Measurement properties of the dizziness handicap inventory: a systematic review

Koppelaar-van Eijsden, H.M.; Schermer, T.R.; Bruintjes, T.D.

Measurement Properties of the Dizziness Handicap Inventory: A Systematic Review

*Hanna Maria Koppelaar-van Eijsden, *†Tjard Roland Schermer, and *‡Tjasse Doewe Bruintjes

*Apeldoorn Dizziness Centre, Gelre Hospitals, Apeldoorn, The Netherlands; †Department of Primary and Community Care, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; and ‡Department of Otorhinolaryngology, Leiden University Medical Center, Leiden, The Netherlands

Objective: To critically appraise and summarize the quality of the measurement properties of the Dizziness Handicap Inventory's (DHI) in adult patients with complaints of dizziness.
Databases Reviewed: Pubmed, Embase, and CINAHL.
Methods: The selected literature databases were systematically searched to identify studies investigating one or more measurement properties of the DHI. From the included studies, relevant data were extracted, their methodological quality was assessed, the results were synthesized and the evidence was graded and summarized according the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology.
Results: The search strategy resulted in 768 eligible publications, 42 of which were included in the review. Overall, evidence on the DHI's content validity was either lacking or limited and of low quality. Moderate evidence was found for inconsistent structural validity, sufficient construct validity and borderline sufficient responsiveness. Based on the studies included, low evidence was found for sufficient reliability of the DHI total score. No evidence synthesis could be done for the DHI's internal consistency due to multidimensionality (i.e., lack of support of the original subscales) and for its measurement error due to a lack of published information on the minimal important change.
Conclusions: The current evidence for a number of measurement properties of the DHI is suboptimal. Because of its widespread use and the current lack of a better alternative, researchers can use the DHI when assessing handicapping effects imposed by dizziness, but they should be aware of its limitations. Moreover, we recommend using the DHI total score only and also to consider adding an instrument with more favorable measurement properties when assessing self-perceived handicap in patients with dizziness.
Key Words: Dizziness—Dizziness Handicap Inventory—Measurement properties—Patient-reported outcome measure—Vestibular research.

Otol Neurotol **43:**e282–e297, 2022.

INTRODUCTION

Dizziness is a common health complaint, with the lifetime population prevalence estimated at 20 to 30% (1). It can impair activities of daily living, cause falls and absence from work, and significantly reduce quality of life (2). In order to establish the effectiveness of treatment in patients with dizziness, patient-reported outcome measures (PROMs) are often used besides more objective measurements (3). A commonly used questionnaire to measure the self-perceived handicapping effects imposed by dizziness is the Dizziness Handicap Inventory (DHI), developed by Jacobson and Newman in 1990 (4). It consists of 25-items that add up to a total score ranging from 0 to 100, a high score indicating greater dysfunction. The original DHI has three subscales; a physical, functional, and emotional subscale. The DHI has been used as a (primary) outcome measure in numerous studies assessing vertigo/dizziness in patients with various causes of dizziness (5,6) and has been translated in several other languages. Many consider the DHI as ''the golden standard'' for measuring self-perceived handicap or quality of life in patients with dizziness, and over the past three decades the DHI has been used in research and routine patient care worldwide (7). Nonetheless, questions have been raised about the DHI's validity and reliability. Several systematic reviews concluded that the DHI should not be the questionnaire of first choice because its measurement properties are not always satisfying and do not meet current standards for contemporary health outcome measurement (8–10). However, these reviews only considered the original article of Jacobson and Newman and did not include studies about the measurement properties of the DHI published after that. Therefore, we conducted a systematic literature review with the aim to critically appraise and summarize the

Address correspondence and reprint requests to Hanna Maria Koppelaar-van Eijsden, M.Sc., Apeldoorn Dizziness Centre, Gelre Hospitals Albert Schweitzerlaan 31, 7334 DZ Apeldoorn, The Netherlands; E-mail: h.van.eijsden@gelre.nl
This study was financially supported by Gelre hospitals.
The authors disclose no conflicts of interest.
Supplemental digital content is available in the text.
DOI: 10.1097/MAO.0000000000003448

quality of the measurement properties of the DHI for adult patients with complaints of dizziness.

METHODS

This systematic review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement (11) and the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) guideline for conducting systematic reviews on PROMs (12–14). We systematically reviewed the validity, reliability and responsiveness of the DHI assessing the following measurement properties: content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness. We use the definitions of the COSMIN taxonomy (15) (Table 1).

Protocol registration: not applicable.

Literature Search

PubMed, Embase, and CINAHL were last searched on January 22, 2021. The search strategy consisted of terms for the PROM ''Dizziness Handicap Inventory'' and terms for ''measurement properties'' using the COSMIN search filter (16) (see Appendix 1, Supplemental Digital Content, http://links.lww.com/MAO/B392, for the full search strategy). Reference lists of included studies were manually checked for additional articles. A search in PsycINFO did not identify studies that were not already found through reference checking.

Eligibility Criteria

An article was included if it described the development or translation of the DHI or assessed one or more measurement properties; the study population consisted of patients diagnosed with a condition that comes with complaints of dizziness or unsteadiness or both; and if it was published in English and full-text available. Articles were excluded if the DHI was only used to measure treatment effect, validate another instrument, or was used as a diagnostic instrument.

Study Selection, Data Extraction, and Hypotheses Formulation

Titles, abstracts, and after that potentially eligible full-text articles were screened by two independent raters (HKE and TS) using the Rayyan QCRI (17). Disagreements were discussed among the two raters until consensus was reached. If necessary a third rater (TB) was consulted.

From each article the following information was extracted: sample size, age, sex, setting, diagnosis, duration of dizziness complaints, mean DHI score at baseline and length of follow-up. Regarding the measurement properties data was extracted about the type of measurement property studied, its outcome, and information on the methodology used. Data were extracted by two of the authors (HKE and TS) and disagreements were discussed until consensus was reached.

To assess the construct validity (i.e., hypotheses testing) and responsiveness we formulated a priori hypotheses. We expected moderate to high (i.e., $r = 0.50–0.89$) correlations between the DHI total score and other instruments measuring self-perceived handicap due to vertigo and/or unsteadiness, low to moderate ($r = 0.26–0.69$) correlations with dizziness, anxiety, depression, health-related quality of life and gait, low correlations ($r = 0.26–0.49$) with balance tests, and very low correlations ($r \leq 0.25$) with vestibular function tests (18).

Risk of Bias and Synthesis of the Results

Assessment of the measurement properties consisted of three phases.

First, methodological quality of each study was assessed using the COSMIN Risk of Bias checklist (13), rating them as ''very good,'' ''adequate,'' ''doubtful,'' or ''inadequate.''

Second, the results of each study were rated against the criteria for good measurement properties (12). Results were rated as either ''sufficient'' (+), ''insufficient'' (−), or ''indeterminate'' (?).

Third, the results of all studies on a particular measurement property were quantitatively summarized and compared against the criteria for good measurement properties to determine whether the measurement property of the DHI was overall

TABLE 1. *Definitions of validity, reliability and the nine measurement properties according to the COSMIN taxonomy*

Measurement Property	Definition
Validity	The degree to which a PROM measures the construct it purports to measure
Reliability	The degree to which the measurement is free from measurement error
Structural validity	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured.
Internal consistency	The degree of the interrelatedness among the items.
Reliability	The proportion of the total variance in the measurements which is due to ''true'' differences between patients.
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured.
Hypotheses testing for construct validity	The degree to which the scores of a PROM are consistent with hypotheses based on the assumption that the PROM validly measures the construct to be measured.
Cross-cultural validity / measurement invariance	The degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version of the PROM.
Criterion validity	The degree to which the scores of a PROM are an adequate reflection of a ''golden standard.''
Responsiveness	The ability of a PROM to detect change over time in the construct to be measured.

COSMIN is an initiative which aims to improve the selection of outcome measurement instruments both in research and in clinical practice by developing methodology and practical tools for selecting the most suitable outcome measurement.

COSMIN indicates COnsensus-based Standards for the selection of health Measurement Instruments; PROM, patient-reported outcome measure.

''sufficient'' (+), ''insufficient'' (−), ''inconsistent'' (±), or ''indeterminate'' (?). The overall quality of the evidence was graded using the Modified GRADE approach (12), which comprises assessment of risk of bias (i.e., methodological quality of the studies), inconsistency (unexplained inconsistency of results across studies), imprecision (total sample size of the studies) and indirectness (evidence from different population(s) than our population of interest). Finally, the overall quality of the evidence from all studies combined was rated as ''high,'' ''moderate,'' ''low,'' or ''very low.''

RESULTS

Study Selection

The combined PubMed, Embase, and CINAHL searches resulted in 768 unique publications, of which 134 articles were selected based on title and abstract. Finally, 42 articles were included (Fig. 1) (4,19–59).

Table 2 shows the patient characteristics of each study and the measurement properties studied. Study populations were heterogeneous in the majority of studies, comprising patients with dizziness and unsteadiness due to vestibular disorders, central dysfunction, multisensory/multifactorial causes, or unspecified causes. Time since onset of the dizziness symptoms varied from the acute stage to $\geq$1 year, although the majority of studies did not describe this parameter. Studies were conducted at either the outpatient clinic of a (tertiary) hospital, an ENT or neurology hospital department, or a specialized dizziness clinic. One study (40) took place at a private physiotherapy practice.

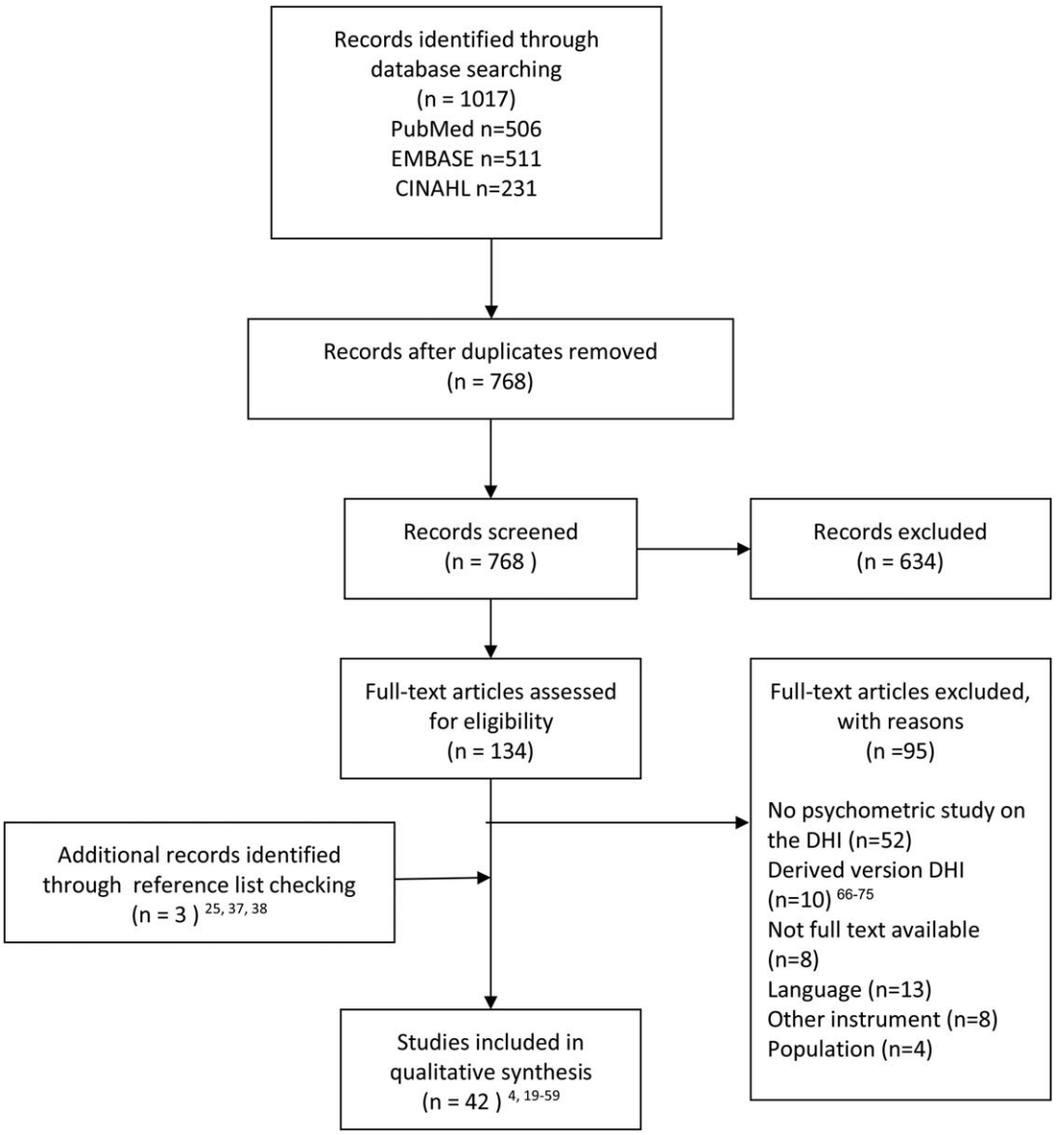

FIG. 1. PRISMA flow diagram of the search. DHI indicates Dizziness Handicap Inventory.(66–75)

TABLE 2. *Study characteristics and measurement properties of the 42 studies included in the quantitative analysis*

Year of Publication, 1st Author	N*	Mean Age Years ± SD (Range)	Sex, % Female	Diagnoses[d]	Time Since Onset Mean ± SD	DHI Baseline Mean ± SD (Range)	Language	Measurement Properties Studied
1990 Jacobson (4)	106	48 ± 15.84 (15–85)	62.3	Not reported	Not reported	32.7 ± 21.9	English	Development, content validity, reliability, measurement error, hypotheses testing
1991 Jacobson (19)	367	48.84 ± 14.51 (17–85)	65.1	Patients for balance function evaluation	Not reported	Not reported	English	Hypotheses testing
1997 Enloe (30)	95	57 ± 14.9 (25–88)	Not reported	Various diseases, 82.1% UVW	Not reported	53.57 ± 20.82 (6–96)	English	Reliability, measurement error, responsiveness
1998 Ryd (65)	22	56.21 ± 13.92 (34–79)	68.2	Various diseases, 68.2% MD	Not reported	Not reported	English	Structural validity
1999 Asmundson (52)	95	39.3 ± 10.5 (23–78)	65	Not reported	Not reported	Not reported	English	Structural validity
2000 Jacobson (55)	22	51.59 ± 19.52	45.5	Patients with normal balance function	Not reported	30.27 ± 20.98	English	Hypotheses testing
	29	49.83 ± 16.37	55.2	UVW	Not reported	39.79 ± 21.15		
	21	56.48 ± 18.00	42.9	BVW	Not reported	49.71 ± 30.00		
2001 Perez (56)	337	50.9 (14–83)	50.4	Various diseases, 37.1% MD, 13.6 BPPV	Not reported	39.02 (0–92)	Spanish	Structural validity
2003 Jarlsäter (57)	15	63 (30–87)	66,7	Vestibular dysfunction	>3 months	Median 52 (22–88)	Swedish	Translation, content validity, reliability
2004 Poon (58)	71	52.86 ± 15.6 (17–85)	66.2	Various diseases, 62% uncompensated VW	>6 months	44.94 ± 24.20 (0–100)	Chinese	Translation, content validity, internal consistency, reliability, responsiveness
2005 Kammerlind (59)	30	56 ± 12	43.3	UVW	13 ± 10 months (1–33)	28 ± 23	Swedish	Reliability, measurement error, hypotheses testing
	20	73 ± 6	55	Central neurological dysfunction	22 ± 20 months (1–84)	35 ± 21		
2006 Vereeck (20)	106	53.6 ± 11.5	49.1	Various diseases, 55.7% vestibular schwannoma	Not reported	34.0 ± 25.7	Dutch	Translation, reliability, measurement error
2006 Vereeck (21)	214	53.9 ± 13.5 (23–87)	48.6	Various diseases, 48.6% vestibular schwannoma	Not reported	35.1 ± 25 (0–96)	Dutch	Hypotheses testing
2007 de Castro (22)	250	61.3 (14–91)	74	Chronic dizziness and diagnostic hypothesis of vestibular syndrome	≥ 3 months	44.87 ± 24.97	Portuguese	Translation, content validity, reliability, measurement error
2007 Mbongo (23)	32	47.6 ± 10.7	50	UVW	1–2 months	26.9 ± 17.7	Not reported	Hypotheses testing
	23	47.1 ± 8.37	65.2		4–7 months	24.3 ± 20.7		
	37	49.2 ± 9.5	59.5		>1 year	25.8 ± 18.8		

TABLE 2 *(Continued)*

Year of Publication, 1st Author	N*	Mean Age Years ± SD (Range)	Sex, % Female	Diagnoses[d]	Time Since Onset Mean ± SD	DHI Baseline Mean ± SD (Range)	Language	Measurement Properties Studied
2007 Vereeck (24)	214	53.9 ± 13.5 (23–87)	48.6	Various diseases, 48.6% vestibular schwannoma	Not reported	35.1 ± 25 (0–96)	Dutch	Structural validity, internal consistency
2009 Karapolat (25)	33	50.09 ± 14.06	66.7	UVW	40.21 ± 42.9 months	49.33 ± 23.91	Turkey	Translation, internal consistency, hypotheses testing, responsiveness
2009 Kurre (26)	127	50.5 ± 13.2 (21–75)	61.4	Various diseases, 40.9%	Various, >12 months 59.1%	44.5 ± 21.6	German	Translation, content validity, internal consistency, reliability, measurement error
2009 Tamber (27)	92	47.2 ± 11.46 (26–64)	70	Various diseases, 64% Vestibular dizziness	58.2 ± 84.1 months (2–481)	39.91 ± 18.95 (4–86)	Norwegian	Translation, content validity, internal consistency, reliability, measurement error, hypotheses testing, responsiveness
2010 Kurre (28)	194	50.6 ± 13.6	61.9	Various diseases, 37.1 UVW	Various, >12 months 58.8%	44.8 ± 22.2	German	Structural validity
2010 Nola (29)	50	51.6 ± 14.5 (25–85)	76	Various diseases, 48% no vestibular disease	2.3 ± 1.2 days (1–5)	40.4 ± 22.9	Italian	Translation, content validity, internal consistency
2011 Goto (31)	176	61.2 ± 13.5	71.6	Various diseases, 47.2% UVW	>1 month	36.42 ± 23.19	Japanese	Translation, structural validity, internal consistency
2012 Alsanosi (32)	50	45	44	Vestibular disorders	>2 months	50.2 ± 23.9 (10–100)	Arabic	Translation, content validity, internal consistency
2014 Friscia (33)	45	56 ± 15 (18–79)	64.4	Various diseases, 44% UVW	76 weeks[a] (4–832)	45 ± 25 (6–92)	English	Responsiveness
2014 Georgieva-Zhostova (34)	97	45.08 ± 13.85	80.4	Various diseases, 20.6% BBPV	<6 months 47.4%, >6 months 52.6%	60.05 ± 20.62 (6–88)	Bulgarian	Translation, content validity, internal consistency, reliability, measurement error, hypotheses testing

TABLE 2 *(Continued)*

Year of Publication, 1st Author	N*	Mean Age Years ± SD (Range)	Sex, % Female	Diagnoses[d]	Time Since Onset Mean ± SD	DHI Baseline Mean ± SD (Range)	Language	Measurement Properties Studied
2014 Jafarzadeh (35)	57	44.5 ± 14.2 (18–70)	35	Vertigo and dizziness	34.3 ± 39.8 months (1 month to 10 years)	Not reported	Persian	Translation, content validity, structural validity, internal consistency, reliability, measurement error
2015 Sarda (36) [e]	–	–	–	–	–	–	Marathi	Translation, internal consistency
2016 Akhilesh (37)	60	40.8 ± 12.2 (18–65)	50	Complaints of vertigo/dizziness	Not reported	Range 0–70	Kannada language (India)	Translation, content validity, internal consistency
2016 Akhilesh (38)	60	45.8 ± 13.1 (29–67)	75	Complaints of vertigo/dizziness	Not reported	Range 0–70	Malayalam language (India)	Translation, content validity, internal consistency
2017 Ardic (39)	2111	38.8 ± 21.3	71.1	Patients admitted with complaints of disequilibrium and vertigo	Not reported	41.62 ± 23.88		Structural validity, internal consistency
2017 Chiarovano (40)	90	65 ± 15 (31–89)	56.7	Various diseases	Not reported	Not reported	English	Construct validity
2017 Colnaghi (42)	316	53.5 ± 15.6	60.8	Various diseases, 34% no diagnosis	1.6 ± 3 years	Not reported	Italian	Translation, content validity, internal consistency
2017 Nikitas (43)	90	54.68 ± 16.74 (19–80)	62.2	Various diseases, 31.1% BPPV	Non-acute onset	Not reported	Greek	Translation, content validity, internal consistency, reliability, hypotheses testing
2018 Yip (44)	799	♀60.0 ± 17 ♂62.5 ± 16	40	Various diseases, 18.4% UVW	Not reported	35 (VW), 52 (central disorders), 40 (functional disorders)[a,b]	German	Hypotheses testing
2019 Formeister (45)	70	56.3 ± 13.8	49	Not reported	Not reported	42.8 ± 23.5	English	Hypotheses testing
2019 Neupane (46)	50	39 ± 10.7 (15–55)	64	Complaints of vertigo/dizziness	Not reported	Not reported	Gujarati Language (India)	Translation, content validity, internal consistency, reliability
2019 Soleimani (47)	113	44.5 ± 11.3	57.5	Various diseases, 44.2% VW	Not reported	33.2 ± 12.8	Iranian	Structural validity, internal consistency, reliability, measurement error, hypotheses testing

TABLE 2 *(Continued)*

Year of Publication, 1st Author	N*	Mean Age Years ± SD (Range)	Sex, % Female	Diagnoses[d]	Time Since Onset Mean ± SD	DHI Baseline Mean ± SD (Range)	Language	Measurement Properties Studied
2019 Szostek-Rogula (48)	230	56.2 ± 13.6 (25–87)	73	Vestibular disorder	Chronic	(4–100)	Polish	Translation, content validity, structural validity, internal consistency
2019 Valancius (49)	108	51.9 ± 16.1	75.9	VW (83.3%) and central vestibular disorders	Various, >12 month 52.7%	36.7 ± 19.6 (n=65)	Lithuania	Translation, content validity, structural validity, internal consistency, reliability, hypotheses testing
2019 van de Wyngaerde (50)	999	57.60 ± 17.70	58.5	Not reported	Not reported	Not reported	English	Structural validity
	992	56 ± 18.05	39.6	Not reported	Not reported	Not reported		
2019 Zmnako (51)	301	44.5 ± 15.2	59.8	Various diseases, 35.9 UVW	17.3 ± 28.8 months	41[b,c]	Central Kurdish	Translation, content validity, internal consistency, reliability, hypotheses testing
2020 Ahmed (53)	202	18–89	29.7	Various diseases	>3 months	Not reported	Urdu	Translation, structural validity, internal consistency, reliability
2020 Vozel (54)	50	61.4 ± 13.8	56.7	Patients treated at our department	Not reported	49.7 ± 22.8	Slovenian language	Translation, content validity, reliability

Articles presented in chronological and thereafter alphabetical order.

BPPV indicates Benign Paroxysmal Positional Vertigo; BVW, Bilateral Vestibular Weakness; DHI, Dizziness Handicap Inventory; MD, Mènière's disease; N*, maximum number of participants in the study; SD, standard deviation; UVW, Unilateral Vestibular Weakness; VW, Vestibular Weakness.

[a]Median.

[b]Approximation, read from figure.

[c]Weighted median.

[d]In case of multiple diagnoses the largest diagnoses group is mentioned in the table.

[e]Article included because of the information in the Methods section about the translation of the DHI, which seems to be not based on the data as reported in this article.

CONTENT VALIDITY

Development of the DHI

Jacobson and Newman (4) developed the DHI around 1990 to evaluate the self-perceived handicapping effects resulting from dizziness and unsteadiness, because an adequate tool to quantify the effects of medical, surgical, and rehabilitative management was lacking. The construct of the DHI is defined as ''precipitating physical factors associating with dizziness as well as the functional and emotional consequences of vestibular system disease.'' No conceptual model underlying the development of this construct was described. Items for the DHI were generated by reviewing case-history reports of patients with dizziness. No description of the diagnoses of these patients was given, and it was not made clear whether the questionnaire applies to patients with acute dizziness, chronic dizziness, or both. No cognitive interviews with patients or health professionals were conducted to establish the relevance, comprehensibility, and comprehensiveness of the questionnaire. Based on these findings and in the light of the current consensus on how to develop a PROM we consider the development of the DHI as ''inadequate.''

Translations of the DHI

Since its introduction the DHI has been translated into 22 different languages (Table 2). A multiple forward and backward translation procedure was used in 13 out of 24 studies (20,22,25,26,29,32,34,36,42,47,49,53,54). Cultural or textual adaptations were made in some translated versions of the DHI (26,32,35,51,58).

Content Validity

No studies were found in which the primary goal was to assess the content validity of the English or translated versions of the DHI. However, the majority of studies which translated the DHI (20 out of 22) included some form of pilot testing for (one or more aspects of) the content validity, that is, relevance, comprehensiveness and/or comprehensibility (22,25–27,29,32,34,35,37,38, 42,43,46–49,51,54,57,58). We rated the methodological quality for these individual studies as ''doubtful,'' mainly because clear descriptions of the process of executing the cognitive interviews were lacking. The description given by Kurre et al. was the most complete, but nevertheless the methodological quality of this study was rated as ''doubtful'' because the analysis was done by a single researcher and no information on the skills of the moderators and interviewer was reported (26).

Relevance

In three studies (26,32,57), it appears that patients were asked about the relevance of all individual items of the DHI. A clear description on how patients were asked about the relevance and what the findings were is lacking in the respective articles. Therefore, the overall rating for ''relevance'' is rated as 'indeterminate' and we could not grade the quality of evidence.

Comprehensiveness

In two studies, information on comprehensiveness of the DHI was reported (26,57). In two other studies, patients were asked about comprehensiveness, but a clear description is lacking (32,49). Patients (n = 14) in the study of Kurre et al. (26) estimated that the DHI covered, on average, 92% (range 50–100%) of their self-perceived disabilities. In contrast patients reported missing items about being in stressful/hectic environments, or when walking longer distances in open areas, climbing a staircase, using a lift, specific driving functions, activities of self-sufficiency, job activities and further emotional aspects (26,57). Overall, based on these findings we rated ''comprehensiveness'' as ''inconsistent'' and rated the overall quality of evidence as ''low.''

Comprehensibility

Comprehensibility of the DHI was addressed in 20 translation studies (22,25–27,29,32,34,35,37,38,42,43, 46–49,51,54,57,58). Overall, the patients in these studies had no difficulty understanding or answering the questions. In two studies, vestibular rehabilitation experts were consulted regarding the comprehensibility of the DHI, but no particularities were reported (35,42). Based on these findings we rated the overall ''comprehensibility'' as ''sufficient'' with low quality of evidence.

Structural Validity

Jacobson and Newman did not describe the type of model the DHI is based on, that is, either a reflective model (which assumes that all items in a scale or subscale are manifestations of one underlying construct and are thus expected to be correlated) or a formative model (in which items in a scale or subscale are not supposed to be correlated) (12). However, given the type of questions we feel that the DHI can be considered as a reflective model and that therefore, the structural validity should be evaluated.

Table 3 presents an overview of the fourteen studies addressing structural validity (24,27,28,31,35,39,41,47–50,52,53,56). Methodological quality is rated as inadequate for six studies, mainly due to small sample size (27,35,47,49,52). One study used Rasch Analysis and most of the other studies used exploratory factor analysis (EFA). Methodological quality was therefore rated as ''adequate'' (24,28,31,39,48,56). Two studies used CFA to verify the structure of the DHI, and the methodological quality of these studies was rated as ''very good'' (50,53). However, the criteria for good measurement properties for structural validity were not met (Table 3). The number of factors found by EFA in the respective studies varied from two to eight and factors explain at least 41% of the variance in DHI scores. One recent study in a large sample using bi-factor analysis (50) showed one general factor comprising all 25 items of the DHI that explained 70% of the variance and three additional factors explaining 5.9, 6.4, and 12.9% of the variance. The original authors described three subscales (physical, functional, and emotional) of the DHI (4). None of the 14 subsequent studies that have looked at the structural validity of the DHI were able to support these original subscales. Based on these findings, we rated the overall ''structural validity'' of the DHI as ''inconsistent'' and graded the quality of evidence as ''moderate.''

Internal Consistency

The structural validity of the DHI is inconsistent and therefore the condition of sufficient structural validity has not been met. Thus, the internal consistency could not be assessed. Cronbach's Alpha for the DHI total score ranges from 0.74 to 0.93 (see Appendix 2, Supplement Digital Content 2, http://links.lww.com/MAO/B392, table A showing the results for 'internal consistency' of the DHI (total score and subscales) as reported by 23 studies) (24–27,29,31,32,34–39,42,43,46–49,51,53,54,58).

Cross-Cultural Validity/Measurement Invariance

None of the 42 studies included assessed cross-cultural validity or measurement invariance. Hence this measurement property of the DHI cannot be rated.

Reliability

Eighteen studies reported on test–retest reliability of the DHI total score (Table 4) (4,20,22,26,27,30,34,35, 43,46,47,49,51,53,54,57–59), with sample sizes varying

TABLE 3. *Structural validity of the DHI*

Reference	N	Type of Factor Analysis	No of Factors*	Factors	Results	Study Quality	Rating
Ryd (41)	22	Rasch	–	–	$	Inadequate	?
Asmudson (52)	95	EFA	3	Disability, postural difficulties, phobic avoidance	53.80%	Inadequate	+
Perez (56)	337	EFA	3	Vestibular handicap, vestibular disability, visuo-vestibular disability	48.32%	Adequate	–
Vereeck (24)	214	EFA	4	Functional, emotional, motion sensitivity, visuo-vestibular	53.80%	Adequate	+
Tamber (27)	92	EFA	2, 3, 8	Not specified	74% (8 factors)	Inadequate	?
Kurre (28)	194	EFA	3	Effect of dizziness on emotion and participation, provoking dizziness postural stability	49.20%	Adequate	–
Goto (31)	176	EFA	5	Functional, emotional, motion sensitivity		Adequate	?
Jafarzadeh (35)	57	EFA	8	Not specified	74%	Inadequate	+
		EFA	3 fixed factors	Not specified	46%		–
Ardic (39)	2111	EFA	2	Not specified	41%	Adequate	–
Soleimani (47)	113	EFA	2	Emotional and physical effects	54.8%	Inadequate	+
Szostek-Rogula (48)	230	EFA	3	Restrictions and disabilities due to vertigo, positional vertigo, visual dependence	49.80%	Adequate	–
Valancius (49)	108	EFA	2	Not specified	44.5%	Inadequate	–
van de Wyngaerde (50)	999	EFA	4	General factor (all items), physical manifestations, emotional impact, catastrophic impact	General factor: 70%, factor 1: 12.9%, factor 2: 5.9%, factor 3: 6.4%	Very good	+
	992	CFA		General factor (all items), physical manifestations, emotional impact, catastrophic impact	RMSEA = 0.070	Very good	–
Ahmed (53)	202	CFA	3	Based on the model of Perez et al. (56)	CFI = 0.94, TLI = 0.91, RMSEA = 0.08	Very good	–

References are placed in chronological order.

CFA indicates Confirmatory Factor Analysis; CFI, comparative fit index; EFA, Exploratory Factor Analysis; *n*, number of participants; TLI, Tucker-Lewis index; *, best fit model according to the corresponding article; Rating '+', sufficient; '−', insufficient; '?', indeterminate; RMSEA, Root Mean Square Error of Approximation, which should be below 0.06 according to COSMIN rating criteria, $; results based on 24 of the 25 questions of the DHI due to an administration error, model of fit parameters not reported.

from 14 (4) to 106 (20). Time between the test and retest ranged from ''the same day'' to ''15 days.'' We considered a time interval of >24 h ''appropriate'' and consequently rated the methodological quality of two studies as ''inadequate'' due to inappropriate (i.e., too short) time interval (4,20). In most of the studies, test conditions were not similar for the test and retest administrations of the DHI (e.g., the first test in the hospital and the retest at home, or as an online questionnaire, first test and retest in different language) with no rationale given to support this choice. The methodological quality of these studies is therefore either rated as ''inadequate'' (27,30,49,58,59), or when the test conditions were unclear as ''doubtful'' (34,35,43,47,54).

Since there is inconsistent information on the structural validity of the DHI and there is no evidence to support the original subscales we decided to merely report available data on reliability of the DHI total scores. The reported intraclass correlation coefficient (ICC) values for the DHI total score ranged from 0.81 to 0.99, which are all well above the minimal norm of 0.70. Based on these findings, we rated the overall ''reliability'' of the DHI total score as ''sufficient'' and graded the quality of evidence as ''low.''

Measurement Error

Ten studies reported data on the DHI's measurement error (Table 4), of which two studies did not report data that reflect measurement error properly according to COSMIN criteria. Jacobson and Newman reported a standard error of 6.23 points and consequently a true change of 18 points, however it is unclear how they calculated this (4). Also the ''± 2xSEM = 9.32'' reported by Enloe et al. (30) is ambiguous. Data on Smallest

TABLE 4. *Reliability and measurement error for DHI total score*

			Reliability ICC (95% CI)			Measurement Error			
Reference	*n*	Time Interval*	DHI Total Score	Study Quality	Rating	SDC	Limits of Agreement Mean Difference (Upper; Lower Limit)	Study Quality	Rating
Jacobson (4)	14	Same day	Not reported	Inadequate	?	Not reported		Inadequate	?
Enloe (30)	20	24–48 h	0.94	Inadequate	+	Not reported		Inadequate	?
Jarlsäter (57)	15	1 week	Not reported	Inadequate	–	Not studied		NA	NA
Poon (58)	49	1 week	0.87	Inadequate	+	Not studied		NA	NA
Kammerlind (59)	30[c]	2 days	0.95 (0.89–0.98)	Inadequate	+	11.2	−3 (−14.2–8.2)	Inadequate	?
	20[c]	2 days	0.94 (0.84–0.98)		+	Not studied		NA	NA
Vereeck (20)	106	1–8 h	0.99 (0.98)	Inadequate	+	8.25	−1.1 (−9.35–7.1)	Inadequate	?
de Castro (22)	25	15 days	0.99[a]	Adequate		4.7[a]	0.64 (−4.1;5.34)	Adequate	?
Tamber (27)	27	48 h	0.90	Inadequate	+	19.7	4 (−15.67–23.67)[b]	Inadequate	?
Kurre (26)	40	5.5 days	0.95 (0.91–0.98)	Adequate	+	12.4	1.8 (−10;14)	Adequate	?
Georgieva-Zhostova (34)	52	1 week	0.94 (0.88–0.97)	Doubtful	+	12.4	2.55 (−9.15;14)	Doubtful	?
Jafarzadeh (35)	30	2 days	0.96 (0.93–0.98)	Doubtful	+	19		Doubtful	?
Nikitas (43)	39	3 days	0.93 (0.85–0.96)	Doubtful	+	Not studied		NA	
Neupane (46)	20	10 days	0.81	Doubtful	+	Not studied		NA	NA
Soleimani (47)	113	13.1 days	>0.90	Doubtful	+	8	−8 to 8	Doubtful	?
Valancius (49)	65	1 week	0.90 (0.81–0.94)	Inadequate	+	Not studied		NA	NA
Zmnako (51)	59	1–5 days	0.93	Adequate	+	Not studied		NA	
Ahmed (53)	30[d]	8–15 days	Not reported	Inadequate	?				
Vozel (54)	50	3 days	0.95 (0.90–0.97)	Doubtful	+	Not studied		NA	NA

Reference are placed in chronological order. ICC values of the DHI sub scores are not reported, but can be requested from the authors.

DHI indicates Dizziness Handicap Inventory; LoA, Limits of Agreement; mean difference ± 1.96 * $SD_{difference}$ = mean difference (upper limit, lower limit) NA; not applicable Rating '+'; sufficient, '−'; insufficient, '?'; indeterminate, SDC; Smallest Detectable Change, 95% CI; 95% Confidence Interval, *; time interval as reported in the method section or in the result section.

[a]Calculated based on the data in the article.

[b]Read from the Bland-Altman plot in the article.

[c]Two different populations neurological (N = 30) and vestibular disorders (n = 20).

[d]Population of bilingual healthy students, English and Urdu versions of the DHI were used.

Detectable Change (SDC) or the Limits of Agreement (LoA) were reported or could be calculated from the presented data in eight studies (Table 4) (20,22,26,27,34,35,47,59). The methodological quality of the studies by Kurre et al. and Castro et al. (22,26) were rated as ''adequate.'' Two studies were rated as ''inadequate'' due to short time interval (20) or inconsistent conditions (59). Three other studies were rated as ''doubtful'' due to inconsistency in the context of the test conditions (34,35,47).

Reported SDC values range from 4.7 to 19.7 points. Sufficient information on the Minimal Important Change (MIC) is lacking, therefore not enough information is available to judge whether the SDC or LoA is smaller than the MIC. Therefore, we rated the overall ''measurement error'' as ''indeterminate'' and we could not grade the quality of evidence.

Criterion Validity

No golden standard for the DHI is available, therefore criterion validity is not applicable.

Construct Validity

Fifteen studies reported data on construct validity (4,19,21,23,25,27,34,40,43–45,47,51,55,59) (Table 5). The methodology of the studies was rated as ''adequate,'' because we assumed that the measurement properties of the comparator instruments were adequate, the correlated constructs were clear, and the statistical methods appropriate. The results were in line with the hypotheses for ''vestibular function'' (93.8%), ''anxiety and depression'' (80%), and ''handicap'' (87.5%). Less than 75% of the results are in line with our hypotheses for gait (69.2%) ''balance'' (57.6%), ''dizziness'' (57.1%), and ''quality of life'' (66.7%). Overall, 70.6% of our hypotheses are in line with the results, therefore we rated the overall ''construct validity hypothesis testing'' as ''borderline sufficient'' with the overall quality of evidence graded as ''moderate.''

Responsiveness

Table 6 shows the five studies which addressed responsiveness (25,27,30,33,58). Enloe et al. (30) used Guyatt's statistic, Poon et al. (58) describes effect sizes and Karapolat et al. (25) used a paired *t*-test, which are according to the COSMIN criteria inadequate methods to assess responsiveness, and therefore we did not include these studies in the evaluation of responsiveness. The methodological quality of the two remaining studies by Tamber et al. (27) and Friscia et al. (33) were rated as

TABLE 5. *Construct validity of the DHI—hypotheses testing*

Reference	n	Measurement	Correlation with DHI	Level of Correlation	Study Quality	Rating
Construct: Handicap						
Kammerlind (59)	30	UCLA-DQ	0.94	Very high	Adequate	−
		Dizziness Beliefs Scale (DBS)	0.64	Moderate		+
	20	UCLA-DQ	0.87	High		+
		Dizziness Beliefs Scale (DBS)	0.70	High		+
Tamber (27)	92	VSS-SF-V-N	0.64	Moderate	Adequate	+
		Disability Scale	0.58	Moderate		+
Zmnako (51)	301	VAS-T (overall handicap)	0.64	Moderate	Adequate	+
Karapolat (25)	33	ABC	−0.618	Moderate	Adequate	+
Construct: Dizziness						
Kammerlind (59)	30	VAS dizziness	0.39 (movement)	Low	Adequate	+
			0.32 (rest)			+
	20	VAS dizziness	0.04 (movement)	Very low		−
			0.05 (rest)			−
Vereeck (21)	214	VVAS	0.71	High	Adequate	−
Tamber (27)	92	VSS-SF-N	0.69	Moderate	Adequate	+
Soleimani (47)	113	VAS vertigo severity	0.44	Low	Adequate	+
Construct: Anxiety and depression						
Tamber (27)	92	VSS-SF-A-N	0.50	Moderate	Adequate	+
Formeister (45)	70	GAD-7	0.67	Moderate	Adequate	+
		PHQ-9	0.63	Moderate	Adequate	+
Soleimani (47)	113	HADS	0.64 (anxiety)	Moderate	Adequate	+
			0.77 (depression)	High		−
Construct: Quality of life						
Tamber (27)	92	COOP/WONCA	0.60	Moderate	Adequate	+
Formeister (45)	70	SF-36	0.66	Moderate	Adequate	+
Soleimani (47)	113	SF-36	0.79	High	Adequate	−
Construct: Gait						
Kammerlind (59)	30	Walking in a figure-of-eight	0.16	Very low	Adequate	−
	20	Walking in a figure-of-eight	−0.10	Very low		−
Vereeck (21)	214	TUG	0.57	Moderate	Adequate	+
		10-m test	0.56	Moderate		+
		Tandem Gait	−0.47	Low		+
		DGI	−0.69	Moderate		+
Tamber (27)	92	Gait speed (m/sec)	−0.36 (preferred)	Low	Adequate	+
			−0.40 (fast)			+
Niktitas (43)	90	FGA	−0.472	Low	Adequate	+
Karapolat (25)	33	TUG	0.16	Very low	Adequate	−
		Gait speed (m/sec)	0.02 (preferred)	Very low		−
		DGI	−0.31	Low		+
		FGA	−0.26	Low		+
Construct: Balance						
Jacobson (19)	367	SOT condition 2–6	−0.35 till − 0.42	Low	Adequate	+(4×)
Kammerlind (59)	30	Standing on one leg	0.10 (EO)	Very low	Adequate	−
		VAS balance problems	0.66	Moderate		−
	20	Standing on one leg	−0.36 (EO)	Low		+
		VAS balance problems	0.48	Low		+
Vereeck (21)	214	Romberg eyes closed with Jendrassik manoeuvre	−0.25	Very low	Adequate	−
		Standing on foam	−0.44 (EO)	Low		+
			−0.44 (EC)			+
		Tandem romberg	−0.46 (EO)	Low		+
			−0.45 (EC)			+
		Single leg stance	−0.51 (EO)	Moderate		−
			−0.42 (EC)	Low		+
		Sum score 7 static balance test	−0.54	Moderate		−
Georgieva-Zhostova (34)	79	Romberg coefficients = ratio CE/OE (on posturography)	0.38	Low	Adequate	+
Chiarovano (40)	90	Test on Wii balance board combined with virtual reality	−0.08	Very low	Adequate	−

TABLE 5 (*Continued*)

Reference	n	Measurement	Correlation with DHI	Level of Correlation	Study Quality	Rating
Yip (44)	84	EC total sway	0.216 (EC)	Very low	Adequate	−
			0.314 (EO)	Low		+
		Romberg ratio (EC/EO)	−0.081	Very low		−
Zmnako (51)	301	CTSIB-T	−0.39	Low	Adequate	+
Mbongo (23)	92	Values of energy in square anterior-posterior plane	0.14	Very low	Adequate	−
	65	Values of energy in square mediolaterale plane	0.15	Very low		−
Karapolat (25)	33	VAS	0.35	Low	Adequate	+
		Five time sit and stand	0.04	Very low		−
		Romberg	−0.26 (EO)	Low		+
			−0.35 (EC)			+
		Tandem Romberg	−0.34 (EO)	Low		+
			−0.22 (EC)	Very low		−
		Foam	−0.32 (EO)	Low		+
			−0.11 (EC)	Very low		−
Soleimani (47)	113	mBBS	0.71	High	Adequate	−
Construct: Vestibular system function						
Jacobson (4)	37	Degree of caloric asymmetry	0.09	Very low	Adequate	+
Jacobson (19)	367	VOR phase at 0.01 Hz	0.11	Very low	Adequate	+
		0.64 Hz chair frequencies	0.12	Very low		+
Yip (44)	618	vHIT (horizontal angular) gain VOR	0.007 (R)	Very low	Adequate	+
			−0.091 (L)			+
			0.013 (asymmetry)			+
		Unilateral (caloric) weakness (UW)	−0.018	Very low		+
		Total caloric respons	0.055	Very low		+
		oVEMP amplitude	−0.034 (R)	Very low		+
			−0.004 (L)			+
			0.016 (asymmetry)			+
		cVEMP amplitude	0.044 (R)	Very low		+
			−0.007 (L)			+
			− 0.008 (asymmetry)			+
Jacobson (55)	72	VOR gain	−0.22 (0.01Hz)	Very low	Adequate	+
			−0.38 (0.32Hz)	Low		−

cVEMP, cervical Vestibular Evoked Myogenic Potential; CTSIB-T, Clinical Test of Sensory Interaction on Balance Total score; DGI, Dynamic Gait Test; DHI, Dizziness Handicap Inventory; EC, Eyes Closed; EO, Eyes Open; FGA, Functional Gait Assessment; GAD-7, anxiety questionnaire; HADS, Hospital Anxiety and Depression Scale; Hz, Hertz; mBBS, modified Berg Balance Score; n, number of included patients for the analysis; oVEMP, Ocular Vestibular Evoked Myogenic Potential; PHQ-9, Patient Health Questionnaire-9; Rating '+', sufficient; '−', insufficient; '?', indeterminate; SF-36, Short Form Health survey-36; SOT, Sensory Organisation Test; TUG, Times Up and Go test; UCLA-DQ, UCLA Dizziness Questionnaire; VAS, Visual Analogue Scale; vHIT, video Head Impuls Test; VOR, Vestibular Ocular Reflex; VSS SF-N, Vertigo Symptom Scale Short-Form-Norway;VVAS, Visual Vertigo Analogue Scale.

''adequate.'' In both studies, the reported area under the curve (AUC) values were >0.70 which indicates good responsiveness. In contrast, only 65% (5 out of 8 correlations) were in line with our predefined hypotheses which is less than the preferred 75%. Therefore, we rated the overall evidence for ''responsiveness'' of the DHI as ''borderline sufficient'' and graded the overall quality of evidence as ''moderate.''

Interpretability and Feasibility

In none of the nine studies reporting data on the distribution of DHI scores floor or ceiling effects were demonstrated (26,27,29,30,34,35,48,49,51). Two studies reported a Minimal Important Change (MIC) for the DHI total score: 3 points with 4 to 6 weeks follow-up (33) and 11 points with 6 months follow-up (27) (Table 6).

DISCUSSION

We systematically reviewed the measurement properties of the DHI applying the PRISMA and COSMIN approaches. The review shows that the current evidence for a number of measurement properties of the DHI is suboptimal. Overall, sufficient information on its content validity is either lacking or limited and of low quality. Moderate evidence was found for inconsistent structural validity, sufficient construct validity, and borderline sufficient responsiveness. Based on the studies included low evidence was found for sufficient reliability of the DHI total score. No evidence synthesis could be done for the DHI's internal consistency and measurement error, and no evidence was found to support the use of the original three subscales (i.e., emotional, physical, functional).

TABLE 6. *Responsiveness and interpretability of the DHI total score*

Reference	n	Follow-Up	Mean Change DHI Total Score	ES	SRM	Correlations Between DHI and Comparative Instruments	AUC (95%CI)	MIC	Study Quality	Rating
Enloe (30)	31	6–8 weeks	11.94 ± 15.60		0,77[b]				NA	NA
Poon (58)	17	6–7 months	12.17		0.53				NA	NA
Tamber (27)	72	6 months	7.08 ± 10.21[a]	0.40[c]	0.69[b]	VSS: 0.57 COOP/ WONCA: 0.56 Disability scale: 0.51 Preferred gait: 0.10 Fast gait: 0.2	0.83 (0.71–0.94)	11	Adequate	AUC > 0.70 (+), hypothesis (3+, 2−)
Friscia (33)	45	4–6 weeks	6 ± 16	0.23	0.36	ABC: 0.66 FES-I: 0.53 VAP: 0.41 GROC: 0.61	0.80 (0.66–0.93)	3	Adequate	AUC >0.7 (+), hypothesis (3+, 1−)
Karapolat (25)	27	4 weeks	14.19[a]	0.59[a]					NA	NA

Reference are placed in chronological order.

ABC, Activities-specific Balance Confidence Scale; AUC (95%CI), Area Under The Curve (95% Confidence Interval); DHI, Dizziness Handicap Inventory; ES, Effect Size; FES-I, Falls Efficacy Scale International; GROC, Global Rating of Changes Scale; MIC, Minimal Important Clinical difference; NA, not applicable; Rating '+', sufficient; '−', insufficient; '?', indeterminate; SRM Standardised Response Mean; VAP, Vestibular Activities and Participation Scale; VSS, Vertigo Symptom Scale.

[a]Calculated based on data in the article—weighted means.

[b]Calculation based on data in the article SRM = mean change score / SD change score.

[c]Calculation based on data in the article ES = mean change score / SD baseline.

Strengths and Weaknesses of the Study

The main strength of this review is that it included not only the original development study by Jacobsen and Newman, but also all the relevant studies on the measurement properties of the DHI in the English language that have been published ever since. We used the extensive guidance of the current COSMIN recommendations (12–14) to ensure that all relevant properties were assessed in a standardized way (3). Clearly some form of subjectivity is inevitable when assessing risk of bias and grading of published evidence.

Comparison with Existing Literature

Several reviews have been published previously about the use of PROMs in patients with dizziness. The most recent review by Stewart et al. (10) also used the COSMIN criteria and evaluated multiple instruments. However, for the main assessment they only included the original article of Jacobson and Newman when evaluating the DHI's measurement properties. Stewart et al. (10) concluded that the DHI has ''robust clinimetric properties,'' but is not the first questionnaire of choice. This differs from our findings, the main discrepancy resulting from a difference in the rating of the content validity. Stewart et al. (10) rated the content validity of the DHI as ''excellent,'' but based on all published evidence we rated the content validity as ''insufficient.'' We rated the content validity as ''insufficient'' mainly due to a lack of cognitive interviews when developing and translating the DHI, which is an important methodological flaw. Moreover, based on the results of our systematic review, questions can be raised about the relevance of all individual items and the comprehensiveness of the DHI, aspects Stewart et al. considered to be adequate. Our findings are in line with the review by Duracinsky et al. (8), who also stated that there are doubts about the validity of the DHI.

In their original publication of the DHI (4), the authors describe the construct and the target population of the instrument in a limited way, which results in ambiguity. This is reflected, for example, in the fact that the DHI is often used to measure quality of life (8) which is a different construct than handicap. The ambiguity in the construct of the DHI is reflected in the structural validity, as multiple factors were found in exploratory factor analysis (4,19,21,23,25,27,34,43–45,47,51,55,59) which could imply multidimensionality and thus more than one underlying construct. The factors found differ from the original subscales and are different in the various studies (Table 3). Therefore, just like Van de Wyngaerde et al. (50), we advise to use only the DHI total score and not the original subscale scores. Another issue, addressed by Ryd and Rheault (41) based on their Rasch Analysis, is that the response categories of the DHI may better be transformed to a dichotomous scale (only ''Yes'' or ''No,'' instead of ''Yes,'' No, or ''Sometimes''). Later articles, however, have not further addressed this issue. A limitation in the development of the DHI is that patients were not involved in generating the different items of the questionnaire. Therefore, it is unclear if all items of the DHI are relevant and sufficiently comprehensive for the construct and what the impact of the underlying

vestibular disorder is, as a different perception of disability could be expected in different conditions (60–62). In most of the included studies, the patient sample consisted of a mix of diagnoses which could cause over- or underestimation of the DHI's psychometric qualities in a more specific population. This is undesirable if a researcher's aim is to establish effectiveness of an intervention in a homogeneous patient population.

Several studies regarding the DHI's test–retest reliability and measurement error were rated as ''inadequate'' due to different test conditions, for instance with a first completion in the hospital and the second at home (27,30,49,58,59). However, those different test conditions are often seen in research as well as in daily practice. Even with these varying test conditions the DHI total score has high and acceptable ICC values (Table 4). And although the DHI is mainly used as an outcome to measure treatment effects of interventions, only a limited number of articles have looked at the DHI's measurement error, responsiveness and interpretability parameters (such as the Minimal Important Change [MIC]), which are the most important measurement properties to draw meaningful conclusions about the (change in) score when evaluating interventions. We found two studies with data on the MIC, which reported very different values for the MIC (3 *versus* 11 points) and in both cases the MIC was smaller than or similar to the SDC. This implies that a clinical important change is within the range of the measurement error and consequently the MIC cannot be assessed.

Implication for Clinical Practice and Research

Our review shows that the measurement properties of the DHI are suboptimal. However, developing and validating a completely new questionnaire is very time consuming and, for that reason, not the most attractive choice. Therefore, we advise the next step to take is to consider other existing instruments that measure self-perceived handicap (especially the ones that have been developed more recently) and to critically evaluate their psychometric properties in the same way we did for the DHI. Although the review by Stewart et al. (10) was not sufficiently comprehensive for this purpose, they recommend the Vestibular Rehabilitation Benefit Questionnaire (VRBQ) (63,64) to assess treatment outcomes in patients with vestibular dysfunction, which scored the highest on the COSMIN criteria. At the same time, because the DHI is still widely used in clinical practice and in research, it is necessary to get an even better understanding of the measurement properties of the DHI. First, qualitative validation studies should focus on the content validity, especially on relevance and comprehensiveness, in order to gain more insight in the validity and completeness of the DHI for different diagnosis groups. Second, further research should focus on measurement error, responsiveness and interpretability parameters because the main reason for using the DHI is treatment evaluation and there is a lack of high-quality studies on these issues.

CONCLUSION

Based on our systematic literature review, we conclude that the current evidence for a number of measurement properties of the DHI is suboptimal. Because of its widespread use and the current lack of a better alternative researchers can use the DHI when assessing handicapping effects imposed by dizziness, but they should be aware of its limitations. Moreover, we recommend to use the total DHI score only and also to consider adding an instrument with more favorable measurement properties when assessing self-perceived handicap in patients with dizziness.

Acknowledgments: We would like to thank Simone Visser and Daphne Smit, clinical librarians at Gelre Hospitals, for their assistance in the literature searches and Hester van der Zaag and Roula Tsonaka for their methodological and statistical advice.

REFERENCES

1. Neuhauser HK, et al. Burden of dizziness and vertigo in the community. *Arch Intern Med* 2008;168:2118–24.
2. Ciorba A, et al. The impact of dizziness on quality-of-life in the elderly. *Eur Arch Oto-rhino-laryngol* 2017;274:1245–50.
3. Viergever K, et al. Questionnaires in otology: a systematic mapping review. *Syst Rev* 2021;10:119.
4. Jacobson GP, Newman CW. The development of the Dizziness Handicap Inventory. *Arch Otolaryngol Head Neck Surg* 1990; 116:424–7.
5. Hall CD, 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;40:124–55.
6. Mcdonnell MN, Hillier SL. Vestibular rehabilitation for unilateral peripheral vestibular dysfunction. *Cochrane Database Syst Rev* 2015;1:CD005397.
7. Mutlu B, Serbetcioglu B. Discussion of the dizziness handicap inventory. *J Vestib Res* 2013;23:271–7.
8. Duracinsky M, et al. Literature review of questionnaires assessing vertigo and dizziness, and their impact on patients' quality of life. *Value Health* 2007;10:273–84.
9. Fong E, Li C, Aslakson R, Agrawal Y. Systematic review of patient-reported outcome measures in clinical vestibular research. *Arch Phys Med Rehabil* 2015;96:357–65.
10. Stewart VM, Mendis MD, Low Choy N. A systematic review of patient-reported measures associated with vestibular dysfunction. *Laryngoscope* 2018;128:971–81.
11. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
12. Prinsen CAC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil* 2018;27:1147–57.
13. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res Int J Qual life Asp Treat Care Rehabil* 2018;27:1171–9.
14. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil* 2018;27:1159–70.
15. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.

16. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil* 2009;18:1115–23.
17. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210.
18. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 2012;24:69–71.
19. Jacobson GP, Newman CW, Hunter L, Balzer GK. Balance function test correlates of the Dizziness Handicap Inventory. *J Am Acad Audiol* 1991;2:253–60.
20. Vereeck L, Truijen S, Wuyts F, Van de Heyning PH. Test-retest reliability of the Dutch version of the Dizziness Handicap Inventory. *B-ENT* 2006;2:75–80.
21. Vereeck L, Truijen S, Wuyts FL, De Heyning V. The dizziness handicap inventory and its relationship with functional balance performance. *Otol Neurotol* 2007;28:87–93.
22. Castro AS, Gazzola JM, Natour J, Ganança FF. Brazilian version of the dizziness handicap inventory. *Pro Fono Versão Bras do dizziness Handicap Invent* 2007;19:97–104.
23. Mbongo F, Tran Huy P, Vidal P-P, de Waele C. Relationship between dynamic balance and self-reported handicap in patients who have unilateral peripheral vestibular loss. *Otol Neurotol* 2007;28:905–10.
24. Vereeck L, Truijen S, Wuyts FL, De Heyning V. Internal consistency and factor analysis of the Dutch version of the Dizziness Handicap Inventory. *Acta Otolaryngol* 2007;127:788–95.
25. Karapolat H. Reliability, validity and sensitivity to change of Turkish Dizziness Handicap Inventory (DHI) in patients with unilateral peripheral vestibular disease. *Int Adv Otol* 2009;5:237–45.
26. Kurre A, et al. Translation, cross-cultural adaptation and reliability of the German version of the Dizziness Handicap Inventory. *Otol Neurotol* 2009;30:359–67.
27. Tamber A-L, Wilhelmsen KT, Strand LI. Measurement properties of the Dizziness Handicap Inventory by cross-sectional and longitudinal designs. *Health Qual Life Outcomes* 2009;7:101.
28. Kurre A, Bastiaenen CH, van Gool CJ, Gloor-Juzi T, de Bruin ED, Straumann D. Exploratory factor analysis of the Dizziness Handicap Inventory (German version). *BMC Ear Nose Throat Disord* 2010;10:3.
29. Nola G, Mostardini C, Salvi V, Ercolani AP, Ralli G. Validity of Italian adaptation of the Dizziness Handicap Inventory (DHI) and evaluation of the quality of life in patients with acute dizziness. *Acta Otorhinolaryngol Ital* 2010;30:190–7.
30. Enloe LJ, Shields RK. Evaluation of health-related quality of life in individuals with vestibular disease using disease-specific and general outcome measures. *Phys Ther* 1997;77:890–903.
31. Goto F, Tsutsumi T, Ogawa K. The Japanese version of the dizziness handicap inventory as an index of treatment success: Exploratory factor analysis. *Acta Otolaryngol* 2011;131:817–25.
32. Alsanosi AA. Adaptation of the dizziness handicap inventory for use in the Arab population. *Neurosciences (Riyadh)* 2012;17:139–44.
33. Friscia LA, Morgan MT, Sparto PJ, Furman JM, Whitney SL. Responsiveness of self-report measures in individuals with Vertigo. *Dizziness Unsteadiness Otol Neurotol* 2014;35:884–8.
34. Georgieva-Zhostova S, Kolev OI, Stambolieva K. Translation, cross-cultural adaptation and validation of the Bulgarian version of the Dizziness Handicap Inventory. *Qual Life Res* 2014;23:2103–7.
35. Jafarzadeh S, Bahrami E, Pourbakht A, Jalaie S, Daneshi A. Validity and reliability of the Persian version of the dizziness handicap inventory. *J Res Med Sci* 2014;19:769–75.
36. Sarda SA, Vanaja CS. Correlating two questionnaires: Vestibular activities and participation (VAP) and dizziness handicap inventory (DHI). *J Hear Sci* 2015;5:33–40.
37. Akhilesh. Development and standardization of dizziness handicap inventory (DHI) in the Indian language Kannada. *Int J Appl Res* 2016;2:340–2.
38. Akhilesh. Development and standardization of dizziness handicap inventory (DHI) in the Indian language Malayalam. *Int J Appl Res* 2016;2:26–8.
39. Ardıç FN, Tümkaya F, Akdağ B, Şenol H. The subscales and short forms of the dizziness handicap inventory: Are they useful for comparison of the patient groups? *Disabil Rehabil* 2017;39:2119–22.
40. Chiarovano E, MacDougall HG, Wang W, Reynolds P. Imbalance: Objective measures versus subjective self-report in clinical practice. *Gait Posture* 2018;59:217–21.
41. Ryd J, Rheault W. Rasch analysis of the Dizziness Handicap Inventory. *J Rehabil Outcomes Meas* 1998;2:17–24.
42. Colnaghi S, Rezzani C, Gnesi M, et al. Validation of the Italian Version of the Dizziness Handicap Inventory, the Situational Vertigo Questionnaire, and the Activity-Specific Balance Confidence Scale for Peripheral and Central Vestibular Symptoms. *Front Neurol* 2017;8:528.
43. Nikitas C, Kikidis D, Katsinis S, Kyrodimos E, Bibas A. Translation and validation of the dizziness handicap inventory in Greek language. *Int J Audiol* 2017;56:936–41.
44. Yip CW, Strupp M. The Dizziness Handicap Inventory does not correlate with vestibular function tests: A prospective study. *J Neurol* 2018;265:1210–8.
45. Formeister EJ, Krauter R, Kirk L, Zhu TR, Rizk HG, Sharon JD. Understanding the Dizziness Handicap Inventory (DHI): A cross sectional analysis of symptom factors that contribute to DHI variance. *Otol Neurotol* 2020;41:86–93.
46. Neupane AK, Kapasi A, Patel N. Psychometric features of Dizziness Handicap Inventory (DHI): Development and standardization in gujarati language. *Int Tinnitus J* 2019;23:86–90.
47. Soleimani R, Jalali MM, Bakhshayesh B, Rashidi Mojdehi P, Ghadiri Asli SMS. Psychometric properties of the Persian Version of Dizziness Handicap Inventory. *Iran J Otorhinolaryngol* 2019;31:359–67.
48. Szostek-Rogula S, Zamysłowska-Szmytke E. Validation of the Polish version of the Dizziness Handicap Inventory. *Med Pr* 2019;70:529–34.
49. Valančius D, et al. Validation and factor analysis of the Lithuanian Version of the Dizziness Handicap Inventory. *J Int Adv Otol* 2019;15:447–53.
50. Van De Wyngaerde KM, et al. The component structure of the Dizziness Handicap Inventory (DHI): A reappraisal. *Otol Neurotol* 2019;40:1217–23.
51. Zmnako SSF, Chalabi YI. Reliability and validity of a Central Kurdish version of the Dizziness Handicap Inventory. *Sci Rep* 2019;9:8542.
52. Asmundson GJG, Stein MB, Ireland D. A factor analytic study of the dizziness handicap inventory: Does it assess phobic avoidance in vestibular referrals? *J Vestib Res Equilib Orientat* 1999;9:63–8.
53. Ahmed A, Aqeel M, Chughtai NA. Indigenous context of vertigo: Translation and validation of dizziness handicap inventory for diagnosis and evaluation of patients in Pakistani hospitals. *Int J Hum Rights Healthc* 2021;14:87–99.
54. Vozel D, Steiner N, Božanić Urbančič N, Mladenov D, Battelino S. Slovenian cross-cultural adaptation and validation of health-related quality of life measures for Chronic Otitis Media (COMQ-12), Vertigo (DHI, NVI) and TINNITUS (THI). *Zdr Varst* 2020;59:120–7.
55. Jacobson GP, Calder JH. Self-perceived balance disability/handicap in the presence of bilateral peripheral vestibular system impairment. *J Am Acad Audiol* 2000;11:76–83.
56. Perez N, Garmendia I, García-Granero M, Martin E, García-Tapia R. Factor analysis and correlation between Dizziness Handicap Inventory and dizziness characteristics and impact on quality of life scales. *Acta Oto-Laryngol Suppl* 2001;545:145–54.
57. Jarlsäter S, Mattsson E. Test of reliability of the Dizziness Handicap Inventory and the activities-specific balance confidence scale for use in Sweden. *Adv Physiother* 2003;5:137–44.
58. Poon DMY, Chow LCK, Hui Y, Au DKK, Leung MCP. Translation of the Dizziness Handicap Inventory into Chinese, validation of it, and evaluation of the quality of life of patients with chronic dizziness. *Ann Otol Rhinol Laryngol* 2004;113:1006–11.
59. Kammerlind A-S, Larsson PB, Ledin T, Skargren E. Reliability of clinical balance tests and subjective ratings in dizziness and disequilibrium. *Adv Physiother* 2005;7:96–107.
60. Graham MK, Staab JP, Lohse CM, McCaslin DL. A comparison of Dizziness Handicap Inventory scores by categories of vestibular diagnoses. *Otol Neurotol* 2021;42:129–36.
61. Lucieer FMP, Van Hecke R, van Stiphout L, et al. Bilateral vestibulopathy: beyond imbalance and oscillopsia. *J Neurol* 2020;267:241–55.

62. Dobbels B, Lucieer F, Mertens G, et al. Prospective cohort study on the predictors of fall risk in 119 patients with bilateral vestibulopathy. *PLoS One* 2020;15:e0228768.
63. Morris AE, Lutman ME, Yardley L. Measuring outcome from Vestibular Rehabilitation, Part I: Qualitative development of a new self-report measure. *Int J Audiol* 2008;47:169–77.
64. Morris AE, Lutman ME, Yardley L. Measuring outcome from vestibular rehabilitation, part II: Refinement and validation of a new self-report measure. *Int J Audiol* 2009;48:24–37.
65. JL R, Rheault W. Rasch analysis of the Dizziness Handicap Inventory. *J Rehabil Outcomes Meas* 1998;2:17–24.
66. Jacobson GP, Calder JH. A screening version of the dizziness handicap inventory (DHI-S). *Am J Otol* 1998;19:804–8.
67. Tesio L, Alpini D, Cesarani A, Perucca L. Short form of the dizziness handicap inventory: Construction and validation through Rasch analysis. *Am J Phys Med Rehabil* 1999;78:233–41.
68. Kammerlind A-S, Bladström M, Svensson K. Test-retest reliability of two short Swedish versions of the Dizziness Handicap Inventory. *Adv Physiother* 2011;13:50–5.
69. Piker EG, Jacobson GP, Tran AT, McCaslin DL, Hale ST. Spouse perceptions of patient self-reported vertigo severity and dizziness. *Otol Neurotol* 2012;33:1034–9.
70. McCaslin DL, Jacobson GP, Lambert W, English LN, Kemph AJ. The development of the vanderbilt pediatric dizziness handicap inventory for patient caregivers (DHI-PC). *Int J Pediatr Otorhinolaryngol* 2015;79:1662–6.
71. Chen W, Shu L, Wang Q, et al. Validation of 5-item and 2-item questionnaires in Chinese version of Dizziness Handicap Inventory for screening objective benign paroxysmal positional vertigo. *Neurol Sci* 2016;37:1241–6.
72. Zanotto D, Mamuyac EM, Chambers AR, et al. Dizziness Handicap Inventory score is highly correlated with markers of gait disturbance. *Otol Neurotol* 2017;38:1490–9.
73. Masuda K, Goto F, Matsunaga T, Okami K, Iida M. Evaluation of a target age of the Dizziness Handicap Inventory for Patient Caregivers(DHI-PC). *Equilib Res* 2019;78:267–73.
74. Tiwari D, Goldberg A, Yorke A, Marchetti G, Alsalaheen B. Measurement properties of 'Dizziness Handicap Inventory-Children and Adolescent' in children post-concussion. *Arch Phys Med Rehabil* 2019;100:e95–6.
75. van Vugt VA, de Vet HCW, van der Wouden JC, van Weert HCPM, van der Horst HE, Maarsingh OR. The 25-item Dizziness Handicap Inventory was shortened for use in general practice by 60 percent. *J Clin Epidemiol* 2020;126:56–64.