the sinusoidal harmonic acceleration (SHA) protocol for gain, phase, symmetry, and artefact using the VOG measurement technique across a wide range of stimulation frequencies.

Methods

Participants

An a priori power analysis using G-Power 3.1.9.2 was completed using a repeated measures analysis of variance (ANOVA) design, assuming individual dependent variables, an alpha level of 0.05, and a large effect size ($f = 0.40$) showed that a total sample size of 15 was sufficient for a power of at least 0.80 (actual power = 0.8213105). In the absence of previous data, a large effect size ($f = 0.40$) was assumed since small or medium effects would not be relevant clinically. A total of 17 participants ages 22–25 years (mean 23.9, *SD* 1.14) were evaluated. Each participant served as his/her own control. Thus, the sample size in the current study was likely sufficiently large to find a significant effect if a large difference between the tasking conditions existed. All of the participants were female and denied any history of vestibular dysfunction, oculomotor dysfunction, musculoskeletal dysfunction, neurologic disorders, ingestion of alcohol within 24 h of the evaluation, use of prescribed or illicit medications known to have an effect on the VOR response using the University of Tennessee Hearing and Speech Centre Dizziness Clinic Adult Case History questionnaire. Each participant passed a 20 dB HL hearing screening from 250 Hz and 8000 Hz. Immittance measures were also completed to rule-out any pre-existing middle ear pathology. This study was approved by the Institutional Review Board at the University of Tennessee Health Science Centre, and all participants signed an informed consent before participation in this study.

Protocol

RCT was completed at the University of Tennessee Health Science Centre, Department of Audiology and Speech Pathology Vestibular and Balance Laboratory. RCT was performed using Interacoustics NyDiag 200 and standard SHA test protocol. Calibration was completed prior to SHA testing for all participants using the standard fixed saccade VOG protocol. Binocular VOG recording with the eyes covered in complete darkness using a stimulation rate of 50 degrees per second for 0.01 (2 cycles), 0.02 (2 cycles), 0.04 (2 cycles), 0.08 (4 cycles), and 0.16 (6 cycles) Hz was completed for each participant. The testing room was also darkened to limit the possibility of light infiltration in the VOG goggles and participants were instructed to confirm complete darkness within the VOG goggles. Standard normative values from the Interacoustics NyDiag 200 software were used (Table 1). The normative range values for gain were: 0.01 Hz (18–65), 0.02 Hz (25–79), 0.04 Hz (31–88), 0.08 Hz (30–86), 0.16 Hz (27–86). The normative range values for phase were: 0.01 Hz (27–59), 0.02 Hz (12–39), 0.04 Hz (4–21), 0.08 Hz (−2 to 13), 0.16 Hz (−10 to 9). The normative range values for symmetry were: 0.01 Hz (−34 to 30), 0.02 Hz (−34 to 26), 0.04 Hz (−32 to 25), 0.08 Hz (−31 to 26), 0.16 Hz (−32 to 26). The order of rotation frequencies was followed in the specified

order of 0.08, 0.04, 0.16, 0.02, and 0.01 Hz. Additional frequencies, 0.32 Hz and 0.64 Hz, were not included as the equipment did not have standardised normative values for 0.32 Hz or the capability of 0.64 Hz. The testing frequencies were not randomised. Standard protocol was followed with respect to the order of testing frequencies in order to vary from fast to slow to fast to slow and ending with the most difficult frequency to tolerate.

Each participant completed 2 visits, one with tasking and one without tasking with the order randomised, each separated by at least a full 24-h day. The initial visit included the hearing screening and informed consent in addition to the experimental condition. Caution was also taken as to not reveal the study hypothesis to the participants during the initial visits to control for any possible information bias prior to the experimental procedure. The hearing screen occurred in a sound-treated booth and pure-tone stimuli were delivered through ER-3A insert earphones with the appropriate sized ER3-14 disposable foam eartips utilising a GSI 61 audiometer (Grason-Stadler, Eden Prairie, MN). Tasking was conducted to be consistent with the Easterday, Plyler, and Doettl (2016) study using the Formby et al. (1992) method of naming categorical items while undergoing minimal interaction with the examiner. Categories included animals, vegetables, states, and furniture in randomised order for each participant. For the no-tasking condition the participant was instructed to keep the eyes open during the testing procedure and re-instructed as needed.

SHA results were evaluated for VOR gain defined as the ratio of the actual experimental movement of the eye over the anticipated or desired movement (Jacobson, Newman, and Kartush 1997). SHA phase is defined as sinusoidal relationship between the eye movement and head movement during and after rotation with respect to time (Martin and Clark 2015). SHA gain symmetry is also typically measured and is a calculation performed to formulate a comparison between the response gain following rightward rotation and leftward rotation or more specifically the percentage of difference between the peak slow component eye velocity produced by the left versus the right directional movements (Jacobson, Newman, and Kartush 1997; Jacobson and Shepard 2016). Artefact during RCT can include blinking, eye closure, etc. (Jacobson, Newman, and Kartush 1997; Martin and Clark 2015). All tracings were then re-evaluated for artefact including eye blinks, eye closure, and other poor morphology with the number of instances of artefact recorded for both tasking conditions and all SHA frequencies and instances of artefact were removed from the tracings.

Statistical analyses

All statistical analyses were completed using IBM SPSS Statistics 25. A two-way repeated measures multivariate analysis of variance (MANOVA) was conducted to evaluate the effects of tasking and SHA frequency on the measured RCT values of gain, phase, symmetry, and the amount of artefact. The independent variables were tasking condition with two values (tasking vs. no tasking) and SHA frequency (0.01, 0.02, 0.04, 0.08, 0.16 Hz). The dependent variables were gain, phase, symmetry, and artefact. Additional follow-up analyses of variance (ANOVA) were then conducted on any significant main effects or interactions noted. As a secondary check of significance, statistical analyses were also completed using separate repeated measures ANOVA collapsed across frequency for the tasking condition on each of the dependent variables of gain, phase, symmetry, and artefact to ensure that any significance that would be present would not be marginalised due to the multiple comparisons in the MANOVA.

Additional post-hoc analysis of individual participant SHA data was also completed to identify significant changes in a participant's VOR gain, phase, symmetry, and artefact (averaged across frequencies) between the non-tasking and tasking conditions. A 95% confidence interval of the SHA frequencies (i.e. 95% critical difference) was calculated for each participant. VOR gain values were collapsed across SHA frequency to establish the baseline value for the no tasking condition and a tasking value. The "tasking effect" was calculated for each measurement by subtracting the no-tasking value from the tasking value for each participant. Differences between the conditions were considered significant if the tasking condition values exceeded the upper or lower baseline critical difference (CD). VOR gain values exceeding the lower CD were considered evidence of gain suppression while values exceeding the upper CD were considered evidence of gain enhancement.

Results

Overall no participants exhibited any patterns of significant abnormalities on RCT findings using the standard normative values (see Table 1). Results for gain, phase, symmetry, and artefact were then averaged across the participants for each tasking condition and SHA frequency (Figures 1–4). Gain, phase, symmetry, and artefact from the MANOVA revealed a significant effect for frequency (p-value 0.043) with no significant effect for tasking condition (p-value 0.096) or frequency-tasking condition (p-value 0.0987) interaction (see Table 2). The follow-up ANOVAs revealed a significant frequency effect for gain (p-value <0.001), phase (p-value <0.001), and artefact (p-value <0.001) with no significant effect for symmetry (p-value 0.209) (see Table 3). It is expected that both gain and phase should have a relationship with SHA frequency as this is well-documented in the literature and represented in the RCT normative data. Additionally, an inverse relationship was noted for artefact and frequency with

Table 1. Normative data for RCT gain, phase, and symmetry from the interacoustics NyDiag 200 software.

	0.01 Hz		0.02 Hz		0.04 Hz		0.08 Hz		0.16 Hz	
Frequency	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Gain	18%	65%	25%	79%	31%	88%	30%	86%	27%	86%
Phase	27°	59°	12°	39°	4°	21°	−2°	13°	−10°	9°
Symmetry	−34%	30%	−34%	26%	−32%	25%	−31%	26%	−32%	26%

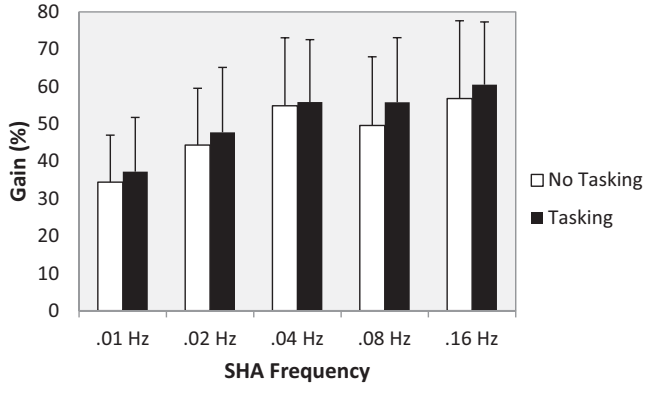


Figure 1. SHA gain values across the SHA frequencies. Standard deviation bars are shown.

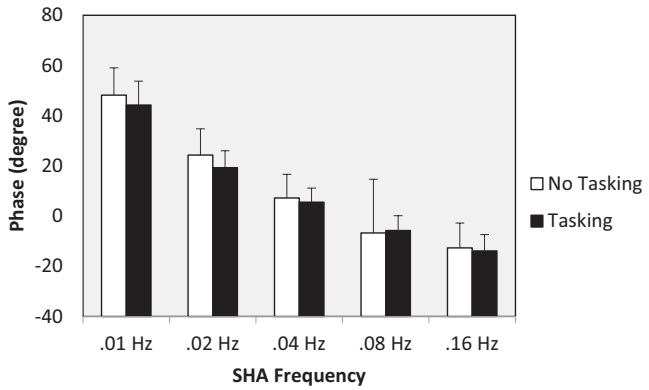


Figure 2. SHA phase values across the SHA frequencies. Standard deviation bars are shown.

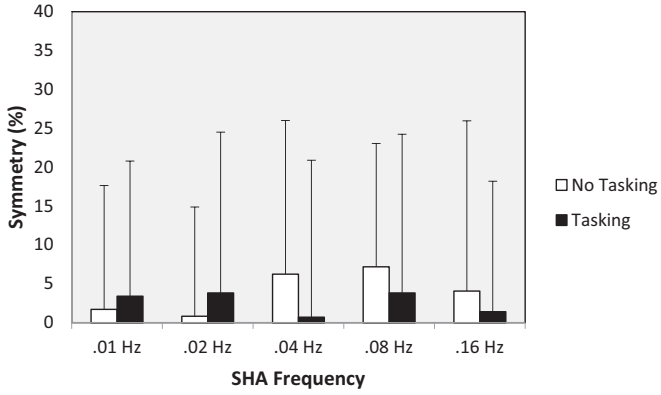


Figure 3. SHA symmetry values across the SHA frequencies. Standard deviation bars are shown.

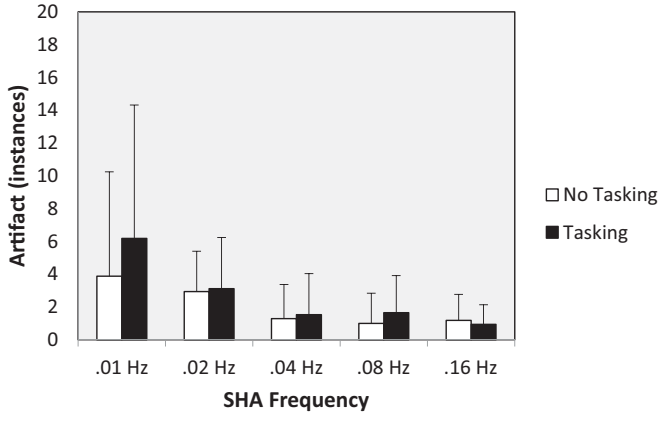


Figure 4. SHA artefact values across the SHA frequencies. Standard deviation bars are shown.

increased artefact for lower frequencies. This inverse relationship is also expected as the testing time is longer as the frequency gets lower. Also, independent repeated measures ANOVA for each SHA value and artefact did not indicate any significant difference (gain p-value 0.147, phase p-value 0.249, symmetry p-value 0.070, artefact p-value 0.519) between the tasking and no tasking conditions.

Although group data did not reveal a significant tasking effect, analysis of individual data indicated tasking did impact some participants (Figure 5). Examination of this data indicated that for gain, participant 3 and participant 10 exhibited an increase in gain in the tasking condition. In contrast participant 6 and participant 12 also exhibited a significant decrease in gain in the tasking condition. For the phase data only participant 3 exhibited a significant change in phase with an increase in phase lag. Examination of the symmetry data indicated that participants 5, 7, 10, 14, 15, 16, and 17 all had significant changes in symmetry values; however, only participant 7 had a significant change in symmetry which resulted in a change in direction. Finally, for artefact participants 1, 2, 5, 10, 11, 12, and 13 all had a significant increase in artefact with participants 15 and 16 exhibiting a decrease in artefact in the tasking condition.

Table 2. Two-way multivariate analysis of variance for RCT measurements for tasking condition and SHA frequency.

	F	df	p	partial η^2	Ω
Frequency	331.191	1	0.043*	1.000	0.756
Tasking Condition	2.476	13	0.096	0.432	0.539
Frequency x Tasking Condition	4.843	1	0.344	0.987	0.122

*$p < 0.05$

Table 3. Analyses of variance results for RCT measures for frequency.

	F	df	p	partial η^2	Ω
Frequency					
Gain	31.022	4	<0.001*	0.660	1.000
Phase	367.763	4	<0.001*	0.958	1.000
Symmetry	1.512	4	0.209	0.086	0.442
Artefact	13.114	4	<0.001*	0.450	1.000

*$p < .05$

Discussion

The purpose of this study was to further evaluate the impact of mental tasking on the VOR responses of phase, gain, symmetry, and artefact using RCT SHA and VOG. The results of this study indicated no significant group effect of mental tasking on gain, phase, symmetry, or artefact. Additionally, the average gain measure (collapsed across frequency) was 48% for the non-tasking and 51% for the tasking condition. Regardless of the p-value obtained the average difference between the conditions of 3% (with a 95% confidence interval of −1.3 to 8%) represent only a very small clinical effect. This study was also unique in that the RCT SHA measures of phase and symmetry were also evaluated for tasking effects. The results of this study did not indicate any significant effect on phase or symmetry values suggesting that mental state may not be a factor in evaluating these measures, at least on a group level. Individual effects on mental tasking were noted with significant variability across participants with regard to the impact of mental tasking, both positive and negative, on the measurementof the VOR.

This study is consistent with the findings of Jacobson et al. (2012) suggesting that SHA gain when measured using VOG is not significantly affected by mental tasking. Additionally, these results exhibited similar findings to Easterday, Plyler, and Doettl (2016) who indicated that mental tasking did not significantly affect the VOR response during caloric irrigation using VOG. Additional previous studies have also shown similar mitigated effects when assessing VOR responses with eyes open (Naito et al. 1963; Torok 1970).

These results and those from previous related studies suggest the effect of mental tasking using VOG may vary from the effect noted using EOG (Collins 1962b; Naito et al. 1963; Torok 1970). Additional studies also describe the significant suppressive effect of eye closure on the VOR which could at least provide some

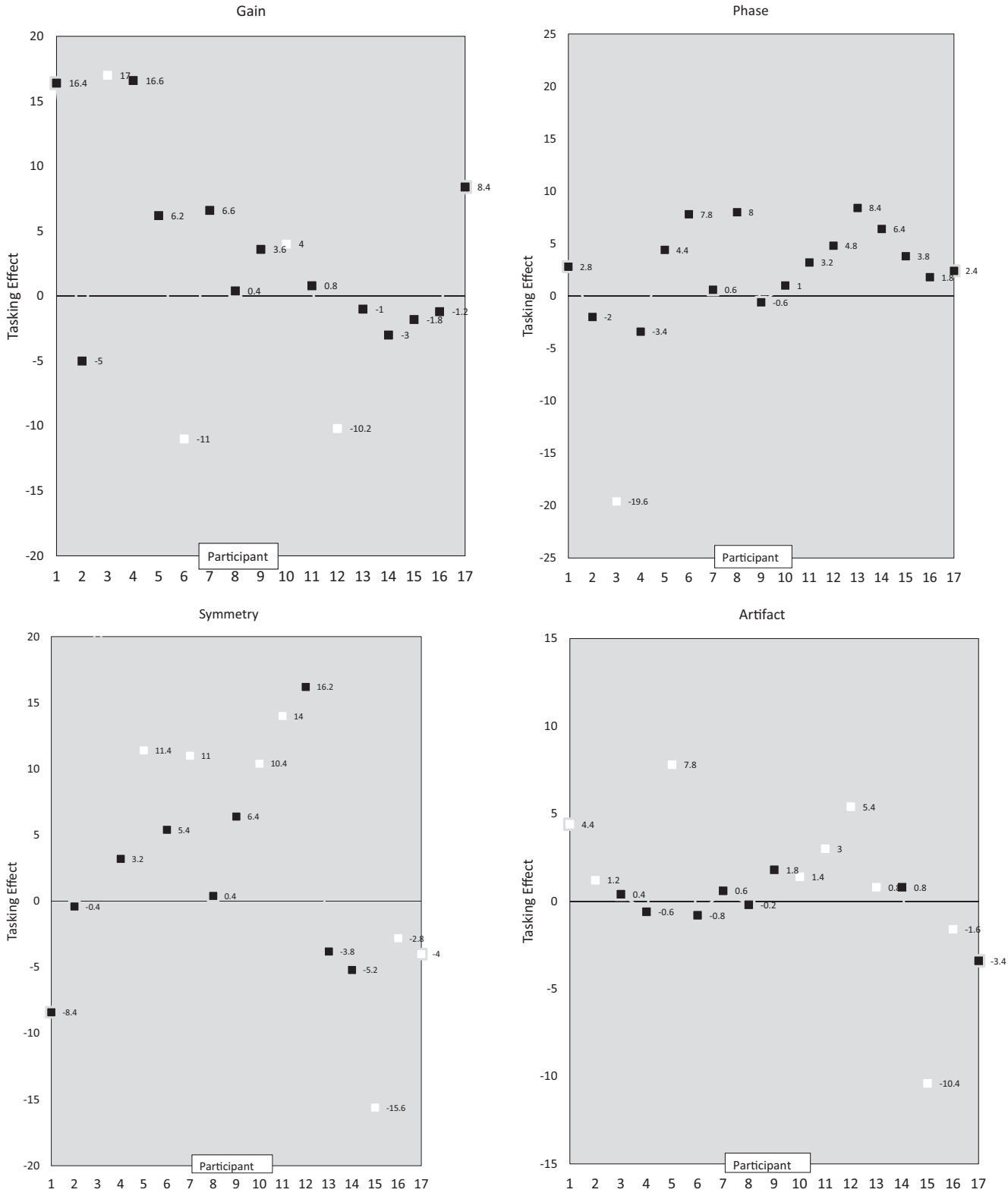


Figure 5. The effect of tasking was calculated for each measurement by subtracting the no-tasking value from the tasking value (No-tasking – tasking) for each participant. 95% confidence intervals were also calculated for each participant to determine if the tasking effect was significant (indicated by the white data points).

explanation of differences reported between tasking effects in EOG versus VOG (Collins 1962b; Goebel et al. 1983; Mahoney, Harlan, and Bickford 1957; Naito et al. 1963; Tjernstrom 1973; Yamanobe et al. 1967). Additionally, Collins (1962b) noted the significance of eye closure even in total darkness reducing the nystagmus response especially when using EOG suggesting that the combination of both EOG and eye closed would have the largest suppressive effects. However, currently there are no studies evaluating the use of mental tasking in the measurement of a VOR response with the specific goal of comparing EOG with eyes open, EOG with eyes closed, and VOG measures making a direct comparison difficult.

While this study did not indicate any significant group effect with mental tasking for SHA values and artefact, individual participant data indicated significant differences for several participants. Individual data revealed a widely variable effect of mental tasking on the SHA values in both directions (i.e. positive and negative). This is also noted in the large standard deviations in the results. Significant individual variability has also been previously reported in the literature for both EOG and VOG studies. However, uniquely this study as well as the previous work, Easterday, Plyler, and Doettl (2016), revealed the possibility of deleterious effects of mental tasking on VOR measures on some participants with increased artefact, decreased gain, and changes in symmetry and phase with tasking on some participants.

One use of mental tasking during VOR evaluation is to provide a control for the mental status of the patient during examination; however, the data from this study suggests the tasking may exist as an additional variable in some people. Interestingly, Jacobson et al. (2012) reported on the significant suppressive effect of auditory stimulation on RCT SHA gain. The mental task used in the current study was categorisation which required some degree of auditory stimulation during the verbal communication between the investigator and the participant. These results are similar to the findings from another recent study from our lab (Easterday et al. 2018) that also evaluated the impact of auditory stimulation on VOR responses. Based on the results of Easterday et al. (2018) and Jacobson et al. (2012) findings it could be postulated that in the cases of noted deleterious effects of mental tasking auditory stimulation have played a role. An excellent follow-up to this study would include the use of mental arithmetic or other alerting tasks that do not involve auditory stimulation.

Although the power analysis indicated adequate power for our sample size (17), these results may not be generalisable to other populations such as the disordered, elderly, or other populations outside the sample size. Additionally, a large effect size was selected during the a priori power analysis to determine sample size: thus, in order to investigate for significance difference using small or medium effect sizes a larger sample size will be needed. All of the participants in this study were female. The possibility of gender effects in RCT SHA measures of gain, phase, and symmetry has been reported, however there is mixed data on the topic with the most recent data suggesting that, at least for SHA, no significant gender effect exists (Wall, Hunt, and Black 1984; Chan et al. 2016; Durney 2016; Li, Hooper, and Cousins 1991; Maes et al. 2008; Peterka, Black, and Schoenhoff 1989; Valente 2007).

Participants in this study also were not screened for the use of nicotine or caffeine prior to their inclusion in the study. Similar to gender, there is debate with regard to nicotine and possible effects on VOR measures (Tibbling and Henriksson 1968; Tibbling 1969). With regard to caffeine, two recent studies suggest minimal to no significant effects (Felipe et al. 2005; McNerney, Coad, and Burkard 2018). Additional limitations include the use of only one type of tasking. Further evaluation using multiple types of tasking is needed. The current study did not address variability of SHA RCT responses over time, the use of mental tasking for serial VOR measurement, SHA VOR suppression testing, or velocity steps protocols and these results should not be used to make inferences on those practices.

Clinical implications

The purpose of this project was to evaluate the effect of mental tasking during RCT SHA using VOG. Results indicated the use of a categorisation alerting task did not have a significant overall effect on gain, phase, symmetry or artefact. Analysis of individual data revealed that tasking had no significant effect for the majority of participants, however in certain cases the tasking did have a significant effect on SHA values. Specifically, the clinician needs to be aware that in some cases mental tasking can also have a deleterious effect on VOR measures when using VOG. Further study is, of course, needed specifically including variability of the VOR response over time, different alerting tasks, velocity steps testing, with a wider age range, and with disordered populations to further evaluate this topic.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Barber, H. O., and G. Wright. 1967. "Release of Nystagmus Suppression in Clinical Electronystagmography." *The Laryngoscope* 77 (6): 1016–1027. doi:10.1288/00005537-196706000-00008

Barr, C. C., W. Schultheis, and D. A. Robinson. 1976. "Voluntary, Non-Visual Control of the Human Vestibulo-Ocular Reflex." Acta Oto-Laryngologica 81 (3–6): 365–375. doi:10.3109/00016487609119974

Chan, F. M., J. Galatioto, M. Amato, and A. H. Kim. 2016. "Normative Data for Rotational Chair Stratified by Age." *The Laryngoscope* 126 (2): 460–463. doi:10.1002/lary.25497

Collins, W. E. 1962a. "Manipulation of Arousal and Its Effects on Human Vestibular Nystagmus Induced by Caloric Irrigation and Angular Accelerations." *Civil Aeromedical Research Institute*.

Collins, W. E. 1962b. "Effects of Mental State upon Vestibular Nystagmus." *Journal of Experimental Psychology* 63 (2): 191–197. doi:10.1037/h0039941

Collins, W. E., G. Crampton, J. Posner, and S. Warburton. 1960. "The Effect of Mental Setupon Vestibular Nystagmus and the Electroencephalogram." *US Army Medical Research Laboratory* 12: 1–15.

Durney, T., Suppression Subtests of Rotational Tetsing for the Towson University Hearing and Balance Center 2016. Normative Data for the Sinusoidal Harmonic Acceleration and Visual. Retrieved from the Maryland Shared Open Access Repository Home. http://hdl.handle.net/11603/2033.

Easterday, M. K., P. N. Plyler, and S. M. Doettl. 2016. "Mental Tasking and Caloric-Induced Vestibular Nystagmus Utilizing Videonystagmography." *American Journal of Audiology* 25 (3): 177–183. doi:10.1044/2016_AJA-15-0070

Easterday, M. K., P. N. Plyler, J. D. Lewis, and S. M. Doettl. 2018. "Effect of Intensity and Speech Stimulus Type on the Vestibulo-Ocular Reflex." *American Journal of Audiology* 30 (9): 792–801. doi:10.3766/jaaa.18016.

Felipe, L., L. C. Simões, D. U. Gonçalves, and P. C. Mancini. 2005. "Evaluation of the Caffeine Effect in the Vestibular Test." *Revista Brasileira de Otorrinolaringologia* 71 (6): 758–762. doi:10.1590/S0034-72992005000600012

Formby, C., L. A. Kuntz, I. M. Rivera-Taylor, N. Rivera-Mraz, D. R. Weesner, N. E. Butler-Young, and A. E. Ahlers. 1992. "Measurement, Analysis, and Modelling of the Caloric Response." Acta Oto-Laryngologica 498: 19–29. doi:10.3109/00016489209136855.

Ganança, M. M., H. H. Caovilla, and F. F. Ganança. 2010. "Electronystagmography versus Videonystagmography." *Brazilian Journal of Otorhinolaryngology* 76 (3): 399–403. doi:10.1590/S1808-86942010000300021

Goebel, J., M. Stroud, L. Levine, and H. Muntz. 1983. "Vertical Eye Deviation and Nystagmus Inhibition during Mental Tasking." The Laryngoscope 93 (9): 1127–1132.

Jacobson, G. P., C. W. Newman, and J. M. Kartush. 1997. *Handbook of Balance Function Testing*. San Diego: Singular Pub. Group.

Jacobson, G. P., E. G. Piker, C. Do, D. J. McCaslin, and L. Hood. 2012. "Suppression of the Vestibulo-Ocular Reflex Using Visual and Nonvisual Stimuli." *American Journal of Audiology* 21 (2): 226–231. doi:10.1044/1059-0889(2012/12-0021).

Jacobson, G. P., and N. T. Shepard. 2016. *Balance Function Assessment & Management*. 2nd ed. San Diego: Plural Publishing.

Kileny, P., B. F. McCabe, and J. H. Ryu. 1980. "Effects of Attention-Requiring Tasks on Vestibular Nystagmus." Annals of Otology, Rhinology *and* Laryngology 89: 9–12. doi:10.1177/000348948008900104.

Li, C. W., R. E. Hooper, and V. C. Cousins. 1991. "Sinusoidal Harmonic Acceleration Testing in Normal Humans." *The Laryngoscope* 101 (2): 192–196.

Maes, L., I. Dhooge, E. De Vel, W. D'haenens, A. Bockstael, H. Keppler, B. Philips, F. Swinnen, and B. M. Vinck. 2008. "Normative Data and Test-Retest Reliability of the Sinusoidal Harmonic Acceleration Test, Pseudorandom Rotation Test and Velocity Step Test." *Journal of Vestibular Research* 18 (4): 197–208.

Mahoney, J., W. Harlan, and R. Bickford. 1957. "Visual and Other Factors Influencing Caloric Nystagmus in Normal Subjects." Archives of Otolaryngology - Head and Neck Surgery 66 (1): 46–53. doi:10.1001/archotol.1957.03830250050005

Martin, F. N., and J. G. Clark. 2015. *Introduction to Audiology*. 12th ed. Harlow: Pearson Education.

Matta, F. V., and J. C. Enticott. 2004. "The Effects of State of Alertness on the Vestibulo-Ocular Reflex in Normal Subjects Using the Vestibular Rotational Chair." *Journal of Vestibular Research* 14 (5): 387–391.

McNerney, K. M., M. L. Coad, and R. Burkard. 2018. "The Influence of Caffeine on Rotary Chair and Oculomotor Testing." *Journal of the American Academy of Audiology* 29 (7): 587–595. doi:10.3766/jaaa.16118

Naito, T., T. Tatsumi, T. Matsunaga, and T. Matsunaga. 1963. "The Effect of eye-closure upon nystagmus." *Acta Oto-Laryngologica. Supplementum* 179: 72–85. doi:10.3109/00016486309138620.

Peterka, R. J., F. O. Black, and M. B. Schoenhoff. 1989. "Age-Related Changes in Human Vestibule-Ocular Reflexes: sinusoidal Rotation and Caloric Tests." *Journal of Vestibular Research* 1 (1): 49–59.

Tibbling, L. 1969. "The Influence of Tobacco Smoking, Nicotine, CO and CO2 on Vestibular Nystagmus." *Acta Oto-Laryngologica* 68 (1–6): 118–126. doi:10.3109/00016486909121549

Tibbling, L., and N. G. Henriksson. 1968. "Effect of Cigarette Smoking on the Vestibular Nystagmus Pattern." *Acta Oto-Laryngologica* 65 (1–6): 518–526. doi:10.3109/00016486809120994

Tjernstrom, O. 1973. "Nystagmus Inhibition as an Effect of Eye-Closure." Acta Oto-Laryngologica 75 (2–6): 408–418. doi:10.3109/00016487309139767

Torok, N. 1970. "The Effects of Arousal upon Vestibular Nystagmus." *Advanced Oto-Rhino-Laryngology* 17: 76–89. doi:10.1159/000385371

Valente, M. 2007. "Maturational Effects of the Vestibular System: A Study of Rotary Chair, Computerized Dynamic Posturography, and Vestibular Evoked Myogenic Potentials with Children." *Journal of the American Academy of Audiology* 18 (6): 461–481. doi:10.3766/jaaa.18.6.2

Wall, C., A. E. Hunt, and F. O. Black. 1984. "Effects of Age, Sex and Stimulus Parameters upon Vestibulo-Ocular Responses to Sinusoidal Rotation." *Acta Oto-Laryngologica* 98 (3–4): 270–278. doi:10.3109/00016488409107563

Wallace, R., and C. Norris. 1966. "Electronystagmography (ENG)—Technique and Significance." *The Laryngoscope* 76.5: 961–969.

Yamanobe, H., S. Saito, R. Kumata, and S. Tsukahara. 1967. "The Effect of Eyelid Closure on Vestibular Nystagmus." *Fukushima Journal of Medical Science* 22: 3–4.

AJA

Research Article

Suppression of the Vestibulo-Ocular Reflex Using Visual and Nonvisual Stimuli

Gary P. Jacobson,[a] Erin G. Piker,[a] Christina Do,[b] Devin L. McCaslin,[a] and Linda Hood[a]

Purpose: To determine to what extent attention directed toward visual, auditory, somesthetic, and imaginary sources would attenuate the vestibulo-ocular reflex (VOR).
Method: Two prospective studies included 16 (Investigation 1) and 5 (Investigation 2) healthy participants (mean age of 24 years in Investigation 1 and 37 years in Investigation 2). VOR gain was assessed with a commercially available rotary chair and was measured in dark both while the subject was tasked with mental alerting exercises and while not being tasked. VOR suppression was measured for the following conditions: (a) visual suppression, (b) auditory suppression, (c) somatosensory suppression, (d) imaginary visual target suppression, and (e) combined auditory and somatosensory suppression.
Results: Attention directed to visual source attenuated the VOR by approximately 85%. Attention directed toward auditory and somatosensory targets (both separately and combined) and attention directed toward an imaginary target suppressed the VOR between 28% and 44%. The extent of VOR suppression that occurred with attention directed toward various nonvisual stimuli was significantly less than the visual suppression of the VOR. The various nonvisual conditions were not statistically different from one another.
Conclusion: The data suggest that it is possible for typical adults to suppress the VOR in the absence of a visual target. That is, the VOR can be attenuated with attention directed toward chair-fixed visual, auditory, somatosensory, and imaginary targets.

Key Words: audiology, balance, physiology, vestibular rehabilitation

The ability to maintain balance, posture, and orientation makes it possible for us to perform automatic tasks, such as standing upright or walking, and more complex activities, such as sports. Inputs from vision, proprioceptive, and vestibular systems are integrated by the central nervous system, which initiates appropriate motor outputs that enable us to have stable balance.

The peripheral vestibular system routes electrical signals to the central vestibular system and from there to the ocular motor pathways to produce compensatory eye movements during movement of the head or head and body together. It is this vestibulo-ocular reflex (VOR) that stabilizes images of interest on the retina during movement of the head. However, in some instances, it is necessary to "suppress" the VOR. For example, when we are sitting in the grandstands, at the level of the net, watching a tennis match, we do not want our eyes to deviate in the direction opposite the movement of the head as we follow the ball back and forth. In this situation, central processes are recruited to suppress the VOR to permit volitional eye movements to keep an image of interest (e.g., the tennis ball in this case) centered on the fovea of the retina during movement of the head (Konrad, Girardi, & Helfert, 1999). The ability to suppress the VOR also is necessary in the presence of disease when one vestibular end organ is impaired. When this occurs, the tonic electrical asymmetry between the functioning and nonfunctioning end organs produces a spontaneous nystagmus that results in an illusory movement of the environment. In this situation, suppression of the VOR (i.e., by centers in the cerebellar midline) restores a tonic symmetry in the vestibular nuclei and a reduction or elimination of both the spontaneous nystagmus and the vertigo.

The parameter used to quantify VOR suppression is VOR gain. VOR gain is defined as the maximum slow-phase eye velocity (i.e., maxSPEV), divided by the velocity of the head. Under perfect conditions, the VOR produces conjugate eye movements that are equal and opposite the direction of the head movement (i.e., 180° out of phase), resulting in a gain value of 1.0.

VOR suppression is best assessed during rotational testing, which is a "natural" stimulus (i.e., a stimulus that produces excitation of one end organ and simultaneous inhibition of the other). The patient is instructed to stare at an illuminated target that travels with the chair as the chair is oscillated side

[a]Vanderbilt University, Nashville, TN
[b]University of Maryland, College Park
Correspondence to Gary P. Jacobson: gary.jacobson@vanderbilt.edu

Editor and Associate Editor: Larry Humes
Received April 18, 2012
Revision received June 18, 2012
Accepted June 24, 2012
DOI: 10.1044/1059-0889(2012/12-0021)

to side. The results of several studies have shown that VOR suppression is largest in magnitude when it is assessed during rotary chair testing. Visual VOR suppression has been reported to vary between 75% and 90% for rotational testing (Brey, McPherson, & Lynch, 2008) compared with during caloric testing, where VOR suppression has been reported to reduce induced nystagmus by 40%–70% (Barber & Stockwell, 1980). An additional advantage of using rotational testing is that the VOR suppression can be quantified at multiple frequencies.

Although it is well known that connections between the vision, eye movement, and vestibular systems underlie both static and dynamic VOR suppression, a series of investigations conducted over the past 35 years have attempted to assess VOR suppression in the absence of vision using nonvisual targets, including "imagined" visual targets, visual afterimages, somatosensory targets, and auditory targets (Barr, Schultheis, & Robinson, 1976; Furst, Goldberg, & Jenkins, 1987; Jones, Berthoz, & Segal, 1984; Moller, White, & Odkvist, 1990). The results of these investigations suggest that attention directed toward nonvisual targets (e.g., auditory and/or somatosensory targets) may attenuate the VOR. None of these investigations have been replicated. Further, no one has offered an explanation for VOR attenuation where nonvisual targets were used for VOR suppression testing. Accordingly, the objectives of the current investigation were (a) to reassess to what extent attention directed to visual and nonvisual targets can suppress the VOR, (b) to determine how comparable are the values we measure compared with those obtained from previous investigations, and (c) to determine what might be the possible origin of VOR suppression if VOR suppression occurs for nonvisual targets.

Investigation 1

Method

The initial investigation included 16 participants (4 men; mean age = 24 years; range = 21–30 years) with no history of otologic or neurological diseases or disorders. Hearing thresholds at 500, 1000, 2000, and 4000 Hz were obtained on all subjects to ensure normal hearing sensitivity during a screening process. Subjects were recruited from within the Division of Audiology of the Department of Hearing and Speech Sciences at Vanderbilt University. Each subject served as his or her own control. All testing was completed at the Balance and Disorders Center in the Division of Vestibular Sciences at the Vanderbilt Bill Wilkerson Center for Otolaryngology and Communication Sciences. The investigation was approved by the Institutional Review Board at Vanderbilt University (IRB00000477), and subjects provided full informed consent.

Subjects were seated in a commercially available, sinusoidal harmonic acceleration (SHA) chair (Micromedical Technologies System 2000). The chair was enclosed in a lightproof room. SHA testing was conducted at frequencies of 0.02 Hz, 0.08 Hz, and 0.32 Hz (i.e., maximum velocity of 50 degrees per second). Infrared video-oculographic recording techniques were used to record eye position. Only VOR gain will be reported in this investigation, as this is a VOR suppression study, and many times there was insufficient nystagmus to calculate VOR phase and symmetry. VOR gain was quantified as a percentage with a range from 0% to 100% (i.e., gain of 1.0 = 100%). To calculate the magnitude of VOR suppression, we recorded the maxSPEV for each frequency in darkness and then for the various conditions in this investigation. VOR suppression was calculated as [1 – (VOR gain in each of the suppression conditions/VOR gain in darkness)] * 100. Communication between the examiner and subject was accomplished with a wireless intercom system.

Each subject completed five conditions at each frequency (i.e., 0.02 Hz, 0.08 Hz, and 0.32 Hz). All conditions were performed in the dark with the subject's eyes open. The conditions are described in Table 1. For Condition 1, subjects were oscillated in the dark while engaged in mental alerting exercises. Condition 2 was the same as Condition 1 without the mental tasking. The other three conditions were the VOR suppression conditions. During Condition 3, subjects were oscillated in the dark and while in motion asked to stare at a laser target that moved with the chair and was presented at center gaze (visual suppression). For Condition 4, an auditory stimulus was played from a sound source affixed to a vertical metal bar that had been mounted on the footplate of the rotary chair (auditory suppression). The sound source was presented at a comfortable loudness (i.e., 62 dBA) and at 0° azimuth. While the chair was moving, the subject was asked to listen and to stare in the dark at the location of the sound source. For Condition 5, the subject was instructed to grasp a vertical bar mounted on the footplate of the rotary chair (somatosensory suppression). Subjects were further requested to stare at the hand grasping the bar. Data were collected from a minimum of two cycles (i.e., complete chair oscillations) for 0.02 and 0.08 Hz and from a minimum of four cycles for 0.32 Hz. Calibration was completed before each trial. Each condition (i.e., darkness with tasking, darkness without tasking, visual suppression, auditory suppression, somatosensory suppression) was replicated once at each frequency (i.e., 0.02, 0.08, and 0.32 Hz) in a randomized order (Latin square randomization). In all, there were 30 trials for each subject (i.e., 3 frequencies × 5 conditions × 2 runs).

Results

Results are shown in Table 2. VOR gain was greatest for Conditions 1 and 2 (darkness with and without tasking), with a three-frequency mean gain of 47.8 and 44.7, respectively. VOR gain was significantly smaller for all three suppression conditions. VOR gain was similar for Conditions 4 and 5 (auditory and somatosensory suppression), with a three-frequency mean gain of 33.1 and 32.6 for Conditions 4 and 5, respectively. VOR gain was smallest for Condition 3 (visual suppression), which resulted in a three-frequency mean gain of 6.4. In general, results suggested that VOR suppression occurred during all three suppression conditions but was greatest for the visual suppression condition.

The data were analyzed with a repeated measures analysis of variance (ANOVA), where VOR gain (i.e., range of 0–100) served as the dependent variable and the variables run (two levels), frequency (three levels), and condition

Table 1. Description of each of the seven conditions used in Investigations 1 and 2.

Condition	Description
Condition 1: Darkness with mental alerting exercises	Subjects were instructed to keep their eyes open in darkness while engaged in a mental alerting task (e.g., repeating back male and female names beginning with different letters of the alphabet). This test served as a control and as an estimate of the maximum VOR response to the oscillatory stimuli.
Condition 2: Darkness without tasking	Subjects were instructed to keep their eyes open in the dark. This condition served as a control for Conditions 3, 4, and 5.
Condition 3: Visual target	Subjects were instructed to keep their eyes open while staring at a chair-mounted red laser light in the dark (i.e., visual suppression of the VOR).
Condition 4: Auditory target	Subjects were instructed to keep their eyes open while staring at, and listening to, a digital sound recorder (Sony IC Recorder), playing "Moonlight Sonata" by Beethoven, at a level of 62 dBA 0° azimuth in the dark. The recorder was attached to the top of the vertical bar.
Condition 5: Somatosensory target	Subjects were instructed to keep their eyes open in darkness while grasping with both hands and staring at the vertical bar mounted on the footplate of the rotary chair.
Condition 6: Imaginary visual target	Subjects were instructed to keep their eyes open and gaze straight ahead at an imaginary light.
Condition 7: Combined auditory and somatosensory target	Subjects were instructed to keep their eyes open and both listen to the auditory stimulus (same stimulus as in Condition 4) and grasp the vertical bar to which the sound source was affixed (same vertical bar as in Condition 5). Subjects were instructed to direct their attention to the location of their hand and sound source on the vertical bar.

Note. VOR = vestibulo-ocular reflex.

(five levels) served as grouping variables. The results revealed that although VOR gain did not decrease significantly between Runs 1 and 2 ($p = .177$), there were significant main effects noted for the variables frequency, $F(2, 30) = 4.23$, $p = .015$; and condition, $F(4, 60) = 156.42$, $p < .001$.

Post hoc analysis (Bonferroni) showed that VOR suppression was poorer (i.e., the VOR gain was greater across conditions) for 0.02 Hz compared with 0.08 Hz ($p = .02$). Post hoc analysis (Bonferroni) also showed that VOR gains for Conditions 1 and 2 (i.e., controls with Condition 1 and without Condition 2 alerting exercises) were significantly larger ($p < .001$) when compared with VOR gain for Condition 3 (visual suppression, $p < .001$), Condition 4 (auditory suppression, $p = .001$), and Condition 5 (somatosensory suppression, $p < .001$). VOR gain for Conditions 4 and 5 (i.e., auditory suppression and somatosensory suppression) were significantly larger than the VOR gain in Condition 3 (visual suppression, $p < .001$).

The mean gain for each of the five conditions and three frequencies across subjects is shown in Table 2. Across frequencies, attending to an illuminated visual target produced, on average, 86% attenuation of the baseline VOR gain. The VOR across frequencies was reduced 28% and 31% of the baseline by directing attention toward auditory and somatosensory targets, respectively. Figure 1 shows the percent reduction in VOR gain for all five suppression conditions and for Condition 1.

Investigation 2

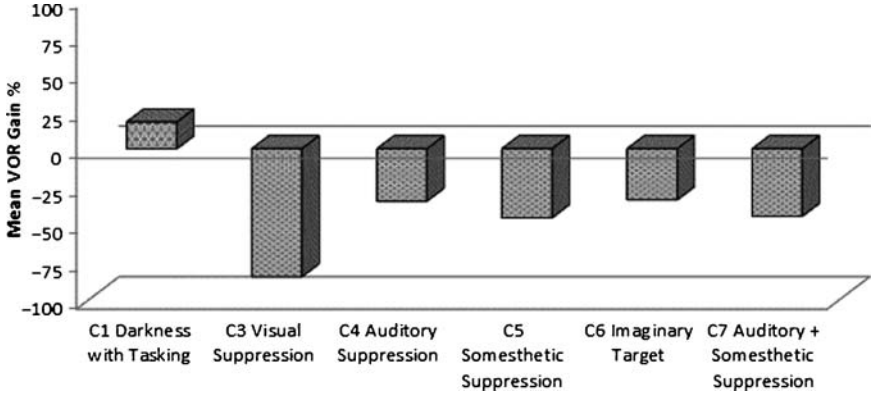
Method

Investigation 2 was a smaller study conducted in an effort to determine whether suppression of the VOR by nonvisual targets is additive. That is, we asked, "If we combine auditory suppression with somatosensory suppression, will the suppression effects be significantly greater than each in isolation?" The recording methods for this investigation were identical to those as in Investigation 1, with the following exceptions: Subjects were 5 adults (2 men, mean age = 37 years)

Table 2. Mean VOR gain (*SD*) for each of the five conditions and three frequencies in Investigation 1.

	Frequency (Hz)			
Condition	0.02 Hz	0.08 Hz	0.32 Hz	Three-frequency mean
1 (Darkness with alerting)	43.1 (12.6)	49.2 (16)	51.2 (18.6)	47.8 (15.9)
2 (Darkness without alerting)	39.8 (8.9)	48.7 (11.3)	45.8 (14.7)	44.7 (12.2)
3 (Visual suppression)	6.2 (2.8)	6.4 (1.5)	6.7 (4)	6.4 (2.9)
4 (Auditory suppression)	31.8 (10.2)	35.1 (12.4)	32.4 (15.3)	33.1 (12.60
5 (Somesthetic suppression)	32.2 (9.8)	33.4 (12.4)	32.3 (13.7)	32.6 (11.80)

Figure 1. Histogram illustrating the percent reduction in the maximum slow-phase nystagmus velocity for visual, auditory, somatosensory, imaginary, and combined auditory and somatosensory suppressors in Investigation 2. Condition 2 (i.e., maximum slow-phase velocity with the subject oscillating in darkness and not performing alerting tasks) served as the baseline. A positive number indicates that VOR gain was greater than baseline, and a negative number indicates the VOR was suppressed.



with no prior history of either otological or neurological impairments. The VOR was assessed at 0.08 Hz only, and two additional testing conditions were added to the original five conditions. In Condition 6 (imaginary visual target), subjects were asked to stare at an imaginary light moving with the chair. For Condition 7, we combined the auditory and somatosensory targets to assess whether there was an additive effect on VOR suppression. For this condition, subjects were instructed to grasp the vertical bar to which the sound source was affixed. Subjects were instructed to direct their attention to the location of their hand and sound source on the vertical bar. As in Investigation 1, the auditory stimulus was presented at 0° azimuth at a comfortable intensity (i.e., 62 dBA). Each condition (i.e., darkness with tasking, darkness without tasking, visual suppression, auditory suppression, somatosensory suppression, imaginary visual target, and combined auditory and somatosensory suppression) was replicated once, and the order of presentation of conditions was randomly assigned.

Results

Results for Investigation 2 are summarized in Table 3. Table 3 shows the mean VOR gain for each of the seven conditions for both Investigations 1 and 2. As with Investigation 1, VOR gain was greatest for Conditions 1 and 2 (darkness with and without tasking) and was smallest for Condition 3 (visual suppression). Compared with Conditions 1 and 2, the VOR gain was significantly smaller for all five suppression conditions. It is interesting to note that VOR gain was similar for Conditions 4–7 (auditory suppression, somatosensory suppression, imaginary target, and combined auditory and somatosensory suppression).

The data were analyzed with an ANOVA, where VOR gain (i.e., range of 0%–100%) served as the dependent variable and condition (seven levels) served as the grouping variable. The results revealed a significant main effect for condition, $F(6, 24) = 24.84$, $p < .001$.

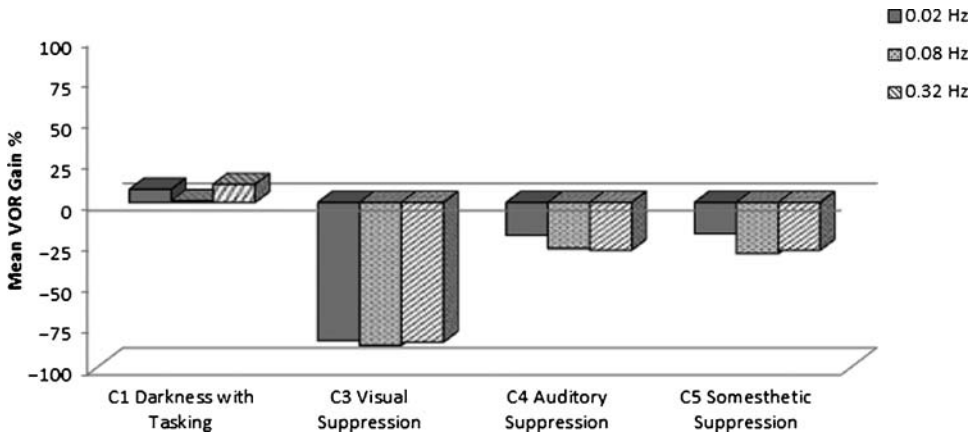
Post hoc analysis (Bonferroni) showed that VOR gains for Conditions 1 and 2 were significantly larger ($p < .001$) when compared with all five VOR suppression conditions (range of ps < .0001–.027). VOR gain was lowest (i.e., VOR suppression was greatest) for Condition 3 (visual suppression, $p < .001$) compared with all other conditions. The VOR gains for Conditions 4–7 were not statistically different from one another ($p = 1.0$). In other words, the VOR was suppressed to the same magnitude with attention directed toward an imaginary target as it was with attention directed toward auditory or somatosensory targets. Further, there was no additive effect of combining visual attention directed toward the auditory and somatosensory targets. As with Figure 1, for Figure 2, we used Condition 2 (i.e., VOR gain with the subject oscillating in darkness and not performing alerting tasks) as the baseline condition. As shown in Figure 2, the VOR was enhanced by 17% when the subject was performing an alerting task (Condition 1). The visual suppression condition (Condition 3) attenuated the VOR by 84%. The VOR was reduced between 34% and 44% of baseline for the other five suppression conditions.

Table 3. Mean VOR gain (*SD*) at 0.08 Hz shown for all conditions in Investigations 1 and 2.

Condition	VOR gain: Investigation 1 (n = 16)	VOR gain: Investigation 2 (n = 5)
1 (Darkness with alerting)	49.2 (16.0)	56.3 (8.9)
2 (Darkness without alerting)	48.7 (11.3)	47.8 (7.9)
3 (Visual suppression)	6.4 (1.5)	7.2 (1.8)
4 (Auditory suppression)	35.1 (12.4)	31.2 (6.1)
5 (Somatosensory suppression)	33.4 (12.4)	25.9 (7.7)
6 (Imaginary target)	N/A	31.6 (7.8)
7 (Combined auditory and somatosensory)	N/A	26.4 (7.5)

Note. N/A = not applicable.

Figure 2. Histogram illustrating the percent reduction in the maximum slow-phase nystagmus velocity for visual, auditory, and somatosensory suppressors in Investigation 1. Condition 2 (i.e., maximum slow-phase velocity with the subject oscillating in darkness and not performing alerting tasks) served as the baseline. A positive number indicates that VOR gain was greater than baseline, and a negative number indicates the VOR was suppressed.



Discussion

The objectives of the current investigation were (a) to reassess to what extent attention directed to visual and nonvisual targets can suppress the VOR, (b) to determine how comparable are the values we measure compared with those obtained from previous investigations, and (c) to determine what might be the possible origin of VOR suppression if VOR suppression occurs for nonvisual targets. Results suggested that using the unsuppressed VOR data as a baseline (i.e., Condition 2, unalerted in darkness), attention directed toward auditory and somatosensory suppressors each attenuated the VOR by 28%–44% across frequencies. However, in comparison, the magnitude of visual suppression of the VOR (i.e., Condition 3, visual suppression) was approximately two to three times greater (i.e., 84%–87% across frequencies) than the suppression achieved when subjects were attempting to stare at auditory or somatosensory targets in darkness. These findings are reasonably consistent with those reported by previous investigators (Barr et al., 1976; Furst et al., 1987; Jones et al., 1984; Moller et al., 1990). For example, visual suppression of induced nystagmus with a chair-fixed target has been reported to range from 93% to 96% (Furst et al., 1987; Moller et al., 1990). Moller et al. (1990) reported that attention directed toward a somatosensory stimulus produced on average 56% suppression of nystagmus (i.e., at frequencies 0.02 Hz, 0.08 Hz, and 0.32 Hz). The same investigators reported that attenuation of the VOR increased only slightly to 63% on average when attention was directed to a combined auditory and somatosensory target (i.e., the target was a handheld loudspeaker).

Previous investigators have suggested that an actual visual target is not required during head movement to produce a reduction in the VOR. For example, both Barr et al. (1976) and Jones et al. (1984) reported that the VOR could be attenuated during rotation in darkness when subjects were asked to fix their gaze on an imaginary visual target. That is, a real visual target was not required for attenuation of the VOR. These findings were similar to those reported by Furst et al. (1987) and Moller et al. (1990). It was the opinion of Moller et al. (1990) that their data supported the contention that modulation of the VOR was not dependent entirely on the visual system but could be potentiated by nonvisual sensory systems. The results of our second investigation showed that the VOR could be suppressed by 44% when subjects were asked to stare at an imagined visual target. This attenuation of the VOR was not significantly different from what occurred when subjects were asked to fix their gaze in darkness on a source of sound or somatosensation. Further, the suppressive effects of having multiple sensory sources to stare at were no different than those for each stimulus in isolation. Thus, these findings are consistent with the interpretation that there is a common process underlying the reduction of the VOR by gaze fixation of all targets in darkness, either visual or nonvisual.

The data in this investigation may provide us with some clues about the processes underlying VOR suppression. That is, the data would appear to support the interpretation that possibly two processes may contribute to suppression of the VOR. The first process may represent activation of other (i.e., in addition to the visual system) modality-specific (i.e., somatosensory and auditory) efferent pathways to the vestibular nuclei that inhibit the VOR. In this regard, it has been explained that corticofugal fibers destined for the vestibular nuclei originate from the premotor cortex; the cingulate cortex; the intraparietal sulcus; the insula; the retroinsular cortex; and the posterior region of the superior temporal cortex (Akbarian, Grusser, & Guldin, 1992, 1994; Faugier-Grimaud & Ventre, 1989; Guldin, Akbarian, & Grusser, 1992; Kawano, Sasaki, & Yamashita, 1980; Leinonen, Hyvarinen, & Sovijarvi, 1980; Nishiike, Guldin, & Baurle, 2000). These are brain regions that are proximal to vestibular, somatosensory, and auditory sensory reception areas (Fukushima, 1997). Electrical stimulation of these regions has been reported to produce both inhibition and facilitation of electrical output from the vestibular nuclei (Fukushima, 1997).

However, others have speculated that these cortical projections to the vestibular nuclei are predominantly inhibitory (Akbarian et al., 1994). Our data would suggest that if multisensory efferent pathways can suppress the VOR, the effect is modest (i.e., roughly one third of the magnitude of visual suppression).

A second process may represent cortical modulation of the VOR through attention directed toward a chair-fixed nonvisual target. We suggest that the effect of attention may represent a separate, nonmodality-specific process, as the magnitude of VOR suppression was not significantly different for auditory targets, somatosensory targets, imagined visual targets, or combinations of auditory and somatosensory targets. In this regard, Moller et al. (1990) also reported that the magnitude of suppression did not increase when attention was directed simultaneously toward both auditory and somatosensory stimuli (i.e., the suppressive effects were not additive). We can observe evidence of attentional effects on the VOR in the clinic during caloric testing. Attention directed toward the thermal stimulus, or the induced vertigo that follows, results in a diminution of the induced nystagmus velocity. Attention directed away from the sensation (i.e., that occurs during alerting exercises) deattenuates the induced nystagmus.

Summary

Our data are consistent with those reported by previous investigators (Barr et al., 1976; Furst et al., 1987; Jones et al., 1984; Moller et al., 1990) and suggest that it is possible for normal adults to suppress the VOR, in the absence of a visual target. That is, in addition to vision, we have shown that the VOR can be attenuated with attention directed toward chair-fixed auditory somatosensory and imagined visual targets. We suggest that this reduction in induced nystagmus may occur because of the effects of corticofugal pathways that are modality specific and destined for the brainstem vestibular center. Additionally, we believe that the data suggest that a nonmodality-specific process, such as directed attention, can explain how gaze in darkness directed toward nonvisual stimuli can result in suppression of the VOR.

Acknowledgments

This investigation was conducted by Christina Do in partial fulfillment of her obligations as a recipient of a National Institutes of Health T-35 award.

References

Akbarian, S., Grusser, O. J., & Guldin, W. O. (1992). Thalamic connections of the vestibular cortical fields in the squirrel monkey (Saimiri sciureus). *Journal of Comparative Neurology, 326,* 423–441.

Akbarian, S., Grusser, O. J., & Guldin, W. O. (1994). Corticofugal connections between the cerebral cortex and brainstem vestibular nuclei in the macaque monkey. *Journal of Comparative Neurology, 339,* 421–437.

Barber, H., & Stockwell, C. (1980). *Manual of electronystagmography.* St. Louis, MO: C. V. Mosby.

Barr, C. C., Schultheis, L. W., & Robinson, D. A. (1976). Voluntary, non-visual control of the human vestibulo-ocular reflex. *Acta Oto-laryngologica, 81,* 365–375.

Brey, R., McPherson, J., & Lynch, R. (2008). Technique, interpretation, and usefulness of whole body rotational testing. In G. Jacobson & N. Shepard (Eds.), *Balance function assessment and management* (pp. 281–317). San Diego, CA: Plural.

Faugier-Grimaud, S., & Ventre, J. (1989). Anatomic connections of inferior parietal cortex (area 7) with subcortical structures related to vestibulo-ocular function in a monkey (*Macaca fascicularis*). *Journal of Comparative Neurology, 280,* 1–14.

Fukushima, K. (1997). Corticovestibular interactions: Anatomy, electrophysiology, and functional considerations. *Experimental Brain Research, 117*(1), 1–16.

Furst, E. J., Goldberg, J., & Jenkins, H. A. (1987). Voluntary modification of the rotatory induced vestibulo-ocular reflex by fixating imaginary targets. *Acta Oto-laryngologica, 103,* 232–240.

Guldin, W. O., Akbarian, S., & Grusser, O. J. (1992). Cortico-cortical connections and cytoarchitectonics of the primate vestibular cortex: A study in squirrel monkeys (*Saimiri sciureus*). *Journal of Comparative Neurology, 326,* 375–401.

Jones, G. M., Berthoz, A., & Segal, B. (1984). Adaptive modification of the vestibulo-ocular reflex by mental effort in darkness. *Experimental Brain Research, 56*(1), 149–153.

Kawano, K., Sasaki, M., & Yamashita, M. (1980). Vestibular input to visual tracking neurons in the posterior parietal association cortex of the monkey. *Neuroscience Letters, 17,* 55–60.

Konrad, H. R., Girardi, M., & Helfert, R. (1999). Balance and aging. *Laryngoscope, 109,* 1454–1460.

Leinonen, L., Hyvarinen, J., & Sovijarvi, A. R. (1980). Functional properties of neurons in the temporo-parietal association cortex of awake monkey. *Experimental Brain Research, 39,* 203–215.

Moller, C., White, V., & Odkvist, L. M. (1990). Plasticity of compensatory eye movements in rotatory tests: II. The effect of voluntary, visual, imaginary, auditory and proprioceptive mechanisms. *Acta Oto-laryngologica, 109,* 168–178.

Nishiike, S., Guldin, W. O., & Baurle, J. (2000). Corticofugal connections between the cerebral cortex and the vestibular nuclei in the rat. *Journal of Comparative Neurology, 420,* 363–372.

AJA

Research Article

Influence of Tasking During Vestibular Testing

Kathryn Makowiec,[a] Kaylee Smith,[a] Ashley Deeb,[a] Erica Bennett,[a] and Jenni Sis[b]

Purpose: The purpose of this study was to investigate the effectiveness of different types of tasking on the measurement of peak slow phase velocity (SPV) for caloric testing and rotary chair testing.
Method: This study evaluated the peak SPV response for caloric testing and rotary chair across five conditions. Three verbal, one tactile, and one condition without tasking were used for both caloric testing and rotary chair. The subjects consisted of 20 young adults (age range: 22–33 years, *M* = 26.65, *SD* = 3.72; seven male, 13 female) with normal vestibular function and no history of ear surgery or vestibular disorder. Study participation consisted of two visits with 24 hr minimum between each, one for caloric testing and one for rotary chair testing. The test completed at each visit was counterbalanced.

Caloric Testing: The caloric irrigations were performed 5 times, with the ears randomized and tasking conditions randomized.

Rotary Chair Testing: Rotary chair sinusoidal harmonic acceleration testing was performed 5 times at 0.08 Hz with the tasking conditions randomized.
Results: Tasking of any kind resulted in significantly larger peak SPV responses when compared to the no tasking condition for rotary chair testing. When comparing each type of tasking, no significant differences were noted. No significant difference was noted when comparing the conditions with tasking to the no tasking condition for caloric testing.
Conclusions: Clinically, either mental or tactile tasking can be utilized as a method to reduce VOR suppression during rotary chair testing. As no difference was found when comparing different verbal tasks to each other, the type of tasking can be catered to the patient. If verbal tasking cannot be completed, the braiding tactile task is a valid substitution. Caloric results varied widely across subjects and did not reach statistical significance, so conclusions on the need for tasking cannot be drawn.

Accurate identification and management of vestibular disorders relies on measurements of the vestibulo-ocular reflex (VOR). The VOR is a compensatory reflex that allows for the stabilization of vision on the retina during active head movement or changes in body position. In response to the vestibular pathway stimulation created by active head movement, the VOR produces a reflexive eye movement called nystagmus. The slow phase velocity (SPV) of this eye movement is equal in magnitude and opposite in direction of the head movement. Stimulation of the VOR is regularly used to identify and quantify peripheral and central vestibular system lesions. When the VOR is induced, the intensity of the nystagmus that is produced in response can be measured.

The VOR can be induced either through active head/body rotation, such as with rotary chair testing, or induced by thermal changes within the ear canal, such as with caloric testing during videonystagmography (VNG) or electronystagmography (ENG). The peak SPV of the nystagmus is most often used clinically to assess vestibular function. When using peak SPV as a measurement of vestibular function clinically, it is important to ensure a robust response is being recorded. One method to ensure a robust response when measuring the VOR is to control for suppression. Suppression of the VOR occurs when the reflex is reduced or eliminated by extra-vestibular factors.

Common causes of VOR suppression during testing include light/visual stimuli and inadequate mental stimulation. Well-supported methods for controlling visual stimuli during testing include having the patient close their eyes or eliminating all light sources in the testing room (Baloh et al., 1977; Karlsen et al., 1980; Jacobson & Newman, 1993). This is either completed through the use of well-fitting video goggles with a cover if completing VNG or through having a patient close their eyes if completing ENG. Rotary chair testing either requires well-fitting video

[a]Division of Audiology, Department of Otolaryngology – Head & Neck Surgery, Henry Ford Health System, Detroit, MI
[b]Ear, Nose, Throat, and Audiology, CHI Health, Omaha, NE
Correspondence to Kathryn Makowiec: kmakowi1@hfhs.org

Editor-in-Chief: Ryan W. McCreery
Editor: Jamie Bogle
Received December 30, 2020
Revision received March 4, 2021
Accepted May 24, 2021
https://doi.org/10.1044/2021_AJA-20-00227

Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

goggles with a cover or an enclosed surround to eliminate light sources.

A common method to provide adequate mental stimulation during testing is mental tasking, which requires the patient to perform a cognitive task while the VOR response is being measured (Collins, 1962; Collins et al., 1960). These cognitive tasks will often be performed in a question–answer format and can include listing, calculations, or be more conversationally based. Cognitive tasks completed within a clinical setting often vary depending on the examiner, the patient, and the clinic.

Multiple studies have shown that mental tasking does matter and will reduce a person's ability to suppress the VOR response (Collins, 1962; Collins et al., 1960; McGovern & Fitzgerald 2008). Generation of the fast phase of nystagmus likely requires steady stimulation of higher-level cortical activity, and mental tasking maintains this steady stimulation (Barin, 2009). In addition to whether to task, the type of mental tasking may also matter. Some studies discussed below have compared different types of mental tasking to each other in an attempt to find an optimal type of mental tasking.

Numerous types of tasking (i.e., conversational, mathematical, active, passive, tactile, and no tasking) have been investigated with variable results. Studies by Kileny et al. (1980) and King et al. (2006) found that verbal tasking resulted in larger measured SPV responses when compared to no tasking. However, studies by Jacobson et al. (2012) and Easterday et al. (2016) found no difference in the VOR response when comparing conditions with tasking to conditions without tasking. In addition to the importance of tasking in general, the effectiveness of different types of tasking has also been investigated. For example, Kileny et al. (1980) reported conversational tasking resulted in larger amplitude responses than a mathematical task. Formby et al. (1992) found that a noninteractive quizzing task with little interaction with the examiner resulted in the largest amplitude responses. Davis and Mann (1987) examined the difference between passive and active tasking and found that active mental tasking resulted in larger amplitude SPV measurements than passive tasking, but no difference was noted between the two types of mental tasking (mathematical and answering questions).

At times within a clinical setting, verbal/aural tasking cannot be completed, such as if a patient has significant hearing loss or is not a native English language speaker. Tactile tasking is a mental alerting option that may be used in these types of situations. In addition to reporting that tasking resulted in larger measured SPV responses than no tasking, King et al. (2006) developed a vibrotactile tasking paradigm to be used during caloric testing. They found no significant difference on the peak SPV of the VOR between verbal and vibrotactile tasking.

The current study had multiple purposes to address the inconsistencies noted in prior literature. The first was to examine if there is a difference between types of mental tasking on the peak SPV response measured during caloric testing and rotary chair testing within the same population. The second, in response to Easterday et al. (2016), is to investigate the effect of tasking versus no tasking on the peak SPV response measured. The third purpose is to confirm that a tactile braiding task can be utilized in place of mental tasking.

The researchers felt this study was of value as previous literature has not compared the impact of different types of tasking on both caloric and rotary chair responses within the same population. Anecdotally within a clinical setting, the researchers have not found that one type of mental tasking works best for all patients, even when the patients have similar education levels, but have found that some sort of tasking typically results in larger responses than when no tasking is used. Additionally, the type of tactile tasking utilized in this study has been used clinically where the researchers practice but has not been validated in previous literature.

Method

Subjects

All procedures were approved by the Henry Ford Health System's Institutional Review Board (IRB # 00000253, Protocol 12721). Participants included 20 young, healthy adults (seven males, 13 females; age range: 22–33 years, M = 26.65, SD = 3.72). To be included in the study, all subjects underwent video head impulse testing (vHIT) to screen for normal lateral semicircular canal and superior vestibular nerve function in the 3–5 Hz frequency range. vHIT testing was completed using the GN Otometrics ICS Impulse equipment. Additionally, subjects with a history of ear surgeries or vestibular disorder(s) were excluded.

Procedure

Study participation consisted of two visits with 24 hr minimum between each, one for caloric testing and one for rotary chair testing. Each visit was 1 hr in duration. The study was conducted within the Henry Ford Hospital Vestibular Lab. The test completed at each visit (i.e., rotary chair or caloric testing) was counterbalanced across subjects. The same clinician performed the testing and tasking for all visits. Five tasking conditions were investigated for both caloric testing and rotary chair testing. These conditions included (a) spatial awareness, such as "describe your home from when you first enter your front door as you walk through"; (b) alphabet/listing, such as "tell me an animal that begins with the letter A, the letter B, etc."; (c) counting/numerical, such as "beginning at 100, count backwards by 3"; (d) tactile, where the subject was provided with three cords and was told to begin braiding the cords together at a certain time and not stop braiding until instructed to do so; and (e) no tasking. Different tasking methods within each condition were used for caloric testing and rotary chair testing to reduce any learning effect. For example, subjects were asked to list female names with a specific letter A–Z during the caloric test and asked to list male names with a specific letter A–Z during the rotary chair

test. The tasking prompts for each condition were consistent across subjects; they were all given the same prompts for the caloric tasking conditions and were all given the same prompts for the rotary chair tasking conditions. Table 1 shows the tasking prompts used for each condition for caloric irrigations and rotary chair. The order of tasking condition was randomized for each subject and each test. There was no payment provided to subjects for participation.

Caloric Testing

GN Otometrics Chartr VNG goggles were comfortably placed on the subject and calibration was completed prior to air caloric irrigations. Additionally, otoscopic examination was completed to ensure no obstructing cerumen in the external ear canals prior to irrigation. The subject was placed in the standard 30° supine position and a total of five warm air caloric irrigations were performed, using a different tasking condition for each irrigation. The caloric irrigations were completed in both the right (R) and left (L) ear, and the order was randomized (for example, one subject underwent R, L, R, R, L, and another subject underwent L, L, R, R, L). The tasking condition assigned to each irrigation was additionally randomized. Randomization of ear and tasking condition ensured habituation would not impact the obtained results. The irrigation temperature was set to 50 °C and the length of each irrigation was 60 s.

All five irrigations were completed within a 1 hr session. The subject was given 4 min between each irrigation with the VNG goggles open to allow the ear to return to body temperature and to allow the caloric response to dissipate before completing the next irrigation. The subjects were instructed to inform the examiner if the caloric irrigations became uncomfortable or if they wished to stop prior to the completion of all five conditions. No subject reported discomfort and all were able to undergo all five caloric irrigations without difficulty. For each condition with tasking, the subject was engaged by the examiner immediately upon the completion of the 60-s irrigation, and the tasking was continued until after the peak SPV amplitude was observed and the response subsided. A minimum of 60 s of recording was completed after the completion of the 60-s irrigation to ensure the peak SPV was recorded. For the condition without tasking, the subject was only instructed to keep their eyes open and the response was recorded until it subsided. Reminders to keep eyes open were provided as needed for each subject. The peak SPV was calculated for each tasking method by the Otometrics VNG software and the experienced clinician reviewed, confirmed, and agreed with the calculated peak.

Rotary Chair Testing

Micromedical System 2000 Rotational Chair goggles were comfortably placed on the subject and calibration was completed prior to rotary chair testing. Rotary chair sinusoidal harmonic acceleration testing was completed 5 times at 0.08 Hz with a 2-min break with the goggles open in between each to allow the response to dissipate prior to continuing. For each condition with tasking, the subject was engaged by the examiner as soon as the chair began moving and tasking was continued the entire time the chair was moving. The condition without tasking only involved instruction and reminders to maintain eyes open throughout. Three cycles were completed for each condition (first cycle dropped automatically by software, analysis completed on second and third cycles). Peak SPV was calculated for each tasking method by the Micromedical Spectrum software and the experienced clinician reviewed, confirmed, and agreed with the calculated peak.

Data analysis

Data analysis was completed using SigmaPlot. The means and standard deviations for peak SPV response were calculated for each tasking condition for both caloric testing and rotary chair testing. Repeated-measures analyses of variance (ANOVAs) were conducted to analyze for significant difference between conditions. Additionally, the first caloric SPV response was compared to the last caloric SPV response for each subject to ensure that the first caloric irrigation did not result in an inaccurately large response due to the novel stimulus and to ensure there was no evidence of habituation to the caloric stimulus and caloric response by the fifth irrigation.

Results

Caloric data was unable to be analyzed from three subjects due to small ear canal size leading to poor irrigations (n = 17). Rotational chair testing was unable to be completed for one subject due to scheduling conflicts (n = 19).

Caloric Testing

Table 2 displays the mean, standard deviation, and range of the peak caloric SPV response for each condition

Table 1. Tasking prompts used for each condition for caloric irrigations and rotary chair.

Tasking condition	Calorics	Rotary chair
Spatial	Directions to today's appointment	Describe home, beginning at front door
Alphabet	Girls names A–Z	Boys names A–Z
Counting/numerical	Beginning at 100, count backwards by 3	Beginning at 0, count upwards by 3
Tactile	Braiding 3 cords together	Braiding 3 cords together
No tasking	N/A – reminders to keep eyes open only	N/A – reminders to keep eyes open only

Note. N/A = not applicable.

Table 2. Mean, *SD*, and range of the peak SPV response for caloric testing for each condition.

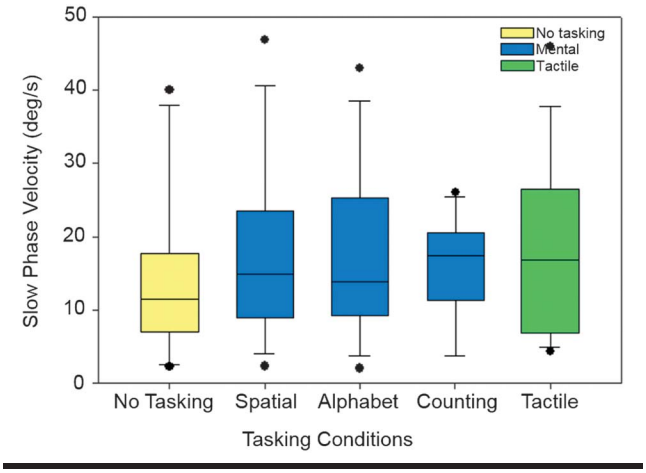
Caloric peak SPV	No tasking	Spatial	Alphabet	Counting/Numerical	Tactile
MEAN	14.47	16.94	17.82	16.47	18.53
ST DEV	10.18	10.19	11.06	9.08	11.60
Range	2–40	2–48	2–43	4–29	5–47

(no tasking, spatial, alphabetic, counting/numerical, tactile). On average, the condition without tasking produced the lowest SPV response; however, a repeated-measures ANOVA revealed no significant differences between the tasking and no tasking methods, $F(4, 64) = 0.78$, $p = .544$. Figure 1 shows the data from Table 2 in graph form; the box-and-whisker plot shows each caloric condition across the *x*-axis and the SPV in deg/s is shown on the *y*-axis. Each box indicates the first, median, and third quartiles with the whiskers showing the range from 10th to 90th percentiles. The outliers are plotted above or below the whiskers. No significant difference in SPV response was noted, $F(1,11) = 2.763$, $p = .125$, when comparing the first caloric response to the last caloric response within each subject, indicating that the data was not inaccurately influenced by which tasking condition was completed with the first irrigation and the last irrigation.

Rotary Chair Testing

Table 3 displays the mean, standard deviation, and range of the peak rotary chair SPV response for each condition (no tasking, spatial, alphabetic, counting/numerical, tactile). A repeated-measures ANOVA revealed a significant effect of tasking on rotary chair results, $F(4,12) = 11.906$, $p = < .001$. Post hoc comparisons using a Bonferroni *t* test indicated that the mean SPV for the condition without tasking was significantly lower than the remaining four tasking conditions ($p < .001$). Figure 2 shows the data from Table 3 in graph form; the box-and-whisker plot shows each rotary chair condition across the *x*-axis and the SPV in deg/s is shown on the *y*-axis. Each box indicates the first, median, and third quartiles with the whiskers showing the range from 10th to 90th percentiles. The outliers are plotted above or below whiskers. No significant difference was noted between the other four tasking conditions (spatial, alphabet, counting/numerical, tactile) when compared to each other.

Figure 1. Box plot of the peak slow phase velocity in deg/s for caloric testing for each condition. Shown are the median (black line) and the range of the 75th to 25th percentile (i.e., first and third quartiles). Outliers that fall outside the whiskers are shown as black dots.



Discussion

The purpose of this study was to assess if the type of mental tasking impacts the peak VOR measurement for caloric and rotary chair testing, if there was a significant difference between conditions with tasking and with no tasking, and if the braiding tactile tasking method utilized is an acceptable mental alerting task replacement. This study compared different forms of mental tasking to analyze if one type of tasking was notably better than the others. No significant difference was noted across the mean peak SPV measurements using the three verbal types of mental tasking (alphabet listing, spatial, and counting/numerical) for both caloric and rotary chair testing. All three of these forms of tasking resulted in robust peak SPV responses for both caloric and rotary chair tests and none were significantly better than the other. This is in disagreement with the research completed by Formby et al. and Kileny et al. Formby et al. (1992) found a significant difference in the peak SPV measurement across types of tasking, with counting/numerical tasking resulting in the lowest measured SPV responses. Kileny et al. (1980) also found a significant difference in types of mental tasking; their study showed that mathematical tasking resulted in significantly lower amplitude caloric responses than conversational based tasking. Our findings are, however, in agreement with what was found by Davis and Mann (1987), who did not note a significant difference between types of active mental alerting tasks.

When comparing the tasking and no tasking conditions within our study, there was a significant difference noted on the peak SPV response for rotary chair testing, where all of the tasking conditions resulted in significantly larger responses than the no tasking condition. This finding is contradictory to the results found by Jacobson et al. (2012), which investigated how visual and nonvisual stimuli could induce VOR suppression, causing a reduction in the responses measured. While it was not the purpose of their study, review of the data showed there was no significant difference in the measured peak response for rotary chair

Table 3. Mean, standard deviation, and range of the peak slow phase velocity (SPV) response for rotary chair testing for each condition.

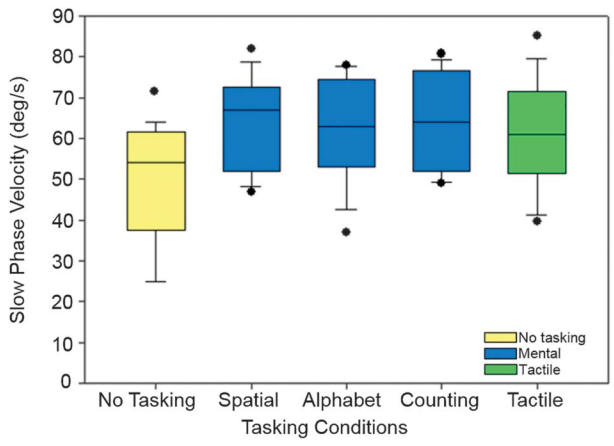
Rotary chair peak SPV	No tasking	Spatial	Alphabet	Counting/numerical	Tactile
M	49.42	62.47	60.58	60.89	62.79
SD	13.82	13.32	12.65	14.50	14.09
Range	25–72	47–84	35–78	48–80	40–88

when comparing mental tasking to no tasking conditions. For the purposes of their study, this indicated that mental tasking did not adequately suppress the VOR.

No significant difference on the measured SPV response for caloric testing was noted when comparing conditions with tasking to the no tasking condition. When looking at individual subject data across all five irrigations, the SPV response for the no tasking condition was the smallest of the responses measured across the five irrigations in seven out of 17 participants. Two of the 17 participants had their largest SPV response in the no tasking condition. The SPV response for the no tasking condition fell somewhere in the middle for the eight remaining subjects, where some mental tasking conditions resulted in a larger response than the no tasking condition and other mental tasking conditions resulted in a smaller response than the no tasking condition.

Our finding of no significant difference in the measured SPV response for caloric testing when comparing the no tasking condition to the conditions with tasking are in agreement with the findings of Easterday et al. (2016). The lack of statistical significance found between the tasking and no tasking conditions on the measured caloric response in our study may be attributed to the large variability seen on the caloric responses within individual subjects and across all subjects within each condition. This can be seen in Figure 1, with both large error bars and outliers plotted. Within individual subjects, caloric responses ranged widely, for example, from 17 deg/s for the smallest response to 47 deg/s for the largest response for one individual, whereas a few subjects had much tighter response ranges, for example from 5 deg/s for the smallest response to 10 deg/s for the largest response.

Figure 2. Box plot of the peak slow phase velocity in deg/s for rotational chair testing for each condition. Shown are the median (black line) and the range of the 75th to 25th percentile (i.e., first and third quartiles). Outliers that fall outside the whiskers are shown as black dots.



The outliers for each condition were further investigated to see if there was an impact of gender or age. The age range for the subjects who represented outliers in any condition was 22–33 years and the mean was 25.86 years. The age range for the study group as a whole was the same (22–33 years) and the mean across all subjects was 26.65 years, so age was not found to be an indicator of why an outlier was an outlier. Of the outliers across conditions, the male to female ratio was 2:5. When looking at all participants, the male to female ratio was 6:11, so gender did not appear to have an impact as to why an outlier was an outlier either. Additionally no one subject was found to be a source of outlier data points across all conditions.

To ensure that the first caloric did not result in an inaccurately large response, as is often noted in clinical settings due to the novel nature of the stimulus/task, and to ensure there was no habituation by the last caloric irrigation, we compared the peak SPV measured for the first irrigation to the peak SPV measured for the last irrigation for each subject. No significant change was noted when comparing the first to the last irrigations across subjects. This allowed us to be confident that the wide range of caloric responses seen within many of our subjects was not due to the first irrigation resulting in a significantly larger response and was not due to habituation to the caloric stimulus and response by the last irrigation.

Within each condition, the caloric responses could also widely range, for example within the alphabet tasking condition, the caloric response ranged from 2 deg/s for one participant to 43 deg/s for another participant. This information is shown in Table 2; all of the conditions, no tasking included, had wide ranges of responses. The counting/numerical task had the smallest range when comparing the smallest to the largest response; 4 deg/s for the smallest response to 29 deg/s for the largest response, which is still a large range across subjects for the same condition. Due to the large variability on the measured caloric responses within this data set, it is difficult to draw conclusions whether tasking had a significant impact on the caloric response.

Our study found no significant difference between the verbal mental tasking exercises (alphabet listing, spatial, and counting/numerical) and tactile tasking within both the caloric and the rotary chair data sets. This is in agreement with King et al. (2006) who reported no significant difference in the peak SPV of the VOR between verbal and vibrotactile

tasking methods. This indicates that the braiding tactile task utilized in this study is an appropriate replacement if verbal mental tasking cannot be utilized. Our data on rotary chair additionally agrees with what was found by King et al., who noted that both verbal and vibrotactile tasking resulted in significantly larger measured responses than the no tasking condition.

Based on the results of our study, we recommend that, in a clinical setting, either mental tasking or tactile tasking be utilized minimally when completing rotary chair testing to reduce suppression and allow for accurate and robust SPV measurements. Given the inconsistencies in the literature regarding whether to task, with some studies showing that tasking results in larger responses and others showing tasking is equivalent to no tasking, and as only two of the subjects had the largest response for calorics in the no tasking condition, we additionally recommend that some method of tasking is utilized during caloric testing, despite the data not reaching statistical significance. This recommendation differs from the recommendation made by Easterday et al. Like us, they found tasking did result in larger amplitude responses on calorics when looking at some individual data, but they recommended that tasking did not need to be used based on the group mean results and due to the increase in eye blink artifact noted when tasking.

As no difference was found when comparing the different verbal tasks to each other, the tasking type can be catered to the patient. Additionally, if verbal tasking is unable to be completed, for example if the patient has significant hearing loss or does not speak English, the braiding task described above can be used as a valid substitution for a verbal task.

It should be taken into consideration that this study and other studies examining the same topic are typically completed on young, normal subjects rather than the average clinical population seen within a vestibular clinic setting. Future research is needed to examine if there is any interaction of age, socioeconomic status, educational level, and underlying vestibular dysfunction on the best tasking method and whether or not tasking changes the peak SPV response that is able to be measured.

References

Baloh, R. W., Solingen, L., Sills, A. W., & Honrubia, V. (1977) Caloric testing. 1. Effect of different conditions on ocular fixation. *Annals of Otology, Rhinology, and Laryngology, 86*(5, Suppl. 3), 1–6. https://doi.org/10.1177/00034894770865s301

Barin, K. (2009). Clinical neurophysiology of the vestibular system. In J. Katz, L. Medwetsky, R. Burkhart, & L. Hood (Eds.), *Handbook of clinical audiology* (pp. 431–466). Lippincott, Williams & Wilkins.

Collins, W. E. (1962). Effects of mental set upon vestibular nystagmus. *Journal of Experimental Psychology, 63,* 191–197. https://doi.org/10.1037/h0039941

Collins, W. E., Crampton, G. H., Posner, J. B., & Warburton, S., III. (1960). The effect of mental set upon vestibular nystagmus and the electroencephalogram. *US Army Medical Research Laboratory, 12,* 1–15.

Davis, R., & Mann, R. (1987). The effects of alerting tasks on caloric induced vestibular nystagmus. *Ear and Hearing, 8*(1), 58–60. https://doi.org/10.1097/00003446-198702000-00011

Easterday, M. K., Plyler, P. N., & Doettl, S. M. (2016). Mental tasking and caloric-induced vestibular nystagmus utilizing Videonystagmography. *American Journal of Audiology, 25*(3), 177–183. https://doi.org/10.1044/2016_AJA-15-0070

Formby, C., Kuntz, L. A., Rivera-Taylor, I. M., Rivera-Mraz, N., Weesner, D. R., Butler-Young, N. E., & Ahlers, A. E. (1992). Measurement, analysis, and modelling of the caloric response. 2. Evaluation of mental alerting tasks for measurement of caloric-induced nystagmus. *Acta Oto-Laryngologica, 498,* 19–29. https://doi.org/10.3109/00016489209136855

Jacobson, G. P., & Newman, C. W. (1993). Background and technique of caloric testing. In G. P. Jacobson, C. W. Newman, & J. M. Kartush (Eds.), *Handbook of balance function testing* (pp. 156–192). Mosby.

Jacobson, G. P., Piker, E. G., Do, C., McCaslin, D. L., & Hood, L. (2012). Suppression of the vestibulo-ocular reflex using visual and nonvisual stimuli. *American Journal of Audiology, 21*(2), 226–231. https://doi.org/10.1044/1059-0889(2012/12-0021)

Karlsen, E. A., Goetzinger, C. P., & Hassanein, R. (1980). Effects of six conditions of ocular fixation on caloric nystagmus. *Archives of Otolaryngology—Head & Neck Surgery, 106*(8), 474–476. https://doi.org/10.1001/archotol.1980.00790320026007

Kileny, P., McCabe, B., & Ryu, J. (1980). Effects of attention requiring tasks on vestibular nystagmus. *Annals of Otology, Rhinology & Laryngology, 89,* 9–12. https://doi.org/10.1177/000348948008900104

King, J. E., Gonzalez, J. E., & Fuller, M. I. (2006). Development of a vibrotactile tasking device for use in vestibular assessment. *Journal of Vestibular Research, 16*(1–2), 57–67.

McGovern, T. N., & Fitzgerald, J. E. (2008). The effect of mental alerting on peripheral vestibular nystagmus during spontaneous, gaze (30 left, 30 right), and body positional (left & right lateral lying) testing using electronystagmography (ENG). *Internal Journal of Audiology, 47*(10), 601–606. https://doi.org/10.1080/14992020802123579

The Influence of Caffeine on Rotary Chair and Oculomotor Testing

DOI: 10.3766/jaaa.16118

Kathleen M. McNerney*
Mary Lou Coad†
Robert Burkard†

Abstract

Background: When patients are given instructions before vestibular function testing, they are often asked to refrain from ingesting caffeine 24 h before testing. However, research regarding the effects of caffeine on the outcome of vestibular function testing is limited.

Purpose: To evaluate whether the results from rotational chair tests are influenced by caffeine.

Research Design: Participants were tested after consuming a caffeinated beverage (i.e., coffee containing ~300 mg of caffeine), as well as after abstaining from caffeinated beverages. The participants underwent oculomotor testing, sinusoidal harmonic acceleration testing, optokinetic testing, visual enhancement/suppression testing, subjective visual vertical/horizontal testing, trapezoidal step testing, and unilateral utricular centrifugation testing.

Study Sample: Thirty healthy young controls aged 18–40 yr (mean = 23.28 yr; 9 males, 21 females) participated in the study.

Data Collection and Analysis: Rotational chair tests were completed with the Neuro Kinetics rotary chair (Pittsburgh, PA). VEST 7.0 software was used to collect and analyze the participants' eye movements (I-Portal VOG; Neuro Kinetics). IBM SPSS was used to statistically analyze the results.

Results: Statistically significant differences were found for the results from several oculomotor tests (i.e., vertical saccades [SCs], horizontal SCs, and optokinetics), whereas the remaining rotational chair tests did not reveal any statistically significant differences between sessions. If a statistically significant difference was found, the participants were then stratified based on the amount of caffeine they consumed on a daily basis. This stratification was accomplished based on the guidelines from the International Coffee Organization. When the data were analyzed based on the stratified groups, statistically significant results remained in the no/low caffeine intake group, whereas no statistically significant results remained in the moderate/high caffeine intake group. Clinically speaking, the largest effect was seen in those individuals who did not typically ingest large amounts of caffeine, whereas the results were not found to be significantly different in those individuals who were typical caffeine consumers. This strengthens the argument that it is not necessary to require that individuals refrain from consuming caffeinated beverages before oculomotor/rotary chair testing as the results from typical caffeine consumers are not significantly affected.

Conclusions: Although statistically significant results were found for a number of the oculomotor function tests, the ingestion of caffeine had little influence on the clinical interpretation of the responses. Therefore, the results from the present study indicate that it is not necessary to require that healthy young

*Department of Speech-Language Pathology, SUNY Buffalo State, Buffalo, NY; †Department of Rehabilitation Science, University at Buffalo, Buffalo, NY

Corresponding author: Kathleen M. McNerney, Department of Speech-Language Pathology, SUNY Buffalo State, Buffalo, NY 14222; Email: mcnernkm@buffalostate.edu

The first author (K.M.) received a New Investigator Research Grant from the American Academy of Audiology's Research Grants in Hearing and Balance program; this grant program is funded by the American Academy of Audiology/American Academy of Audiology Foundation Research Grants in Hearing and Balance Program.

Data collection and analysis was completed when the first author was working in the Department of Rehabilitation Science at the University at Buffalo.

This study was presented orally at the American Balance Society Meeting, March 6, 2013, Scottsdale, AZ.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

individuals abstain from caffeine before undergoing rotary chair/oculomotor testing. Further research is necessary to determine whether there is also a limited effect of caffeine on rotary chair/oculomotor test results from older individuals, as well as individuals diagnosed with a vestibular impairment.

Key Words: caffeine, rotary chair, rotational chair, vestibular testing

Abbreviations: C = caffeine; HC = high caffeine user; LC = low caffeine user; NC = no caffeine; OPK = optokinetic; RC = rotary chair; SCs = saccades; SHA = sinusoidal harmonic acceleration; SP = smooth pursuit; SVH = subjective visual horizontal; SVV = subjective visual vertical; TST = trapezoidal step test; UUC = unilateral utricular centrifugation; VE = visual enhancement; VNG = videonystagmography; VS = visual suppression

INTRODUCTION

Rotational chair testing can be used clinically in the assessment of the function of the vestibular system. It is capable of providing information that is not obtained via standard videonystagmography (VNG). For example, caloric testing cannot provide information about frequencies above 0.003 Hz (i.e., the frequency of the convection currents in the endolymph resulting from thermal stimulation), which is an extremely low frequency not typically activated during everyday head movements (Barin, 2008). The information obtained from rotational chair testing can provide an indication as to how the vestibular system functions in response to higher frequency accelerations of the head; e.g., from 0.01 to 1.28 Hz, or sometimes 2 Hz (Brey et al, 2008). In addition, evaluation of the phase of eye movements obtained through rotational chair testing can provide information regarding how well someone has compensated for their vestibular impairment (Rubin, 1982).

Tests of oculomotor function are used to evaluate the central pathways which are imperative to the normal function of the vestibular–ocular reflex. These tests (which include smooth pursuit [SP], saccades [SCs], and optokinetics [OPKs]) require one to be able to accurately follow an illuminated dot (or dots) which are projected in front of them. Previous research has shown that the results from oculomotor tests can be affected by severe fatigue or the inability to attend to the stimulus (Hale et al, 2015).

Patients are often asked to refrain from ingesting anything that contains caffeine before undergoing routine vestibular testing (BayCare Clinic Ear; Brey et al, 2008; ENTCare). Previous studies in our laboratory have shown that caffeine has minimal effects on calorics, vestibular-evoked myogenic potentials, and the sensory organization test (McNerney et al, 2014a,b). The present study furthered the evaluation of caffeine on the results from vestibular function testing by examining the effects of caffeine on additional tests which can be administered in the Neuro Kinetics (Pittsburgh, PA) rotary chair.

METHODS

Thirty individuals between 18 and 40 yr of age (9 males, 21 females; mean = 23.28 yr), served as the participants in the present experiment. None of the participants reported a history of vestibular or balance dysfunction. The participants were tested during two separate sessions, which lasted ~2–3 h each. During the caffeine (C) session, individuals were asked not to drink caffeine the morning of the test and before data collection, and they were asked to drink 16 oz. of Starbucks Breakfast Blend coffee which contained ~300 mg of caffeine (McCusker et al, 2003; Caffeine Informer, 2014). A higher amount of caffeine was chosen to determine whether there were any effects of larger amounts of caffeine on rotary chair tests. This would then allow generalization of the results to individuals who drank more than two cups of coffee with lower amounts of caffeine. The participants finished the coffee within ~30 min, which coincided with the peak absorption time (i.e., ~30 min), and the sessions were completed well within the half-life of caffeine, i.e., 2.5–10 h (Dernaro and Benowitz, 1991). During the no caffeine (NC) sessions, the participants were asked not to consume caffeine for 24 h before testing. Testing sessions were counterbalanced across the participants and were separated by at least 1 day. The participants were also asked to keep a caffeine diary over a 7-day time period, which allowed us to evaluate how much caffeine they drank on a daily basis. The participants were asked to insert the number and type/brand of each drink they consumed (i.e., espresso and espresso drinks, brewed coffee, and black tea). An "other" column was included in the event that a particular participant consumed a beverage that was not included on the list. The exact amount of daily intake was then computed by the experimenters using online resources (Mayo Clinic, 2014; Wilstar, 2014). This information was then used to separate the participants into a low caffeine (LC) intake group, which consisted of individuals who consumed no/low amounts of caffeine per day (i.e., 0–200 mg) or a high caffeine (HC) intake group, which consisted of individuals who consumed moderate/high amounts of caffeine per day (i.e., >200 mg). Criteria for stratifying the participants into LC versus HC intake groups were based on data from the International Coffee Organization (2012). This stratification of groups was used to analyze the results from the C versus NC session as a function of caffeine intake when statistically significant results were revealed (i.e., the C versus NC results were compared in the LC intake group and

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

the C versus NC results were compared in the HC intake group).

The participants were also asked to answer a caffeine withdrawal questionnaire before the NC session, which assessed the severity of 14 common caffeine withdrawal symptoms, e.g., fatigue, fogginess, and irritability (Ozsungur et al, 2009). The participants indicated whether they were experiencing a particular symptom, as well as the severity of that symptom on an 11-point Likert scale (e.g., 0 = no symptom present versus 10 = experiencing this particular symptom on a severe scale).

Rotary Chair Testing

The participants were secured in a Neuro Kinetics rotary chair. I-Portal VOG eye goggles (Neuro Kinetics) were used to record the movements of each individual eye (video-oculography). VEST 7.0 software was used to collect and analyze the participants' eye movements (Neuro Kinetics). The participants first performed the following oculomotor tests.

SP

This test evaluates the ability to track an object with smooth eye movements. This is a test of oculomotor function. SP was tested in the horizontal plane at 0.10 Hz (three cycles), 0.030 Hz (three cycles), 0.050 Hz (four cycles), and 0.75 Hz (six cycles), and in the vertical plane at 0.10 Hz (three cycles), 0.30 Hz (three cycles), and 0.50 Hz (four cycles). The parameters analyzed were gain, phase (°), and asymmetry (%) of eye movements.

SCs

This test evaluates movement of the eyes in response to rapid "jumps" of an illuminated dot. Sixty SCs presented at random times and displacements were presented in the horizontal as well as the vertical planes. Peak velocity (°/sec), latency (sec), accuracy (%), and duration (sec) of eye movements were analyzed. Saccade duration was calculated by computing the difference between when the movement of the eye starts to when the eye reaches target (i.e., stopping) position.

Full-Field Optokinetics (OPK)

This test measures the nystagmus created by repeated stimuli (i.e., Dots which encompass the participant's entire visual field. It resembles the type of lights emitted from a disco ball.) moving in front of the participant. The stimuli were presented at either 20, 40, or 60°/sec. Ramp up/down time was 0.5 sec and peak time was 10 sec for each stimulus speed. Eye velocity gain (normalized to 20°/sec) during each stimulus speed was collected and analyzed.

The participants then underwent the following rotary chair (RC) tests.

Trapezoidal Step Test (TST)

During this evaluation, the horizontal vestibular–ocular reflex is tested. The participants underwent an acceleration phase of 0.8 sec until they reached a peak velocity of 100°/sec. The participants were then rotated for an average of 60 sec. In healthy individuals, when the specified velocity is reached and maintained without further acceleration, the participant falsely perceives that the chair is slowing down, and the evoked nystagmus will eventually stop. Once this occurred, the participants underwent a deceleration step to a complete stop. During this phase, the participants incorrectly perceive that they are rotating in the opposite direction. The participants' eyes were again monitored and recorded for up to 60 sec after the rotary chair was stopped. The participants then underwent the second phase of the TST with the chair rotating in the opposite direction (Shepard, 2009). The parameters analyzed included peak velocity (°/sec), decay time (sec), and gain.

Sinusoidal Harmonic Acceleration (SHA)

During this evaluation, the horizontal semicircular canal is tested in response to repetitive sinusoidal motion of the rotary chair. Several frequencies were tested, including 0.02 Hz (two cycles), 0.04 Hz (three cycles), 0.08 Hz (three cycles), 0.64 Hz (eight cycles), and 1.28 Hz (nine cycles). Gain, phase (°), and asymmetry (%) of the eye movements were measured in relation to the chair movements.

Subjective Visual Vertical/Horizontal (SVV/SVH)

This is a test of otolith function. Healthy individuals are generally very good at setting the line very close to 0°, i.e., on average within 1°–2° (Bronstein, 2008). During this evaluation, the participants were presented with a straight line at an angle of −20° to +20° to true vertical, and −35° to +35° to true horizontal for SVV and SVH, respectively. The participants were then asked to adjust this line using a push button located on either handle of the rotary chair until it was as close to vertical (six trials) and as close to horizontal (six trials) as possible.

Unilateral Utricular Centrifugation (UUC)

This is a test of utricular function, and it can provide independent information on the function of each utricle separately. During this evaluation, the participants were rotated until they reached a maximum velocity of 300°/sec (ramp up time = 60 sec; peak time = 375 sec; ramp down time = 100 sec). They were then shifted to the right 4 cm, back to the center, and then to the left 4 cm. During each shift of the chair, the participants

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

were asked to perform multiple trials of SVV (i.e., ideally up to three). The average of the SVV deviations during each shift of the chair were collected and analyzed.

Visual Enhancement/Suppression (VE/VS)

During the VE test, the participants are rotated in the chair at 0.64 Hz while OPK stimuli are illuminated on the wall. In a healthy individual, the gain of the eye movement should be greater compared with when the participant is simply being rotated in the dark. The parameters analyzed included eye gain, asymmetry (%), and phase (°). During the VS test, the participants are rotated in the chair and are asked to remain focused on an illuminated dot which spins with the participant. A healthy individual will be able to suppress any nystagmus which is evoked from the movement of the chair. The results from eye gain were analyzed.

For further information regarding the above tests, please see Brey et al (2008).

Statistical Analysis

Statistical analyses via paired *t*-tests were initially completed with IBM SPSS Statistics 20. When statistically significant results were found, individuals were allocated into different groups based on their weekly caffeine consumption (i.e., LC intake group versus HC intake group). Effect size was calculated via a Cohen's *d* for paired tests ($d = [M_{\text{difference}}/\text{SD}_{\text{difference}}]$). A Cohen's *d* of 0.20 or less would be equivalent to a small effect size, a Cohen's *d* of 0.50 would be equivalent to a medium effect size, and a Cohen's *d* of 0.80 would be equivalent to a large effect size (Cohen, 1988; Nolan and Heinzer, 2011). Statistical analyses of the data, which were conducted during the revision of the article, were completed with version 24 of the IBM SPSS Statistics software.

RESULTS

Caffeine Diary

Weekly caffeine consumption varied from 0 to 4,358 mg of caffeine per week. The weekly caffeine intake was divided by 7 to estimate the amount of daily caffeine intake among the participants. The daily caffeine intake ranged from 0 to 623 mg (mean = 162 mg; SD = 141). The NC/LC intake group was composed of 22 participants whose caffeine consumption ranged from 0 to 183 mg per day, whereas the moderate/high caffeine intake group was composed of eight participants whose daily caffeine intake ranged from 231 to 623 mg of caffeine per day.

Caffeine Withdrawal Questionnaire

Overall, the caffeine withdrawal questionnaire revealed very low caffeine withdrawal scores. Of the thirty participants, severity ratings from 27 participants were averaged (the caffeine diary from three participants revealed no caffeine intake and therefore were not included in the severity rating analysis). Out of 140 possible points, the mean of all of the symptom severity ratings combined across all of the participants was 11.41, whereas the average severity rating within participant ranged from 0 to 59. The most commonly reported symptom was tiredness (N = 19), followed by decreased energy/activeness (N = 18), sleepiness (N = 17), and decreased alertness/attentiveness (N = 16). For more information regarding the caffeine withdrawal questionnaire, please see McNerney et al (2014a,b).

Oculomotor Tests

SP

SP testing was completed in the vertical as well as the horizontal planes. The results are displayed in Table 1. Statistical analysis did not reveal any statistically significant differences between the C and NC sessions for any of the frequencies tested, regardless of plane of stimulation.

SCs

The participants underwent SC testing in the horizontal as well as the vertical plane. The mean of the values from the left and right eyes was computed for

Table 1. Smooth Pursuit

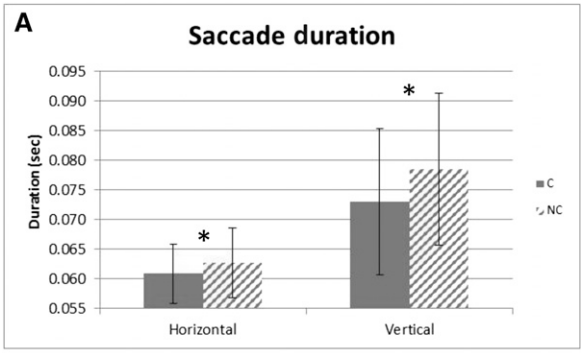
	Horizontal				Vertical		
	0.10 Hz	0.30 Hz	0.50 Hz	0.75 Hz	0.10 Hz	0.30 Hz	0.50 Hz
				C			
Gain	1.00 ± 0.02	1.00 ± 0.02	0.99 ± 0.04	0.94 ± 0.07	1.01 ± 0.03	1.01 ± 0.06	0.96 ± 0.08
Asymmetry (%)	0.61 ± 0.47	1.19 ± 1.29	1.41 ± 1.55	2.42 ± 1.88	2.08 ± 1.80	2.36 ± 2.46	4.23 ± 3.21
Phase (°)	0.42 ± 0.32	0.94 ± 0.83	2.83 ± 1.82	4.28 ± 2.89	1.64 ± 1.26	1.91 ± 2.00	3.75 ± 2.60
				NC			
Gain	1.01 ± 0.03	1.00 ± 0.03	0.99 ± 0.04	0.93 ± 0.08	1.01 ± 0.04	1.02 ± 0.09	0.96 ± 0.10
Asymmetry (%)	0.58 ± 0.54	1.17 ± 0.84	1.48 ± 1.42	2.79 ± 2.81	2.66 ± 2.12	3.39 ± 3.82	3.67 ± 3.51
Phase (°)	0.71 ± 0.82	1.29 ± 1.45	2.45 ± 1.85	4.18 ± 2.85	1.94 ± 1.64	2.00 ± 2.06	2.69 ± 2.08

Note: Mean ± SD results for SP in the horizontal (SPH) as well as the vertical (SPV) planes.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

latency, duration, amplitude, peak velocity, and accuracy. Absolute values were obtained for SC amplitude, as well as for peak velocity before the computation of the means across participants. The results from statistical analyses in the horizontal plane revealed statistically significant differences between the C and NC session for saccade duration [$t(29) = -2.13, p = 0.042$; $M_{difference} = -0.002$, $SD_{difference} = 0.005$; Cohen's $d = 0.39$] and peak velocity [$t(29) = 3.499$, $p = 0.002$; $M_{difference} = 16.78$, $SD_{difference} = 26.26$; Cohen's $d = 0.64$], whereas the results obtained in the vertical plane revealed statistically significant differences for SC duration only [$t(29) = -3.17$, $p = 0.004$; $M_{difference} = -0.006$, $SD_{difference} = 0.01$; Cohen's $d = 0.58$]. The results from SC duration and peak velocity are shown in Figures 1A and B. SC latency, amplitude, and accuracy did not reveal any statistically significant differences between the two sessions. The duration of both the horizontal and vertical SCs was slightly longer in the NC versus the C session ($\Delta = 0.002$ sec for horizontal SCs and $\Delta = 0.006$ sec for vertical SCs), and horizontal peak velocity was higher in the C versus NC session (C = 375.03°/sec versus NC = 358.26°/sec).

When the participants were allocated into the LC versus HC intake groups, a statistically significant result remained for vertical saccade duration in the LC group [$t(21) = -3.04, p = 0.006$; $M_{difference} = -0.007$, $SD_{difference} = 0.01$; Cohen's $d = 0.65$], whereas no significant differences were found in the HC group. In regard to the horizontal plane, no statistically significant results were obtained for either group. In contrast, horizontal peak velocity revealed a statistically significant difference in the LC intake group [$t(21) = 3.08$, $p = 0.006$; $M_{difference} = 17.87$, $SD_{difference} = 27.13$; Cohen's $d = 0.66$], whereas no significant differences were found in the HC group.



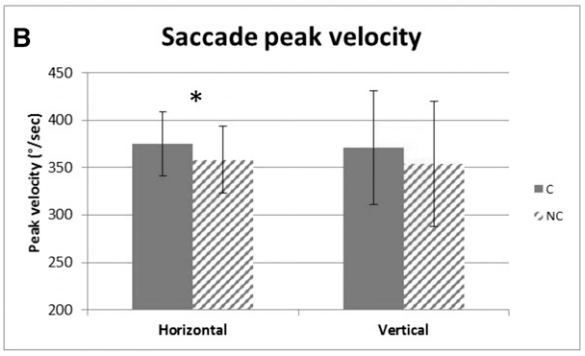


Figure 1. (A) and (B) display the results from SC duration and peak velocity in the horizontal as well as the vertical plane. Significant results are indicated by an * ($p \leq 0.05$).

Optokinetics

Eye velocity gain (normalized to 20°/sec) recorded in response to optokinetic stimuli is shown in Figure 2. Statistical analysis via paired *t*-tests revealed that OPK eye velocity gain was significantly different for the 40°/sec [$t(29) = 2.25$, $p = 0.033$; $M_{difference} = 0.04$, $SD_{difference} = 0.09$; Cohen's $d = 0.41$] and 60°/sec [$t(29) = 2.37$, $p = 0.024$; $M_{difference} = 0.04$, $SD_{difference} = 0.08$; Cohen's $d = 0.43$] stimuli. The mean data indicate that individuals had higher OPK eye velocity gain values in the C versus the NC session for all of the stimuli presented (i.e., 20, 40, and 60°/sec). When the participants were separated into an LC intake group versus an HC intake group, a statistically significant difference remained in both the 40°/sec [$t(21) = 2.8$, $p = 0.011$; $M_{difference} = 0.05$, $SD_{difference} = 0.08$; Cohen's $d = 0.60$] and the 60°/sec [$t(21) = 2.5$, $p = 0.021$; $M_{difference} = 0.05$, $SD_{difference} = 0.09$; Cohen's $d = 0.53$] conditions in the LC intake group, whereas no statistically significant differences remained in the HC intake group.

Rotary Chair Tests

SHA

During SHA, the participants were tested at frequencies of 0.02, 0.04, 0.08, 0.64, and 1.28 Hz. Average eye gain, asymmetry (%), and phase (°) are listed in Table 2. Statistical analyses did not reveal any statistically significant differences between the C and NC sessions.

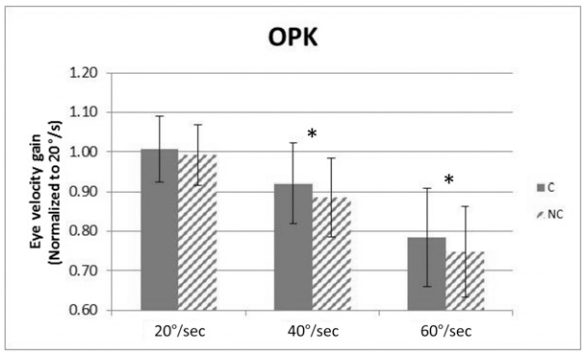


Figure 2. The average eye velocity gain (normalized to 20°/sec) from full-field optokinetic testing for 20, 40, and 60°/sec. Significant results are indicated by an * ($p \leq 0.05$).

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 2. Sinusoidal Harmonic Acceleration

	C			NC		
	Gain	Asymmetry (%)	Phase (°)	Gain	Asymmetry (%)	Phase (°)
0.02 Hz	0.44 ± 0.12	9.26 ± 6.43	22.11 ± 5.94	0.45 ± 0.11	7.90 ± 4.30	23.58 ± 6.22
0.04 Hz	0.51 ± 0.15	8.14 ± 7.26	10.15 ± 4.65	0.52 ± 0.14	6.89 ± 5.31	10.68 ± 5.16
0.08 Hz	0.53 ± 0.19	7.06 ± 5.82	3.96 ± 3.63	0.55 ± 0.18	7.71 ± 6.08	4.39 ± 3.02
0.64 Hz	0.58 ± 0.18	8.74 ± 6.57	9.51 ± 4.88	0.61 ± 0.16	8.47 ± 6.98	7.73 ± 3.90
1.28 Hz	0.90 ± 0.15	4.21 ± 5.07	14.54 ± 6.46	0.92 ± 0.13	2.79 ± 1.98	12.79 ± 6.19

Note: Mean ± SD results for SHA testing.

VE/VS

Eye gain, asymmetry (%), and phase (°) results in response to the VE test and eye gain in response to the VS test can be found in Table 3. Statistical analyses did not reveal any statistically significant differences between the C and NC sessions for any of the recorded measurements.

TST

Table 4 displays eye gain, peak velocity (°/sec), and decay time (sec) in response to the TST. Statistical analyses did not reveal any significant differences between the C and NC sessions for any of the comparisons.

SVV/SVH

Table 5 displays the results from the SVV and SVH tests. The average of the six trials from the relevant plane was computed. An additional SVV test was added to the protocol to evaluate whether SHA testing influences the results of SVV testing. The participants were evaluated on six SVV trials before undergoing SHA and six SVV trials after undergoing SHA. A paired *t*-test of the SVV results pre- versus post- SHA testing found no statistically significant differences between the two measurements. When evaluating the individual differences in degrees between the estimations made pre- versus post-SHA, the largest individual difference in the C session was 2.21°, and the largest individual difference in the NC session was 2.61°. A paired *t*-test comparing the results between the C and the NC sessions for SVV, pre- versus post-SHA, also did not reveal any statistically significant differences. In addition, analysis of the SVH results did not reveal any statistically significant differences between the two sessions.

UUC

Table 6 displays the results from SVV while the participants were undergoing UUC. The average of the SVV deviations during each shift of the chair were collected and analyzed. Comparison of the mean data from the C versus NC sessions via paired *t*-tests did not reveal any statistically significant differences between the two sessions.

DISCUSSION

Rotational chair testing provides information about vestibular function that cannot be assessed via traditional VNG testing. Although in the present study, oculomotor tests were included in the battery of tests that were administered in the rotary chair, they are typically administered during VNG testing. The only statistically significant results that were found in the present study were obtained during oculomotor testing, which included SCs and optokinetics. The remaining tests administered in the rotary chair did not reveal any statistically significant differences between the C and NC sessions. This provides support for the argument that it is not necessary to require healthy young adults to abstain from drinking caffeine before undergoing rotational chair tests.

Caffeine is absorbed relatively quickly after consumption and is circulated through the body including the brain. It has been shown that caffeine can increase neuronal activity (which can result in increased arousal and attention) by binding to A1 receptors, thereby promoting the release of neurotransmitters such as glutamate, dopamine, and acetylcholine (Einöther and Giesbrecht, 2013). Caffeine has also been shown to increase reaction time as well as accuracy in a variety of studies (Fine et al, 1994; Lorist et al, 1994; Smith et al,

Table 3. Visual Enhancement/Visual Suppression

	C			NC		
	Gain	Asymmetry	Phase	Gain	Asymmetry	Phase
VE	1.07 ± 0.08	1.19 ± 1.16	4.91 ± 1.71	1.06 ± 0.07	1.30 ± 1.17	4.42 ± 1.70
VS	0.14 ± 0.05			0.15 ± 0.06		

Note: Mean ± SD results for the VE as well as VS tests.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 4. Trapezoidal Step Test

	Pre-RR	Post-RR	Pre-RL	Post-RL
		Peak Velocity (°/sec)		
C	−65.22 ± 20.27	58.99 ± 14.24	60.19 ± 19.54	−61.18 ± 17.77
NC	−64.12 ± 15.92	63.00 ± 14.86	64.09 ± 15.15	−62.15 ± 13.19
		Decay Time (sec)		
C	15.68 ± 4.51	15.45 ± 4.14	12.94 ± 3.25	14.98 ± 3.94
NC	15.04 ± 4.43	14.63 ± 4.23	11.97 ± 3.98	15.70 ± 5.71
		Gain		
C	0.66 ± 0.20	0.59 ± 0.14	0.60 ± 0.20	0.61 ± 0.18
NC	0.64 ± 0.16	0.63 ± 0.15	0.65 ± 0.15	0.63 ± 0.13

Notes: Mean ± SD results for the TST. Pre-RR = pre-rotary right; Post-RR = post-rotary right; Pre-RL = pre-rotary left; Post-RL = post-rotary left.

1994; Haskell et al, 2005; 2008; Maridakis et al, 2009; Smith, 2009; Einöther and Giesbrecht, 2013). It is therefore not surprising that the tests that revealed a statistically significant difference between the C and NC sessions were tests of central function that have been shown to be affected by attention and fatigue (Hale et al, 2015). The results from oculomotor tests typically revealed faster responses (i.e., higher peak velocity, shorter duration) and higher gain values in the C versus the NC session. It is logical to suggest, therefore, that this is the result of the stimulant property of caffeine. Although statistically significant results were found, it is important to consider the "clinical significance" of these findings.

Normative data are not currently available in the VEST software that was utilized to collect and analyze the data from the Neuro Kinetics chair. Therefore, to determine whether there was a clinically significant difference between the sessions, data from individual participants were evaluated from each session to determine if the results fell within 2 SD of the group mean data obtained during the NC session (Table 7). For any given condition that revealed a statistically significant result, there were no more than two participants who fell outside of the 2 SD range (with the exception of the gain in response to the 60° OPK stimulus, in which three participants fell outside of the group NC session mean ±2 SD range). Further analysis of the data revealed that there were only two participants whose results fell outside of the 2 SD range for more than one condition/session. One of the participants displayed two responses that were outside of the 2 SD range during the NC session and two responses that were outside of the range during the C session. Three of the four outlier responses were obtained during the OPK testing. This particular participant was found to be a low caffeine drinker as per the guidelines stated earlier in this article, and provided low caffeine withdrawal ratings (total severity rating of 7). As oculomotor testing relies heavily on patient/participant participation/attentiveness, it is possible that this particular participant would have needed further instruction and/or addition trials to obtain accurate recordings as the results of OPK gain were consistently below the 2 SD range, regardless of caffeine intake. The other participant fell outside of normal limits for horizontal SC duration during the NC session and horizontal SC peak velocity during the C session. This particular participant was also categorized as a low caffeine drinker and provided low caffeine withdrawal ratings (total severity rating of 1). This is the only participant who fell outside of the 2 SD range for the horizontal SCs. As the horizontal peak velocity for this participant was more than 2 SD below the mean in the "caffeine" condition, and given the known stimulant properties of caffeine, it is unlikely that the results are because of the ingestion of caffeine. If this observation was likely to be due to the stimulant properties of caffeine and not to random variation in the data, we would have expected an increase in peak velocity during the C session and not the observed decrease. In summary, despite the statistically significant differences displayed during several tests of oculomotor function (Figures 1 and 2), there were no clinically significant differences found when comparing the results from both sessions. This would support the conclusion that it is not necessary to require that healthy young adults abstain from ingesting caffeine before undergoing tests of oculomotor function administered in the rotary chair.

Table 5. Subjective Visual Vertical/Horizontal

	C	NC
SVV—Pre-SHA	−0.11° ± 1.71°	−0.53° ± 1.49°
SVV—Post-SHA	−0.10° ± 1.56°	−0.32° ± 1.13°
SVH	−1.07° ± 1.10°	−0.65° ± 1.05°

Note: Mean ± SD for the SVV and SVH tests.

Table 6. Unilateral Utricular Centrifugation

	R	C	L
C	−2.07° ± 3.49°	0.24° ± 2.61°	3.00° ± 3.69°
NC	−2.41° ± 4.33°	0.45° ± 2.89°	3.16° ± 4.30°
	0.33°	−0.21°	−0.16°

Note: Results from SVV while the patient was undergoing UUC, when positioned to the right (R), center (C), and left (L).

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 7. Oculomotor Test Analysis

NC				
SCs			OPK gain (normalized to 20°/sec)	
	Horizontal	Vertical		
Duration (sec)	Peak Velocity (°/sec)	Duration (sec)	40°/sec	60°/sec
S#2(+)	—	S#30(+)	S#8(−), S#30(−)	—
C				
SCs			OPK (normalized to 20°/sec)	
	Horizontal	Vertical		
Duration (sec)	Peak Velocity (°/sec)	Duration (sec)	40°/sec	60°/sec
—	S#2(−)	—	S#30(−)	S#6(+), S#21(+), S#30(−)

Notes: Display of the participants who were outside of the 2 SD range for each of the oculomotor tests which revealed a statistically significant difference between the C and NC sessions. An (−) indicates that no participant fell outside of the 2 SD range. A (+) or (−) after the participant number indicates whether the participant fell above or below the 2 SD range.

Further support of the above conclusion is provided through the comparison of the statistical analysis of the results from the LC versus HC groups. When statistically significant results were found between the C and NC sessions for any given measure, the participants were then separated into two groups based on the amount of caffeine they consumed per day. The LC group consisted of 22 individuals, whereas the HC group consisted of eight individuals. Ideally, we would have liked to have equal or near-equal numbers for the LC and HC groups. As we did not assess the amount of caffeine that each individual consumed on a daily basis before enrolling them in the study, we were only able to obtain this information from the caffeine diary that the participants were asked to complete between the C and NC sessions. Despite this, the results revealed that statistically significant differences remained in the LC intake group, whereas no statistically significant differences remained in the HC intake group. It is possible that the lack of significance for the HC intake group was because of a lack of statistical power. However, individuals who do not drink caffeine would not likely start to on the day of the test (i.e., those in the LC group), and individuals who normally ingest moderate to high amounts of caffeine on a daily basis would likely continue with their normal routine (i.e., those in the HC group). This again strengthens the argument that it does not appear necessary to require that healthy young adults abstain from drinking caffeine before undergoing tests of oculomotor function administered in the rotary chair.

CONCLUSIONS

The present study evaluated whether the results from rotary chair tests are influenced by whether an individual ingests caffeine before undergoing testing. Given that statistically significant results were found during tests that are typically completed during a VNG evaluation but not for tests which require a rotary chair to be administered (i.e., SHA, VS/VE, and TST), it does not support the requirement of having young healthy adults refrain from drinking caffeine before undergoing rotary chair testing. In addition, although some of the results from oculomotor testing did reveal some statistically significant differences, none of the changes displayed would be classified as "clinically significant" changes in results. Future research is necessary to determine if the same results would occur in individuals who have been diagnosed with a vestibular impairment, as well as in older individuals. Also, these results are only generalizable to those who consume a moderate dose of caffeine before vestibular assessment. A dose/response curve of caffeine consumption for the various tests of vestibular and oculomotor function might be enlightening.

REFERENCES

Barin K. (2008) Background and technique of caloric testing. In: Jacobson GP, Shephard NT, eds. *Balance Function Assessment and Management*. San Diego, CA: Plural Publishing.

BayCare Clinic's Ear, Nose, and Throat. Retrieved from http://www.baycare.net/media/133993/ent_eng_instructions.pdf. Accessed August 21, 2012.

Brey RH, McPherson JH, Lynch RM. (2008) Technique, interpretation, and usefulness of whole body rotational testing. In: Jacobson GP, Shephard NT, eds. *Balance Function Assessment and Management*. San Diego, CA: Plural Publishing.

Bronstein AM. (2008) Tests of otolith function and vestibular perception. In: Jacobson GP, Shepard NT, eds. *Balance Function Assessment and Management*. San Diego, CA: Plural Publishing, 435–446.

Caffeine Informer. (2014) The Complete Guide to Starbucks Caffeine. http://www.caffeineinformer.com/the-complete-guide-to-starbucks-caffeine. Accessed August 29, 2014.

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

Dernaro C, Benowitz N. (1991) Caffeine metabolism: disposition in liver disease and hepatic-function testing. Totowa, NJ: Humana Press, 513–539.

Einöther SJL, Giesbrecht T. (2013) Caffeine as an attention enhancer: reviewing existing assumptions. *Psychopharmacology (Berl)* 225(2):251–274.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

ENTCare. Retrieved from http://www.entcare.org/PatientEducation/PatientEducation/Electrostagmography.html. Accessed August 21, 2012.

Fine BJ, Kobrick JL, Lieberman HR, Marlowe B, Riley RH, Tharion WJ. (1994) Effects of caffeine or diphenhydramine on visual vigilance. *Psychopharmacology (Berl)* 114(2):233–238. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=7838913.

Hale T, Trahan H, Parent-Buck T. (2015) Evaluation of the patient with dizziness and balance disorders. In: Katz J, Chasin M, English K, Hood L, Tillery K, eds. *Handbook of Clinical Audiology*. Philadelphia, PA: Wolters Kluwer, 408.

Haskell CF, Kennedy DO, Milne AL, Wesnes KA, Scholey AB. (2008) The effects of L-theanine, caffeine and their combination on cognition and mood. *Biol Psychol* 77(2):113–122. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=18006208.

Haskell CF, Kennedy DO, Wesnes KA, Scholey AB. (2005) Cognitive and mood improvements of caffeine in habitual consumers and habitual non-consumers of caffeine. *Psychopharmacology (Berl)* 179(4):813–825. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=15678363.

International Coffee Organization. (2012) www.ico.org. Accessed August 21, 2012.

Lorist MM, Snel J, Kok A, Mulder G. (1994) Influence of caffeine on selective attention in well-rested and fatigued subjects. *Psychophysiology* 31(6):525–534 http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=7846213.

Maridakis V, O'Connor PJ, Tomporowski PD. (2009) Sensitivity to change in cognitive performance and mood measures of energy and fatigue in response to morning caffeine alone or in combination with carbohydrate. *Int J Neurosci* 119(8):1239–1258. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=19922353.

Mayo Clinic. (2014) Caffeine: how much is too much? http://www.mayoclinic.com/health/caffeine/nu00600. Accessed July 30, 2014.

McCusker RR, Goldberger BA, Cone EJ. (2003) Caffeine content of specialty coffees. *J Anal Toxicol* 27(7):520–522.

McNerney K, Coad ML, Burkard R. (2014a) The influence of caffeine on calorics and cervical vestibular evoked myogenic potentials (cVEMPs). *J Am Acad Audiol* 25(3):261–267.

McNerney KM, Coad ML, Burkard RF. (2014b) The influence of caffeine on the sensory organization test. *J Am Acad Audiol* 25(6):521–528.

Nolan S, Heinzer T. (2011) *Essentials of Statistics for the Behavioral Sciences*. 1st ed. New York, NY: Worth Publishers.

Ozsungur S, Brenner D, El-Sohemy A. (2009) Fourteen well-described caffeine withdrawal symptoms factor into three clusters. *Psychopharmacology* 201(4):541–548.

Rubin W. (1982) Harmonic acceleration tests as a measure of vestibular compensation. *Ann Otol Rhinol Laryngol* 91(5 Pt 1):489–492. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med2&AN=6982652.

Shepard NT. (2009) Evaluation of the patient with dizziness and balance disorders. In: Katz J, Medwetsky L, Burkard R, Hood L, eds. Handbook of Clinical Audiology. 6th ed. Baltimore, MD: Wolters Kluwer | Lippincott Williams & Wilkins, 467–496.

Smith A. (2009) Effects of caffeine in chewing gum on mood and attention. *Hum Psychopharmacol* 24(3):239–247.

Smith A, Kendrick A, Maben A, Salmon J. (1994) Effects of breakfast and caffeine on cognitive performance, mood and cardiovascular functioning. *Appetite* 22(1):39–55. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=8172489.

Wilstar. (2014) Caffeine content of popular drinks. http://wilstar.com/caffeine.htm. Accessed July 30, 2014.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Efficient Use of Vestibular Testing

Steven A. Zuniga, MD*, Meredith E. Adams, MD, MS

KEYWORDS

• Balance • Dizziness • Vertigo • Vestibular disorders • Vestibular testing

KEY POINTS

- The majority of vestibular disorders may be readily diagnosed from a thorough clinical history and physical exam.
- Objective assessment of vestibular function may provide insights that facilitate diagnostic refinement and inform management decisions in specific clinical scenarios.
- An understanding of the data provided by vestibular tests, as well as their inherent limitations, enables the efficient utilization of these test modalities.

INTRODUCTION

The dizzy patient presents a formidable diagnostic challenge. Dizziness is one of the most common symptoms prompting clinical evaluation.[1] Afflicted patients experience poor quality of life, risk of falling with injury, and frequently require sick leave, job changes, or disability.[1–4] Clinicians must consider a wide range of benign to life-threatening etiologies within the time constraints of modern practice. The demands are compounded by the absence of a comprehensive clinical practice guideline for the dizziness diagnostic evaluation.

The admixture of distressed, at-risk patients, diagnostic uncertainty, and time pressure generates an explosion of test acquisition.[5–7] Vestibular tests consistently rank among the most billed office procedures associated with otolaryngologic practices in the United States,[8] and are also frequently obtained by other specialists and primary care clinicians.[9] However, there is no "standard practice," and vestibular test use varies markedly between geographic regions, medical specialties, and individual practices.[9–11]

Choosing Wisely Canada, a campaign to reduce unnecessary tests and treatments through the development of common-sense guidelines, specifically recommends, "[d]on't order specialized audiometric and vestibular neurodiagnostic tests in an attempt

Department of Otolaryngology—Head and Neck Surgery, University of Minnesota, 420 Delaware Street, MMC 396, Minneapolis, MN 55455, USA
* Corresponding author.
E-mail address: zunig085@umn.edu

Otolaryngol Clin N Am 54 (2021) 875–891
https://doi.org/10.1016/j.otc.2021.05.011
oto.theclinics.com
0030-6665/21/© 2021 Elsevier Inc. All rights reserved.

Abbreviations	
BPPV	Benign paroxysmal positional vertigo
cVEMP	Cervical VEMP
GEN	Gaze-evoked nystagmus
IVN	Inferior vestibular nerve
oVEMP	Ocular VEMP
PT	Physical therapy
SCDS	Superior semicircular canal dehiscence
SCV	Slow component velocity
SHA	Sinusoidal harmonic acceleration
SPV	Slow phase velocity
SVN	Superior vestibular nerve
VEMP	Vestibular evoked myogenic potential
VNG	Videonystagmography
VOR	Vestibulo-ocular reflex

to screen for peripheral vestibular disease."[12] Rather, the diagnosis should be guided by the presenting symptoms and office examination, and tests "should only be ordered if clinically indicated."[12] Accordingly, this article aims to help clinicians apply an accessible decision-making rubric to identify clinical scenarios that may and may not benefit from data derived from specific vestibular function tests.

MAKING WISE CHOICES

The etiology of dizziness can often be identified with a detailed history and focused physical examination without additional testing.[13–17] To illustrate, predictive accuracy of 78.5% for the final diagnosis was achieved by a previsit questionnaire that differentiated dizziness causes by episode description, symptoms (including auditory) characteristic of peripheral versus central etiologies, and general and emotional health.[18] Diagnostic criteria for common disorders, including Meniere's disease and benign paroxysmal positional vertigo (BPPV), do not include vestibular test data.[19,20] Best practices for using history and exam for dizziness diagnosis are presented elsewhere in this edition.

We also need to count the costs of our diagnostic approaches. Beyond the obvious monetary costs borne by patients, systems, and payors,[9,10] patients undergoing testing will experience morbidity and opportunity costs.[21] For example, videonystagmography (VNG) with caloric testing can induce nausea/vomiting, headaches, and other residual symptoms that preclude a return to work or activities of daily living for several days.[21] These tests may also require the involvement of additional family members (for the drive home) and finding child care alternatives, thus creating collateral social and financial burdens.

How then does a clinician determine if vestibular testing is clinically indicated? The decision-making checklist developed by Dr. William Follansbee[22] is particularly helpful in this regard. Prior to obtaining testing, a differential diagnosis is generated from the symptoms and exam. To determine what tests to order, *if any*, clinicians should: (1) define the specific question they are asking with the test; (2) determine if they truly need to know the answer because it will refine the diagnosis and/or affect patient management or outcomes; (3) identify the test that will best answer the specific question.

What Questions May Vestibular Tests Answer?

1. Site of Lesion. Vestibular tests may localize lesions to central versus peripheral vestibular pathways and by laterality and topography (eg, superior vs inferior vestibular nerve, otolithic organ vs semicircular canal).

2. Extent of Lesion. Vestibular tests may quantify the severity of vestibular hypofunction over the vestibulo-ocular reflex (VOR) frequency range, characterizing disease stage/progression to inform treatment decisions (eg, ablation).
3. Level of Compensation. Following acute vestibular loss, vestibular tests can assess compensation status to direct uncompensated individuals to vestibular physical therapy (PT)[23] and track progress over time.
4. Functional Integration of Sensory Inputs. A complex network of sensory inputs must be successfully integrated to maintain stance and gait. Tests including computerized dynamic posturography assess visual, vestibular, and postural contributions to balance.

What Questions May Vestibular Tests NOT Answer?

Vestibular tests offer more insight into vestibular physiology (eg, VOR function) than pathophysiology and do not typically provide a specific diagnosis (eg, Meniere's disease).[24] For example, VNG reveals nystagmus generated by asymmetric peripheral vestibular system stimulation, but supplementary information is required to elucidate the cause. Vestibular tests also do not determine the level of disability resulting from vestibular impairment. Patient-reported disability measures (eg, dizziness handicap inventory) and vestibular test results are poorly correlated,[25] potentially due to variations in physiologic and behavioral adaptation.

VESTIBULAR PHYSIOLOGY: KEY CONCEPTS FOR TEST SELECTION AND INTERPRETATION

Visual–Vestibular Interaction

Vestibular tests interrogate the function of the vestibular system largely through the measurement and interpretation of eye movements. A common goal of eye movements is to optimize visual acuity by directing and maintaining an object of interest on the fovea.[26] When the head is static and the object moves, central visual tracking systems are used. The *saccade* control system generates fast voluntary and involuntary eye movements that focus objects in the visual periphery on the fovea.[26] The *smooth pursuit* system maintains images of small, slowly moving objects on the fovea, driven by retinal slip from visual motion.[24] The peripherally-mediated VOR maintains fixation on a stable target during head movements by generating eye movements that are equal in velocity but opposite in direction to head movements (**Fig. 1**). During simultaneous movement of the head and visual surround, a complex interaction between vestibular and visual ocular control systems is necessary.

Nystagmus is generated by a combination of peripherally and centrally mediated eye movements.[26] The VOR holds images on the retina by producing reflexive compensatory eye movements during head rotations, but VOR-driven eye rotation is limited by anatomic constraints. As the eyes approach their orbital limit, central processes "reset" the position of the eyes, quickly moving them in the opposite direction and directing the gaze toward the oncoming visual scene. The slow phase of nystagmus is driven by tonic asymmetry in the neural activity of the vestibular system, and its velocity (degrees/s) is a common vestibular test outcome measure. The centrally generated fast phase movement is more discernible to observers and is used to name the direction of nystagmus (see **Fig. 1**).

Contributions from the Brainstem and Cerebellum

The VOR is maximally efficient and demonstrates nearly perfect gain (ie, eye velocity approximates head velocity) at frequencies between 0.05 and 6 Hz (**Fig. 2**).[27] VOR efficiency declines at higher and lower frequencies. Low-frequency transduction is

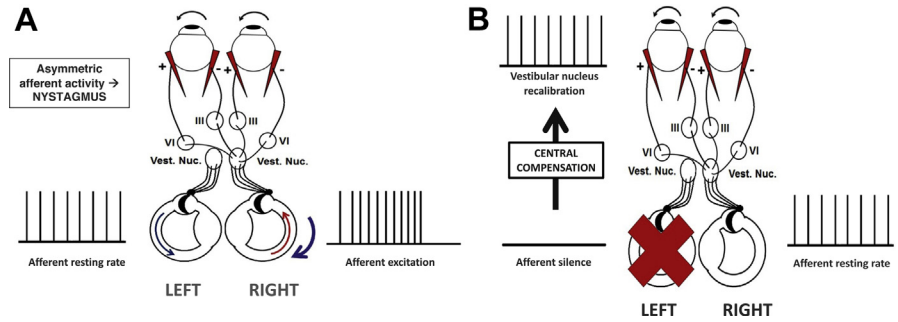


Fig. 1. Horizontal Vestibulo-ocular reflex (VOR). (*A*) While the head is at rest, vestibular afferents demonstrate an approximately symmetric baseline firing rate of 10 to 100 action potentials per second. This tonic neural activity affords bidirectional sensitivity to the system; excitatory head movements increase and inhibitory head movements decrease the firing rate, respectively.[24] The direction of cupular deflection determines whether action potential frequency increases or decreases with head motion, and the complementary arrangement of the canals ensures that an increase in firing rate for one results in a decrease in its coplanar mate. Stimulation of the horizontal semicircular canal produces eye movements in the plane of that canal (Ewald's first law) that are equal in velocity but opposite in direction to the associated head movement. In this case, sustained rightward head motion stimulates and inhibits the right and left horizontal canal afferents, respectively, resulting in leftward slow phase eye movement (*curved arrows*), and right beat (fast phase) nystagmus. (*B*) Acute left peripheral hypofunction creates a similar asymmetry in vestibular afferent firing rates with corresponding eye movements. Central static compensation occurs following an acute vestibular lesion and involves tonic rebalancing of the resting activity of the vestibular nuclei. This minimizes the tonic firing rate asymmetry in the second-order neurons originating in the vestibular nuclei.

improved by the perseveration of raw rotational signals by brainstem vestibular nuclei by the central *velocity storage* mechanism.[28] Velocity storage may be transiently diminished with acute unilateral vestibulopathy, leading to prerotary and postrotatory nystagmus (rotary chair) and head-shake nystagmus.

The *neural integrator* is a brainstem mechanism that allows the eyes to be held in an eccentric position for visual acuity instead of drifting back to neutral gaze position from elastic restoring forces of the orbit.[28] With acute unilateral vestibulopathy, the neural

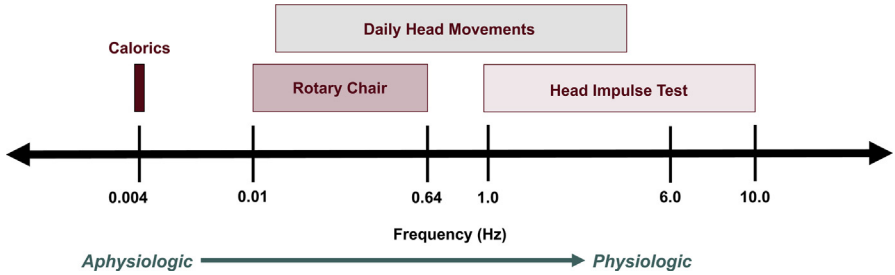


Fig. 2. Vestibulogram. In isolation, the VOR optimally functions between head rotation frequencies of 0.05 and 6 Hz, relying on supplementary central processing to improve function across its dynamic range (ie, velocity storage). Clinicians may assess the integrity of the horizontal VOR across its frequency range using caloric testing (low frequencies), rotary chair testing (low to middle frequencies), and head impulse testing (high frequencies).

integrator's ability to hold eccentric gaze is diminished (ie, unilaterally inhibited), and the eyes drift back to a central position. Corrective saccades are needed to maintain eye position, generating gaze-evoked nystagmus (GEN). The combination of GEN and VOR-driven nystagmus from a peripheral vestibular lesion results in increased nystagmus intensity when looking away from the lesion (ie, toward the nystagmus fast phase) and decreased intensity when looking toward the lesion. This nystagmus intensifies when visual fixation is denied. The effect is known as *Alexander's law* and is useful in distinguishing nystagmus of peripheral from central origin[29] and assessing for compensation.

Adjustment to asymmetries in peripheral vestibular input or compensation for insults within the central vestibular pathways relies on the adaptive plasticity of the cerebellum and brainstem nuclei.[30] Vestibular compensation occurs in 2 stages: (1) static compensation, which occurs in the absence of head movements and involves recalibration of the resting tonic activity of the vestibular nuclei; and (2) dynamic compensation, which is driven by persistent disequilibrium and motion-provoked vertigo.[24]

VIDEONYSTAGMOGRAPHY

VNG employs high-speed infrared cameras and sophisticated algorithms to record and measure eye movements in response to visual or vestibular stimuli. The VNG battery interrogates the oculomotor and vestibular systems and detects pathologic (spontaneous, gaze, positional, and positioning) nystagmus. VNG can be used to identify the site and extent of vestibular lesions and compensation status.

Oculomotor Testing (Central Lesions)

The oculomotor test battery interrogates central pathways responsible for generating voluntary and involuntary eye movements including saccades and smooth pursuit. The head is static during these tests permitting oculomotor system assessment *independent* of the peripheral vestibular system. The timing, speed, and accuracy of eye movements are compared with those of visual target stimuli. Abnormalities of saccades or smooth pursuit represent a dysfunction of neurologic substrates originating anywhere from supranuclear central control centers through the extraocular muscles.[26] Abnormalities of latency and velocity primarily result from dysfunction in the pontine reticular formation and brainstem structures[31]; abnormalities of accuracy often result from dysfunction of the vestibulocerebellum.[32] Oculomotor tests are influenced by age, alertness, medications/substances, and ophthalmologic pathology.[24] The reader is referred to Leigh and Zee[26] for a detailed analysis of neural pathways underlying oculomotor abnormalities.

Spontaneous Eye Movements (Central or Peripheral Lesions)

Eye movements are recorded during visual fixation on a static target and in darkness with eyes open, removing fixation, in midline gaze (primary position) and gazing 30° off midline in each direction (left, right, up, down).[33] Persistent nystagmus in place of steady gaze is considered to be abnormal. *Spontaneous nystagmus* may be observed while the patient looks straight ahead. Nystagmus unobserved in the primary position that appears with eccentric gaze is *GEN*. *End-gaze nystagmus* manifests as small-amplitude nystagmus when gazing more than 30° from midline and is seen in normal individuals.[33]

Nystagmus may be of central or peripheral origin. *Ewald's first law* describes the stereotypic eye movements resulting from semicircular canal stimulation. This canal-fixed frame of reference predicts the ocular movements the peripheral system

is capable of producing (Fig. 3). Peripheral nystagmus typically has a horizontal component and is direction-fixed, while pure vertical, pure torsional, or direction-changing nystagmus is of central origin until proven otherwise. In contrast to central nystagmus, peripheral nystagmus intensifies when fixation is removed and should follow Alexander's law.[24] Bilateral superior canal dehiscence can result in simultaneous bilateral canal stimulation, representing a potential peripheral source of transient down-beat nystagmus.[30]

Gaze testing may also reveal abnormal eye movements with fixation that have equal velocity and amplitude in all directions, collectively referred to as saccadic intrusions and oscillations,[26] and implicate brainstem or cerebellar pathology.[24] Leigh and Zee (2015)[26] comprehensively review their pathogenesis.

Positional and Positioning Testing

Position*al* tests change labyrinthine orientation to gravity, uncovering imbalances in peripheral or central neural pathways. Nystagmus is recorded in static supine/pre-caloric, head right/left, and/or body right/left positions. Representing the most common VNG finding,[24] positional nystagmus is abnormal if it is present in more than half of the positions, direction changing, or average peak slow phase velocity (SPV) exceeds 4°/sec.[33] It may be localizing when considered with other VNG results (discussed below).

Dynamic position*ing* tests include tests for BPPV under video-oculography. The diagnosis of BPPV is clinical, based on history and positive Dix-Hallpike or roll tests, and does not require VNG.[34] VNG recordings may facilitate challenging diagnoses (eg, laterality of horizontal canalithiasis or cupulolithiasis). Vestibular testing is warranted when patients with positional vertigo have atypical nystagmus, another vestibular pathology is suspected, following frequent BPPV recurrences, or failure of repositioning maneuvers.[34]

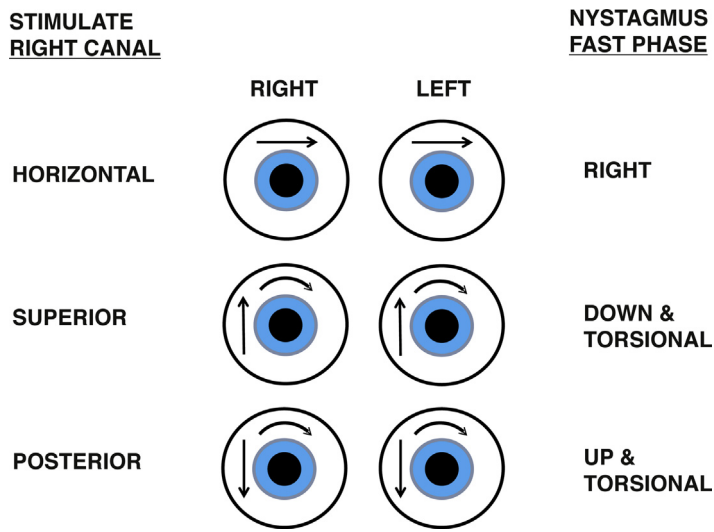


Fig. 3. Compensatory eye movements in response to canal stimulation (right ear). Arrows indicate slow phase compensatory eye movements. In considering whether nystagmus is of central or peripheral origin, it is important to consider whether the peripheral vestibular system is capable of producing the observed pattern of nystagmus.

Caloric Testing (Peripheral Lesions)

Bithermal caloric testing has long been the consensus standard for quantitative evaluation of peripheral vestibular function. Patients lie supine, head elevated 30°, placing the horizontal canal in the earth-vertical plane. Cool and warm irrigations of water or air are alternately delivered to each ear canal.[24] While preventing visual fixation, eye movements are recorded during and after each irrigation. Irrigations alternately stimulate (warm) or inhibit (cool) the horizontal canal VOR in an aphysiologic frequency range (0.003–0.008 Hz) (see **Fig. 2**).[35]

The peak SPV of caloric-induced nystagmus is used to generate the main outcome measures of the caloric test.[36] The *caloric weakness* is a comparison of responses between right and left ears. Responses are compared between ears because absolute responses to calorics are highly variable between individuals. The *directional preponderance* is a comparison of responses to irrigations yielding right versus left beat nystagmus and is analogous to the asymmetry outcome in rotary chair.

What do we learn from an abnormal caloric test? Caloric weaknesses result from lesions along the horizontal VOR pathway, including the horizontal semicircular canal, superior vestibular nerve, vestibulocochlear nerve root entry zone, and vestibular nucleus. Although exact localization is difficult, calorics isolate 1 vestibular periphery from the other,[35] defining laterality (sidedness) and quantifying the severity of vestibular hypofunction based on the percent weakness.

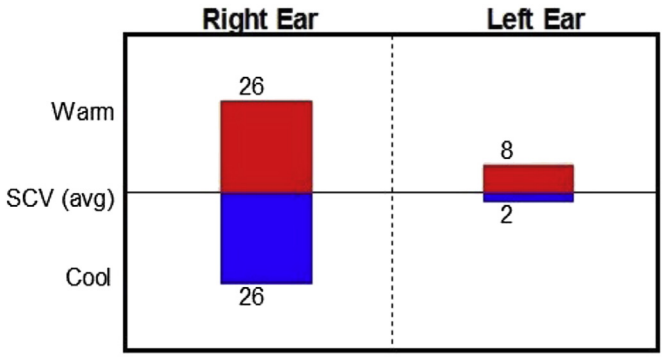
When unilateral weakness is identified, spontaneous and positional nystagmus tests help clinicians assess compensation status. On these tests, peripheral nystagmus follows Alexander's law and disappears with static compensation (**Figs. 4** and **5**). Persistent spontaneous nystagmus of greater than 2 to 3°/s suggests incomplete compensation[24] that may improve with therapy. Directional preponderance most often results from spontaneous nystagmus (eg, right beat nystagmus produces right DP), which characterizes acute uncompensated vestibular hypofunction or irritative lesions.[24]

Bilaterally reduced or absent horizontal VOR function is suggested when the total eye speed (ie, the sum of warm and cool caloric responses) per ear is less than 6°/sec.[37] However, rotary chair or vHIT is needed to determine whether there is a functionally significant residual function at higher physiologic frequencies.

What do we learn from a normal caloric test? Normal caloric results do not "rule out" vestibular disease. Calorics stimulate the horizontal canal VOR; lesions of the otolith organs, vertical canals, or other neural pathways (eg, superior semicircular canal dehiscence (SCDS), inferior vestibular neuritis) would not produce caloric asymmetry.[24] Further, peripheral disorders do not always cause canal hypofunction (eg, Meniere's disease).[38,39] Laboratories set cutoffs for unilateral weakness to avoid false positives, but the level of asymmetry sufficient to produce symptoms is unknown. A milder asymmetry may be significant in a symptomatic patient.[24]

ROTARY CHAIR TESTING

Rotary chair testing is a calibrated, midfrequency test of the horizontal VOR and its superior vestibular nerve (SVN) afferents (see **Fig. 2**). The common test paradigms are sinusoidal harmonic acceleration (SHA) and step testing. In SHA protocols, the chair rotates back and forth at frequencies 0.01 to 0.64 Hz at a set velocity.[24] Step testing consists of chair acceleration to sustained velocity followed by equal deceleration to a full stop[24]; the test is repeated in the opposite direction. Outcome measures are gain, phase or time constant (for SHA and step testing, respectively), and symmetry. Gain refers to the ratio of peak eye velocity to peak head velocity. Phase represents the



Outcome Measure	Formula	Calculated Values	Normal values
Unilateral Weakness	$\frac{(RW+RC)-(LW+LC)}{RW+RC+LW+LC} \times 100\%$	68%	≤25%
Directional Preponderance	$\frac{(RW+LW)-(RC+LC)}{RW+RC+LW+LC} \times 100\%$	10%	≤30%

Fig. 4. VNG results from a patient with persistent disequilibrium 3 months following acute vestibular syndrome, consistent with vestibular neuritis. Maximal slow component velocity (SCV) resulting from each caloric irrigation is presented (red, warm irrigation; blue, cool irrigation). Calculation of key outcome measures is illustrated, and normative values presented. Findings are consistent with left peripheral hypofunction. LC, left cool; LW, left warm; RC, right cool; RW, right warm; SCV, slow component velocity.

difference in time between the head and eyes reach peak velocity. Symmetry is a comparison of slow phase eye velocity during rightward versus leftward movements.

While some labs routinely employ rotary chair for vestibular deficit detection,[40] our decision-making rubric favors selective use. Rotational tests activate the VOR in a physiologic manner but, during rotation, the simultaneous push–pull (ie, excitation–inhibition) of semicircular canal pairs limits inferences about unilateral deficits and laterality. One intact labyrinth is sufficient to maintain normal VOR gain. Rotary chair is most useful for confirming and quantifying the severity of *bilateral* vestibular

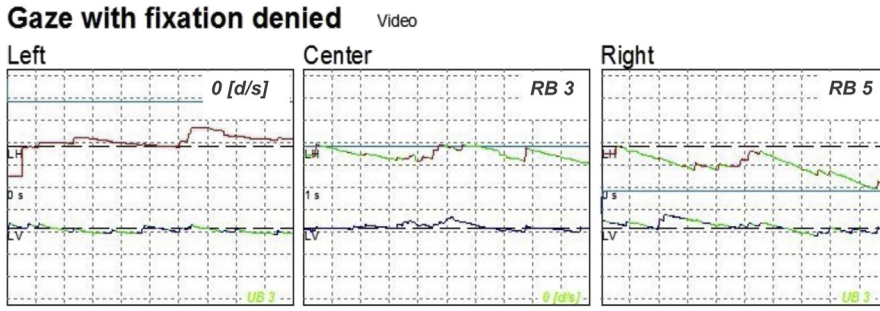


Fig. 5. Gaze testing in a patient with persistent disequilibrium 3 months following acute vestibular syndrome. Caloric testing revealed a 70% left caloric weakness. VNG tracings during gaze testing show spontaneous right beat nystagmus in center gaze (3 deg/s) that follows Alexander's law, intensifying to 5 deg/s when looking toward the fast phase (right). This is consistent with incomplete compensation. d/s = degrees/s; RB = Right Beat.

hypofunction. Bilateral loss of VOR function manifests as reduced VOR gain (<0.1 on SHA) and phase leads greater than 68° or time constant less than 5 seconds[37] (**Fig. 6**). Test results can reveal ototoxicity and direct PT, as verifying intact or reduced but viable VOR gain at high frequencies informs vestibular rehabilitation.[24] A secondary use of rotary chair is the assessment of compensation. Unilateral losses may manifest as borderline low gain and slight phase leads,[24] but these findings are not localizing. Rather, like spontaneous nystagmus on VNG, the asymmetry value reflects compensation, pointing to either the weaker side or an irritative lesion. Rotary chair is costly but may be used in populations poorly suited for calorics (eg, children) or vHIT (eg, limited neck range of motion).[37]

VIDEO HEAD IMPULSE TEST

The head impulse test is a bedside test of VOR integrity.[41] The patient is instructed to fixate on a stable visual target while the examiner applies small amplitude, high peak velocity head rotations in each of the paired semicircular canal planes. During vHIT, similar impulses are administered while patients wear tight-fitting video-oculography goggles outfitted to simultaneously measure eye and head velocity.[24]

The primary vHIT outcome parameters are gain (eye velocity/head velocity) and the response profile (saccades). Normal VOR function drives eye movement to keep up with head movement, producing a gain near 1.0. With VOR impairment, the eyes initially move *with* the head until the individual generates a refixation saccade to correct eye position back onto the target. This manifests as reduced gain and repeatable saccades after the impulse (overt) and/or during the impulse (covert) (**Fig. 7**).[42,43]

With its high-frequency physiologic stimulus (1-6 Hz) (see **Fig. 2**), vHIT affords repeatable, relatively quick, quantitative measures of VOR function for the 6 semicircular canals and their corresponding vestibular afferents, permitting more granular lesion localization.[44,45] The diagnostic accuracy of vHIT for detecting a VOR deficit (defined as a gain of <0.68) has been validated for horizontal[44] and vertical[45] semicircular canal hypofunction, with sensitivity and specificity (compared with a scleral search coil technique) as high as 100%.[46] Importantly, caloric testing may be abnormal while vHIT remains normal, a phenomenon commonly seen in Meniere's disease,[47] suggesting vHIT is complementary to calorics. vHIT assesses dynamic semicircular canal function at high frequencies (but does not reveal diagnoses),[48] and may serve as a first test of VOR function. Additional testing at lower frequencies may be indicated based on results (eg, normal vHIT despite symptoms). While the investigation is ongoing, vHIT may also assess compensation following vestibular loss, with better outcomes being noted for patients who demonstrate conversion from overt to covert saccades following a vestibular loss.[24,49,50]

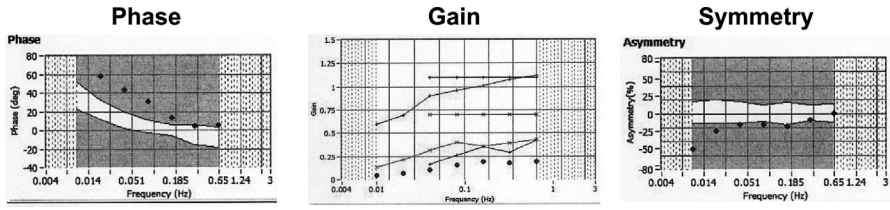


Fig. 6. Rotary chair results for a patient with imbalance and oscillopsia 6 weeks following therapy with intravenous gentamicin. There is a pronounced phase lead and low gain values across the frequency range, consistent with bilateral vestibular hypofunction.

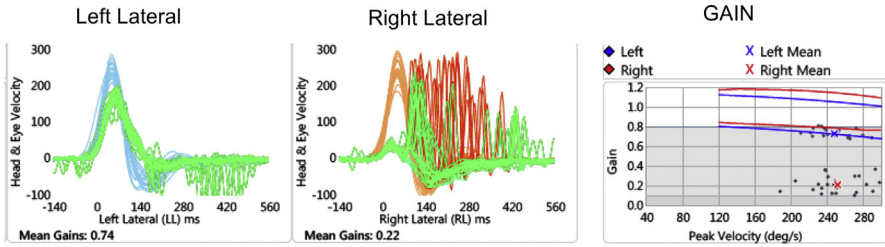


Fig. 7. vHIT response profile and gain measurements for the horizontal canal from a patient with episodic vertigo. On the right, eye movements (*green*) fail to keep up with head movements (*orange*), resulting in low gain (0.22) and corrective saccades (*red*) as the patient refixates on the target. The left ear demonstrates higher gain than the right (0.74) and no corrective saccades.

VESTIBULAR-EVOKED MYOGENIC POTENTIALS

VEMPs assess otolith function. Repetitive sound stimuli are sequentially applied to each ear. Resultant electromyographic (EMG) activity is recorded in target muscles. Cervical VEMP (cVEMP) measures relaxation potentials in the ipsilateral contracted sternocleidomastoid, generated primarily by the saccule and inferior vestibular nerve

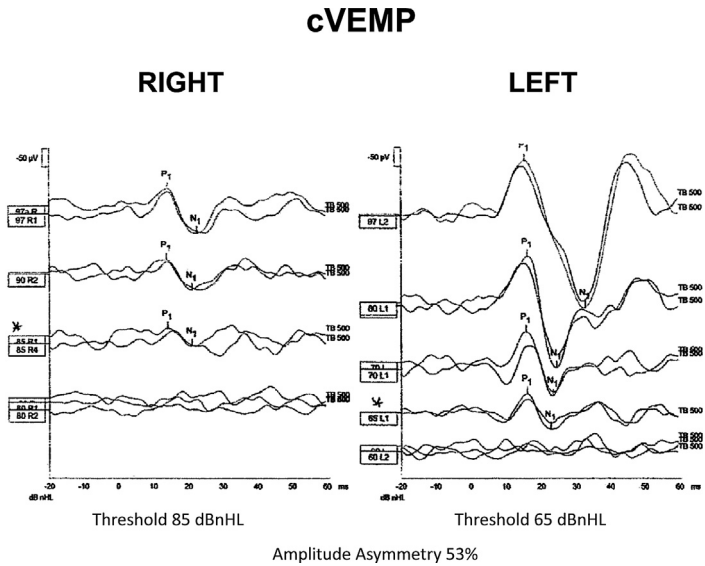


Fig. 8. Cervical vestibular evoked myogenic potential (cVEMP) response profile for right and left ears in a patient with left superior semicircular canal dehiscence syndrome. The stimulus threshold for the right ear is reduced (65 dBnHL) and an abnormal amplitude asymmetry (53%) is present between the right and left ears. Using a cVEMP cutoff threshold value of 85 dBnHL or less has been demonstrated to result in a sensitivity of 86% and a specificity of 90%.[61] (*Data from* Zuniga MG, Janky KL, Nguyen KD, Welgampola MS, Carey JP. Ocular Versus Cervical VEMPs in the Diagnosis of Superior Semicircular Canal Dehiscence Syndrome. *Otol Neurotol*. 2013;34(1):121-126. https://doi.org/10.1097/MAO.0b013e31827136b0.)

Table 1
Key tests, applications, and findings of common clinical questions from differential diagnosis

		Vestibular Function Tests and Pertinent Findings			
Clinical Scenarios	**Sample Clinical Questions**	**VNG**	**Rotary Chair**	**vHIT**	**VEMP**
Diagnostic refinement					
Acute vestibular syndrome	Is there a unilateral peripheral vestibular lesion?	**Caloric Asymmetry** **Spontaneous, gaze, or positional nystagmus obeys Alexander's law**	(+/−) Low gain Phase lead	**Reduced ipsilesional VOR gain** **Ipsilesional corrective saccades**	Absent/reduced oVEMP (SVN lesion) or cVEMP (IVN lesion)
	Has vestibular compensation occurred?[a]	**Spontaneous, gaze, positional, post head shake nystagmus**	**Asymmetry**	(+/−) Conversion of overt to covert saccades	NA
Episodic vestibular syndrome	Positional vertigo: BPPV vs central positional vertigo? Which canal is affected?	Positioning ± positional tests: nystagmus direction or duration (in) consistent with canal excitation	NA	NA	NA
	Is there a third window lesion?	NA	NA	NA	**Low thresholds, Amplitude asymmetry**
	Meniere's Disease: Which ear(s) are active? What is the level of hypofunction?	+/− **Caloric asymmetry** **Spontaneous, positional nystagmus (may point to irritative lesion or hypofunction)**	Asymmetry (may point to irritative lesion or hypofunction)	Gain may be normal even with caloric weakness	Asymmetric air conduction cVEMP (absent or reduced) and oVEMP (intact)
Chronic dizziness	Is there bilateral vestibular hypofunction? [b]	**Low total eye speed bilaterally** **Poor response to ice water caloric**	**Low gain** **Phase lead**	**Reduced VOR gain & saccades bilaterally**	(+/−)Absent/ reduced oVEMP (SVN), cVEMP (IVN)
	Is dizziness nonorganic or aphysiologic?	NA	NA	NA	NA

(*continued on next page*)

Table 1
(*continued*)

Clinical Scenarios	Sample Clinical Questions	Vestibular Function Tests and Pertinent Findings			
		VNG	**Rotary Chair**	**vHIT**	**VEMP**
Profound diagnostic uncertainty	What is happening?	NA	NA	NA	NA
Vestibular ablation: candidacy & efficacy assessment	How much vestibular function exists in each ear?	**Caloric asymmetry** **Total eye speed (to assess for bilateral weakness)**	Low gain if bilateral hypofunction	**VOR gain** **Corrective saccades**	oVEMP (SVN), cVEMP (IVN) response amplitude
Surgical Considerations[c]	Which ear has better vestibular function?	**Caloric asymmetry**	NA	**VOR gain** **Corrective saccades**	oVEMP (SVN), cVEMP (IVN) response amplitude
Clinical resources	Time to complete test in minutes (Mean [SD])[21,44]	71 [23]	26 [13]	10–15	~90 (estimate)
Billing – 2021 Medicare fee schedule	CPT Code(s) (Reimbursement Non-Facility/Facility)	92540 ($109.71/$112.01) 92537 ($42.59/$42.57)	92526 ($113.68/$121.43)	No CPT code available	92517 ($87.23/$43.97) 92518 ($81.30/$43.97) 92519 ($135.39/$65.95)

Denotes vestibular test modality with greatest potential to afford clinically meaningful results given the specific clinical scenario and question.

Abbreviations: BPPV, benign paroxysmal positional vertigo; CPT, current procedural terminology; cVEMP, cervical vestibular evoked myogenic potential; IVN, inferior vestibular nerve; NA, not applicable; oVEMP, ocular vestibular evoked myogenic potential; SVN, superior vestibular nerve; UVH, unilateral vestibular hypofunction; VEMP, vestibular evoked myogenic potential; vHIT, video head impulse test; VNG, videonystagmography; VOR, vestibulo-ocular reflex.

[a] Pertinent findings refer to evidence of an uncompensated lesion.

[b] Extent of frequency range involvement dependent on extent of lesion.

[c] Examples may include decisions regarding initial, bilateral, or contralateral cochlear implantation or stapes surgery, or postoperative assessments of symptomatic patients.

through the vestibulocollic reflex.[51] Ocular VEMP (oVEMP) measures excitation potentials in the contralateral inferior oblique, generated primarily by the utricle and superior vestibular nerve.[52] The outcome measures of cVEMP and oVEMP are their thresholds, peak-to-peak amplitudes, and latencies.

VEMP testing is most valuable for affirming the diagnosis of SCDS (**Fig. 8**).[53] In the presence of a third mobile window, otolith activation by sound is *enhanced*.[54] On the affected side(s), lower intensity sound stimuli induce the response (ie, lower VEMP threshold), and response amplitude may be abnormally increased (ie, amplitude asymmetry).[53] For the diagnosis of SCDS, the reported sensitivity and specificity of cVEMP are 42% to 91% and 90% to 100%, respectively, and of oVEMP are 62% to 100% and 73% to 100%, respectively.[55,56] VEMPs also detect *loss* of otolith function, manifesting as absent, reduced, or asymmetric cVEMP and/or oVEMP responses. While the clinical utility of loss-of-function applications is incompletely defined,[53] VEMPs may characterize vestibular nerve division(s) affected by vestibular neuritis or schwannoma,[54] ears with Meniere's disease,[57] and residual function after surgery/ablation.

SELECTING A VESTIBULAR TEST BATTERY

While clinical history is the cornerstone of vestibular investigation, vestibular function tests have clinically significant implications for diagnosis and management in specific scenarios. **Table 1** summarizes key tests, applications, and findings organized by common clinical questions that arise after clinicians craft a differential diagnosis. When diagnostic criteria are satisfied on clinical grounds (eg, BPPV, Meniere's), vestibular tests provide little additional information to bolster diagnostic confidence. However, as outlined, there are scenarios for which tests provide insight into lesion sites, severity, and compensation status, aiding in diagnostic refinement and management selection. The degree of vestibular impairment in unilateral pathology and the functional capacity of the contralateral vestibular system impacts management decisions, particularly regarding ablative therapies. Tests also inform decisions regarding initial or sequential otologic surgeries that risk vestibular loss (eg cochlear implantation, stapedectomy). Patients reporting vestibular symptoms before or after surgery may benefit from quantitative vestibular assessment, as documented hypofunction may contraindicate contralateral procedures. The majority of vestibular lesions initially impact the lower frequencies with preservation of the mid to high frequencies.[24] Patients with presumed bilateral vestibular areflexia require vestibular testing for diagnostic confirmation and quantification of injury extent. Determination of partial or complete vestibular loss across the frequency range informs the optimal treatment strategy (ie vestibular rehabilitation vs vestibular substitution for partial or complete loss, respectively).

SUMMARY

Despite the prevalence of dizziness and ample availability of diagnostic modalities, many affected individuals do not receive the prompt or accurate diagnoses needed to facilitate proper clinical management.[58–60] While most vestibular disorders may be diagnosed and managed based on data derived from history and exams, there are clinical scenarios in which vestibular function tests prove useful for diagnostic refinement and management decisions. An understanding of the insights that may be provided by vestibular tests and the limitations of the modalities facilitates efficient utilization.

CLINICS CARE POINTS

- The majority of vestibular disorders may be diagnosed based on history and physical exam findings.
- Vestibular function testing does not typically reveal the diagnosis but may provide insights that allow for diagnostic refinement and may inform management decisions.
- After formulating a differential diagnosis, tests should be selected to answer specific clinical questions rather than to "screen" for vestibular disease.
- Vestibular tests provide information about vestibular lesion localization, severity, compensation, and functional status.
- VNG and calorics provide data regarding laterality and severity of vestibular hypofunction and compensation status.
- Rotary chair may diagnose and/or confirm bilateral vestibular hypofunction and compensation status.
- vHIT provides an accessible means of assessing the function of all 6 semicircular canals and both superior and inferior vestibular nerve divisions.
- VEMP affords adjunctive data primarily efficacious in confirming the diagnosis of third window lesions including superior canal dehiscence.
- An understanding of vestibular physiology and the benefits and limitations of vestibular testing is requisite to the efficient use of this technology.

DISCLOSURE

The authors have no conflicts of interest to disclose. This work was supported by NIDCD R21DC016359 (M.E. Adams).

REFERENCES

1. Agrawal Y, Carey JP, Della Santina CC, et al. Disorders of balance and vestibular function in US adults: Data from the National Health and Nutrition Examination Survey, 2001-2004. Arch Intern Med 2009;169(10):938–44.
2. Lin HW, Bhattacharyya N. Balance disorders in the elderly: Epidemiology and functional impact. Laryngoscope 2012;122(8):1858–61.
3. Lin HW, Bhattacharyya N. Otologic diagnoses in the elderly: Current utilization and predicted workload increase. Laryngoscope 2011;121(7):1504–7.
4. Lin HW, Bhattacharyya N. Impact of dizziness and obesity on the prevalence of falls and fall-related injuries. Laryngoscope 2014;124(12):2797–801.
5. Saber Tehrani A, Coughlan D, Hsieh Y-H, et al. Rising Annual Costs of Dizziness Presentations to US Emergency Departments (P06.002). Neurology 2013;80(7 Supplement):689–96.
6. Kim AS, Sidney S, Klingman JG, et al. Practice variation in neuroimaging to evaluate dizziness in the ED. Am J Emerg Med 2012;30(5):665–72.
7. Newman-Toker DE, McDonald KM, Meltzer DO. How much diagnostic safety can we afford, and how should we decide? A health economics perspective. BMJ Qual Saf 2013;22(SUPPL.2):ii11–20.
8. American Academy of Otolaryngology - Head and Neck Surgery. Top 100 ENT CPT codes for 2015. Available at: https://www.entnet.org/wp-content/uploads/2021/06/non-facility-2021.pdf.

9. Adams ME, Yueh B, Marmor S. Clinician Use and Payments by Medical Specialty for Audiometric and Vestibular Testing among US Medicare Beneficiaries. JAMA Otolaryngol Head Neck Surg 2019;146(2):143–9.
10. Adams ME, Marmor S, Yueh B, et al. Geographic Variation in Use of Vestibular Testing among Medicare Beneficiaries. Otolaryngol Head Neck Surg 2017; 156(2):312–20.
11. Piker EG, Schulz K, Parham K, et al. Variation in the Use of Vestibular Diagnostic Testing for Patients Presenting to Otolaryngology Clinics with Dizziness. Otolaryngol Head Neck Surg 2016;155(1):42–7.
12. Canadian Society of Otolaryngology - Head & neck Surgery. Choosing Wisely Canada. 2020. Available at: https://choosingwiselycanada.org/otolaryngology/%0D%0A.
13. Phillips JS, Mallinson AI, Hamid MA. Cost-effective evaluation of the vestibular patient. Curr Opin Otolaryngol Head Neck Surg 2011;19(5):403–9.
14. Phillips JS, Fitzgerald JE, Bath AP. The role of the vestibular assessment. J Laryngol Otol 2009. https://doi.org/10.1017/S0022215109005611.
15. Grill E, Strupp M, Müller M, et al. Health services utilization of patients with vertigo in primary care: A retrospective cohort study. J Neurol 2014;261(8):1492–8.
16. Zhao JG, Piccirillo JF, Spitznagel EL, et al. Predictive capability of historical data for diagnosis of dizziness. Otol Neurotol 2011;32(2):284–90.
17. Jaynstein D. HINTS for differentiating peripheral from central causes of vertigo. J Am Acad Physician Assist 2016;29(10):56–7.
18. Roland LT, Kallogjeri D, Sinks BC, et al. Utility of an abbreviated dizziness questionnaire to differentiate between causes of vertigo and guide appropriate referral: A multicenter prospective blinded study. Otol Neurotol 2015;36(10): 1687–94.
19. Von Brevern M, Bertholon P, Brandt T, et al. Benign paroxysmal positional vertigo: Diagnostic criteria. J Vestib Res Equilib Orientat 2015;25(3–4):105–17.
20. Lopez-Escamez JA, Carey J, Chung WH, et al. Diagnostic criteria for Menière's disease. J Vestib Res Equilib Orientat 2015;25(1):1–7.
21. Kelly EA, Stocker C, Kempton CM, et al. Vestibular testing: Patient perceptions, morbidity, and opportunity costs. Otol Neurotol 2018;39(10):1222–8.
22. Follansbee WP. Diagnostic Decision Making: Time to Take Two Steps Back. Available at: http://ddmchecklist.upmc.com/.
23. Hall CD, Herdman SJ, Whitney SL, et al. Vestibular rehabilitation for peripheral vestibular hypofunction: An evidence-based clinical practice guideline: From the American physical therapy association neurology section. J Neurol Phys Ther 2016;40(2):124–55.
24. Jacobson GP, Shepard NT, Barin K, et al. Balance function assessment and management. In: Jacobson GP, Shepard NT, Barin K, et al, editors. Rotational Vestibular Assessment. 3rd edition. San Diego (CA): Plural Publishing; 2021. p. 283–332.
25. Yip CW, Strupp M. The Dizziness Handicap Inventory does not correlate with vestibular function tests: a prospective study. J Neurol 2018;265(5):1210–8.
26. Leigh RJ, Zee DS. The neurology of eye movements. In: Zee DS, editor. Diagnosis of Nystagmus and Saccadic Intrusions. 5th edition. New York: Oxford University Press; 2015. p. 658–768.
27. Goldberg JM, Wilson VJ, Cullen KE, et al. The vestibular system: a Sixth sense. New York: Oxford University Press; 2012. https://doi.org/10.1093/acprof:oso/9780195167085.001.0001.

28. Jellinger KA. The Neurology of Eye Movements 4th edn. Eur J Neurol 2009;16(7): e132.
29. Robinson DA, Zee DS, Hain TC, et al. Alexander's law: Its behavior and origin in the human vestibulo-ocular reflex. Ann Neurol 1984;16(6):714–22.
30. Curthoys IS, Michael Halmagyi G. Vestibular compensation: A review of the oculomotor, neural, and clinical consequences of unilateral vestibular loss. J Vestib Res 1995;5(2):67–107.
31. Walton MMG, Mustari MJ. Abnormal tuning of saccade-related cells in pontine reticular formation of strabismic monkeys. J Neurophysiol 2015;114(2):857–68.
32. Quaia C, Lefèvre P, Optican LM. Model of the control of saccades by superior colliculus and cerebellum. J Neurophysiol 1999;82(2):999–1018.
33. Shepard NT. Practical management of the balance disorder patient. In: Telian SA, editor. San Diego (CA): Singular Pub. Group; 1996.
34. Bhattacharyya N, Gubbels SP, Schwartz SR, et al. Clinical Practice Guideline: Benign Paroxysmal Positional Vertigo (Update). Otolaryngol Head Neck Surg 2017;156(3_suppl):S1–47.
35. Shepard N, Jacobson G. The caloric irrigation test. Handb Clin Neurol 2016;137: 119–31.
36. JONGKEES LB, MAAS JP, PHILIPSZOON AJ. Clinical nystagmography. A detailed study of electro-nystagmography in 341 patients with vertigo. Pract Otorhinolaryngol (Basel) 1962;24:65–93.
37. Strupp M, Kim JS, Murofushi T, et al. Bilateral vestibulopathy: Diagnostic criteria consensus document of the classification committee of the barany society. J Vestib Res Equilib Orientat 2017;27(4):177–89.
38. Palomar-Asenjo V, Boleas-Aguirre MS. S??nchez-Ferr??ndiz N, Perez Fernandez N. Caloric and Rotatory Chair Test Results in Patients with M??ni??re's Disease. Otol Neurotol 2006;27(7):945–50.
39. Yacovino DA, Hain TC, Musazzi M. Fluctuating vestibulo-ocular reflex in ménière's disease. Otol Neurotol 2017;38(2):244–7.
40. Arriaga MA, Chen DA, Cenci KA. Rotational chair (ROTO) instead of electronystagmography (ENG) as the primary vestibular test. Otolaryngol Head Neck Surg 2005;133(3):329–33.
41. Halmagyi GM, Curthoys IS. A Clinical Sign of Canal Paresis. Arch Neurol 1988; 45(7):737–9.
42. Weber KP, Aw ST, Todd MJ, et al. Head impulse test in unilateral vestibular loss: vestibulo-ocular reflex and catch-up saccades. Neurology 2008;70(6):454–63.
43. Halmagyi GM, Curthoys IS, Cremer PD, et al. The human horizontal vestibulo-ocular reflex in response to high-acceleration stimulation before and after unilateral vestibular neurectomy. Exp Brain Res 1990;81(3):479–90.
44. MacDougall HG, Weber KP, McGarvie LA, et al. The video head impulse test: Diagnostic accuracy in peripheral vestibulopathy. Neurology 2009;73(14): 1134–41.
45. MacDougall HG, McGarvie LA, Halmagyi GM, et al. The Video Head Impulse Test (vHIT) Detects Vertical Semicircular Canal Dysfunction. PLoS One 2013;8(4): e61488.
46. Curthoys IS, Manzari L. Clinical application of the head impulse test of semicircular canal function. Hearing, Balance and Communication 2017;15(3):113–26. https://doi.org/10.1080/21695717.2017.1353774.
47. Kaci B, Nooristani M, Mijovic T, et al. Usefulness of video head impulse test results in the identification of meniere's disease. Front Neurol 2020;11:581527.

48. Curthoys IS, Halmagyi GM. What does head impulse testing really test? JAMA Otolaryngol Head Neck Surg 2019;145(11):1080.
49. Sjögren J, Fransson P-A, Karlberg M, et al. Functional Head Impulse Testing Might Be Useful for Assessing Vestibular Compensation After Unilateral Vestibular Loss. Front Neurol 2018;9(NOV):979.
50. MacDougall HG, Curthoys IS. Plasticity during vestibular compensation: The role of saccades. Front Neurol 2012. https://doi.org/10.3389/fneur.2012.00021.
51. Colebatch JG, Halmagyi GM, Skuse NF. Myogenic potentials generated by a click-evoked vestibulocollic reflex. J Neurol Neurosurg Psychiatry 1994;57(2):190–7.
52. Todd NPMA, Rosengren SM, Aw ST, et al. Ocular vestibular evoked myogenic potentials (OVEMPs) produced by air- and bone-conducted sound. Clin Neurophysiol 2007;118(2):381–90.
53. Fife TD, Colebatch JG, Kerber KA, et al. Practice guideline: Cervical and ocular vestibular evokedmyogenic potential testing: Report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. Neurology 2017;89(22):2288–96.
54. Rosengren SM, Colebatch JG, Young AS, et al. Vestibular evoked myogenic potentials in practice: Methods, pitfalls and clinical applications. Clin Neurophysiol Pract 2019;4:47–68.
55. Noij KS, Rauch SD. Vestibular Evoked Myogenic Potential (VEMP) Testing for Diagnosis of Superior Semicircular Canal Dehiscence. Front Neurol 2020;11:695.
56. Fife TD, Satya-Murti S, Burkard RF, et al. Vestibular evoked myogenic potential testing Payment policy review for clinicians and payers. Neurol Clin Pract 2018;8(2):129–34.
57. Taylor RL, Wijewardene AA, Gibson WPR, et al. The vestibular evoked-potential profile of Ménière's disease. Clin Neurophysiol 2011;122(6):1256–63.
58. To-Alemanji J, Ryan C, Schubert MC. Experiences Engaging Healthcare When Dizzy. Otol Neurotol 2016;37(8):1122–7.
59. Fife D, Fitzgerald JE. Do patients with benign paroxysmal positional vertigo receive prompt treatment? Analysis of waiting times and human and financial costs associated with current practice. Int J Audiol 2005;44(1):50–7.
60. Bhattacharyya N, Baugh RF, Orvidas L, et al. Clinical practice guideline: Benign paroxysmal positional vertigo. Otolaryngol Head Neck Surg 2008;139(5 SUPPL. 4). https://doi.org/10.1016/j.otohns.2008.08.022.
61. Zuniga MG, Janky KL, Nguyen KD, et al. Ocular Versus Cervical VEMPs in the Diagnosis of Superior Semicircular Canal Dehiscence Syndrome. Otol Neurotol 2013;34(1):121–6.

Comments Submitted by Interested Parties on a Pending AMA Guides® Editorial Change Proposal

Instructions: By submitting comments on this form regarding an Editorial Change Proposal, I attest that I have read AMA Guides® Editorial Change Proposal and Submission Requirements and will use them as the primary points of consideration when submitting the Comment Form. As an interested party, I understand that my comments are limited to the original editorial change proposal.

Name or Topic of Proposal: ENT and Related Structures

Individual or Organization Submitting Comments: Steven Mandel MD

Date: 7/8/2022

I. General Criteria for Guides Editorial Changes

- The proposed change is carefully drafted and conforms to the prevailing style of the *AMA Guides 6th Edition*;
- The terminology and the analytical frameworks used in the proposal are consistent with the World Health Organization's International Classification of Functioning, Disability, and Health (ICF);
- The structure and content of the proposed editorial change ensures that impairment ratings are transparent, clearly stated, and reproducible, to insure physician interrater reliability;
- The clinical soundness of the proposed editorial change is demonstrated with the best available evidence except in the case of minor editorial changes.

1. **Does the requested procedure meet the AMA Guides® Editorial Change Proposal and Submission Requirements?**

   ☒Yes
   ☐No
   If No, please explain. (1500 character limit)

2. **Does the submitted literature adequately support the Editorial Change Proposal?**
   ☒Yes
   ☐No
   ☐N/A

   If No, please explain. (1500 character limit)

3. **Are you aware of contradictory literature related to the Editorial Change Proposal?**

   ☒Yes
   ☒No
   ☐N/A

   If Yes, please include a maximum of five (5) articles when submitting this form. Articles in full text or PDF formats are required. Citations only will not be considered.

4. **Do you support this Editorial Change Proposal?**

   ☒Yes
   ☐No

   If No, please provide the rationale for lack of support, citing the specific criteria not met shown at the top of this form. (1500 character limit)

5. **Does the Editorial Change Proposal have any impact on other *AMA Guides* content that may not have been recognized or considered, or conflict with other precedents in the AMA Guides that might affect usage?**

   ☒Yes
   ☐No

   If Yes, please explain. (1500 character limit)

Please provide additional commentary related to the editorial change proposal.

Rotary Chair comments:
What are the normative values and measures of abnormalities ?
Are there extablished values for different ages? And standards for all populations?
Is is necessary to do testing beyond the physical exam to determine vestibular impairment?
Is trhe Head Shaking test sufficient to establish vestibular impairment?

ENG- VNG
What is the correlation between the clinical exam and VNG and ENG testing ?
Is both the VNG and ENG acceptable for impairment ratings ?
The VNG and ENG are not meant to make a diagnosis in the absence of the History and physical.

Mastication
Age and Cognition are associated with changes in mastication
What are the standards for food textures ie .. food diameter
Effect of occlusal conditions, loss of teeth, dentures
What are the effects of oral rehabilitation and MMI?

Dysphagia
What are the symptoms?
What are the accurate tests? Real time MRI/ scintigraphy
Was manometry performed ?
Has the individual under gone swallowing rehabilitation?
Differentiate oropharyngeal and esophageal dysphagia

Deglutination
Sleep related issues w/o sleep apnea
History of aspiration
Effects of Botox
Correlation wih Laryngeal exam and function

Comments Submitted by Interested Parties on a Pending AMA Guides® Editorial Change Proposal

Instructions: By submitting comments on this form regarding an Editorial Change Proposal, I attest that I have read AMA Guides® Editorial Change Proposal and Submission Requirements and will use them as the primary points of consideration when submitting the Comment Form. As an interested party, I understand that my comments are limited to the original editorial change proposal.

Name or Topic of Proposal: ENT changes

Individual or Organization Submitting Comments: Kathryn L Mueller

Date: 6/6/2022

I. General Criteria for Guides Editorial Changes

- The proposed change is carefully drafted and conforms to the prevailing style of the *AMA Guides 6th Edition*;
- The terminology and the analytical frameworks used in the proposal are consistent with the World Health Organization's International Classification of Functioning, Disability, and Health (ICF);
- The structure and content of the proposed editorial change ensures that impairment ratings are transparent, clearly stated, and reproducible, to insure physician interrater reliability;
- The clinical soundness of the proposed editorial change is demonstrated with the best available evidence except in the case of minor editorial changes.

1. **Does the requested procedure meet the AMA Guides® Editorial Change Proposal and Submission Requirements?**

   ☒Yes
   ☐No
   If No, please explain. (1500 character limit)

2. **Does the submitted literature adequately support the Editorial Change Proposal?**
   ☒Yes
   ☐No
   ☐N/A

   If No, please explain. (1500 character limit)

3. **Are you aware of contradictory literature related to the Editorial Change Proposal?**

   ☐Yes
   ☒No
   ☐N/A

   If Yes, please include a maximum of five (5) articles when submitting this form. Articles in full text or PDF formats are required. Citations only will not be considered.

4. **Do you support this Editorial Change Proposal?**

   ☒Yes
   ☐No

   If No, please provide the rationale for lack of support, citing the specific criteria not met shown at the top of this form. (1500 character limit)

5. **Does the Editorial Change Proposal have any impact on other *AMA Guides* content that may not have been recognized or considered, or conflict with other precedents in the AMA Guides that might affect usage?**

   ☐Yes
   ☒No

   If Yes, please explain. (1500 character limit)

Please provide additional commentary related to the editorial change proposal.

I was the section editor for the ENT chapter and worked closely with Dr Sataloff on the initial chapter. I believe all of the suggested changes are needed and appropriate.

I note that Dr Sataloff has additionally asked for the ability to consider work function in the rating. I do not agree with that as generally speaking work disability is a separate question. However I do think that use of the functional tools, as in the PROMIS tools, would go a long way toward allowing additional functional consideration for the loss of specific abilities that affect an individual's overall quality of life and ability to function in specific circumstances essential for that patient.