B-ENT, 2006, 2, 75-80

Test-retest reliability of the Dutch version of the Dizziness Handicap Inventory

L. Vereeck*,**, S. Truijen***, F. Wuyts**,**** and P. H. Van de Heyning*****

*Division of Neuro- and Psychomotor Physical Therapy, Department of Health Care, University College Antwerpen, Van Aertselaerstraat 31, 2170 Merksem, Belgium; **Vestibular Function Laboratory, Department of Otorhinolaryngology, Antwerp University Hospital, Wilrijkstraat 10, 2650 Edegem, Belgium; ***Department of Health Care, University College Antwerpen, Van Aertselaerstraat 31, 2170 Merksem (Antwerpen), Belgium; ****Faculty of Science (Department of Biomedical Physics), University of Antwerp, Groenenborgerlaan 171, 2020 Antwerpen, Belgium; *****Faculty of Medicine, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk (Antwerpen), Belgium

Key-words. Dizziness; unsteadiness; disability; handicap; questionnaire

Abstract. *Test-retest reliability of the Dutch version of the Dizziness Handicap Inventory.* In the last fifteen years the Dizziness Handicap Inventory (DHI) has gained wide acceptance as a useful measure of handicap resulting from dizziness and unsteadiness. The objective of this study was to calculate measurement error and test-retest reliability of the Dutch version of the DHI. The translation into Dutch was performed according to a double translation method. One hundred and six patients with balance problems (range 26-78 years), scheduled for vestibular rehabilitation, completed the DHI on two occasions on the same day. The test-retest reliability was excellent, with Intraclass Correlation Coefficients ranging from 0.94 to 0.99 for DHI sub-scores and DHI total score. Most weighted kappa values (xW) exceeded 0.80 indicating substantial item per item test-retest reliability. Over 80% agreement was noted for all items except for item 8 (74%). Item 8 asks whether the subject feels dizzy or unsteady while doing ambitious activities like sports, dancing and household activities. Consequently measurement errors were calculated, suggesting that, after an intervention, the pre-treatment DHI total score should at least decrease with 12 points (lower bound 99% confidence interval for a true change) before the intervention could be said to be effective for an individual patient. Based on these data, the Dutch version of the DHI showed itself to be a highly reliable instrument to assess the self-perceived handicap imposed by vestibular system diseases.

Introduction

In the last fifteen years the Dizziness Handicap Inventory (DHI) has gained wide acceptance as a useful measure of handicap resulting from dizziness and unsteadiness.¹⁻⁴ It has been used to describe the self-perceived handicapping effects imposed by vestibular system diseases in specific or non-specific patient populations and it has been used as an outcome measure in clinical trials to document the effect of medical, surgical, or rehabilitative interventions.

The DHI is a 25-item, validated, self-reported questionnaire designed to evaluate the precipitating physical factors associated with dizziness and unsteadiness as well as the functional and emotional consequences of vestibular system disease.⁵ Each item is answered with No (0 points), Sometimes (2 points) and Yes (4 points). Scores on the DHI range from 0 to 100 and can be further subdivided into physical (28 points), functional (36 points) and emotional (36 points) subscores. The higher the score, the greater is the perceived handicap.

Because the consequences of disorders of the vestibular system affect different domains of everyday life, the DHI is often used in conjunction with other assessment tools. Conventional vestibulometric techniques, computerized dynamic posturography and functional balance performance can be useful in identifying the presence and nature of a disorder, but they provide little information regarding the impact the disorder has upon the individual's day-to-day life. Therefore it is very useful to have access to a condition-specific health status measure for persons with vestibular disease.

Through the years the DHI has been translated into different languages.⁶⁹ Screening versions have been proposed and several authors have assessed its internal consistency and factor structure.^{1,8-12} Its relation with vestibular function tests, balance performance measures and generic health status assessments have extensively been documented.^{9,13-21} Psychometric properties of the Dutch translation have not yet been published. The purposes of the present investigation were twofold. In a prospective study we wanted to document test-retest reliability of the Dutch version of the DHI and calculate its measurement error.

Materials and Methods

The DHI was translated and adapted from its original version to the Dutch language following an established double translation method.22 A professional translator and a bilingual Antwerp University staff co-worker did an initial English to Dutch translation. This translation was assessed item-byitem by staff members of the ENT department at the Antwerp University Hospital resulting in a final version. This version was then translated into English by two different unrelated translators of whom one was a native English speaker frequently involved in back translations. A similar process of assessment of the correlation between the original and final English versions was done. The final Dutch version of the Dizziness Handicap Inventory is shown in Table 1.

From August 2003 to February 2005, patients with balance system disease referred for vestibular rehabilitation were included in a prospective study to determine the intra-subject scoring variability of their self-perceived balance disability. Patients were asked to complete the DHI on two occasions i.e. prior to and after vestibular testing, clinical balance testing and the drawing up of a customised vestibular rehabilitaprogramme. The time tion between test and DHI-retest session ranged from 1 to 8 hours. Subjects were recruited with oral informed consent obtained during

their attendance at the vestibular function laboratory. No one declined the invitation to participate.

The DHI was completed independently by each subject, or, if he/she was unable to read the questions, these were read to the subject by either a physical therapist or an accompanying family member. The examiner was present to elucidate problems concerning the content of the questions. Immediately after completion, the questionnaire was checked for missing answers.

The independent samples t test was used to compare the DHI total scores of men and women and the paired samples t test was used to see whether the DHI scores differed from the retest scores.

Test-retest reliability was assessed by comparing the DHI test-results prior to and after vestibulometric and balance testing. Test-retest reliability was determined for all ordinal (trichotomous scaled) items by calculating weighted Kappa values and percentage of agreement. Testretest reliability for the subtotals and total score was examined by means of Intraclass Correlation Coefficients (ICC). Item per item reliability was established when the weighted Kappa value exceeded 0.60 as indicated by Landis and Koch (< 0.00: poor; 0.00-0.20: slight; 0.21-0.40: fair; 0.41-0.60: moderate; 0.61-0.80: substantial; 0.81: almost perfect) or when more then 80% agreement was observed.23 Test-retest reliability of the subscales and DHI total score was reached when the ICC was 0.85 or higher.24

Measurement error was determined by calculating the lower $(\hat{d} - t_{(n-1; 0.01)}.SD(d))$ and upper $(\hat{d} + t_{(n-1; 0.01)}.SD(d))$ bounds of the prediction interval using d (= mean difference between test and retest), $t_{(n-1; 0.01)}$ (t-distribution for n-1 degrees of freedom and 0.01 level of significance) and SD(d) (= Standard Deviation of the differences).²⁵

Version 12.0 of SPSS (SPSS. Inc., Chicago, IL, USA) and StatXact[®] 7 (Cytel Inc., Cambridge, MA, USA) were used for the analysis.

Results

One hundred and six patients were included in the reliability study. The means, standard deviations, standard error of the mean and ranges of age, DHI test and retestscores for the total sample and men and women separately are shown in Table 2. Of these 106 patients, 59 had a vestibular schwannoma (prior to or after tumour resection). 16 had Menière's disease, 10 had a documented vestibular hypofunction, 10 had bilateral vestibular hypofunction and 11 subjects had balance problems without a specific diagnosis and vestibular testing with electronystagmography within normal limits. The total DHI score for the total sample prior to vestibular testing was 34 (\pm 25.7) with men having lower DHI scores (men: 29 (\pm 22.7); women: 39 (\pm 27.7); independent samples t test: p = 0.04). The DHI total score and the physical, emotional and functional sub-scores for the vestibular schwannoma group were 20.1 (± 19.2), 7.7 (± 7), 4.6 (± 6.5) and 7.7 (± 8.3) , respectively. Mean differences between DHI test and retest scores are given in Table 3. These differences were small but significant for the total score (paired samples t test: p =

Table 1

Dutch version of the Dizziness Handicap Inventory

Wilt u van d Indier	Duizeligheid Handicap Inventaris (DHI) oel van deze vragenlijst is te bepalen in hoeverre u moeilijkheden ondervindt door uw probleem van duizelig u de vragen beantwoorden met ja, nee of soms door in het overeenkomstig vakje een kruis te schrijven. B e vragen moet u steeds voor ogen houden dat ze betrekking hebben op uw probleem van duizeligheid en in n u een situatie die we beschrijven niet hebt ervaren, probeer dan te denken aan een vergelijkbare situatie iden en antwoord voor die situatie.	ij het b 1stabili	eantwo teit.	oorden
		ja	neen	soms
P1	Neemt uw probleem toe wanneer u naar boven kijkt?			
E2	Voelt u zich gefrustreerd door uw probleem?			
F3	Beperkt u het reizen door uw probleem (zowel op privé- als op beroepsvlak)?			
P4	Neemt uw probleem toe wanneer u in de supermarkt tussen de rekken loopt?			
F5	Is het moeilijk om uit bed te komen door uw probleem?			
F6	Beperkt uw probleem ingrijpend uw sociale leven (uit eten gaan, naar de film, gaan dansen,)?			
F7	Wordt lezen bemoeilijkt door uw probleem?			
P8	Neemt uw probleem toe wanneer u meer actief bent zoals bij sporten, dansen, het huishouden doen (poetsen, de vaat wegzetten,)?			
E9	Bent u, door uw probleem, bang om het huis te verlaten zonder dat iemand u vergezelt?			
E10	Door uw probleem, voelt u zich beschaamd in bijzijn van anderen?			
P11	Neemt uw probleem toe door snelle hoofdbewegingen?			
F12	Vermijdt u hoogtes door uw probleem?			
P13	Neemt uw probleem toe bij het omdraaien in uw bed?			
F14	Door uw probleem, is het moeilijk om inspannend werk te doen in huis of in de tuin?			
E15	Door uw probleem, bent u bang dat mensen zouden denken dat u dronken bent?			
F16	Door uw probleem, kunt u moeilijk alleen wandelen?			
P17	Neemt uw probleem toe bij het wandelen op het voetpad?			
E18	Door uw probleem, kunt u zich moeilijk concentreren?			
F19	Door uw probleem, hebt u moeilijkheden om in het donker in uw huis te lopen?			
E20	Door uw probleem, heeft u angst om alleen thuis te blijven?			
E21	Voelt u zich gehandicapt door uw probleem?			
E22	Heeft uw probleem voor spanning gezorgd in uw relatie met familie of vrienden?			
E23	Voelt u zich depressief door uw probleem?			
F24	Heeft uw probleem invloed op uw verantwoordelijkheden in uw beroep of uw taken thuis?			
P25	Neemt uw probleem toe wanneer u zich bukt?		1	

P: physical; F: functional; E: emotional

Original reference:

Jacobson GP, Newman CW. Arch Otoloaryngol Head Neck Surg. 1990;116:424-427.

0.008) and the emotional subscale (paired samples *t* test: p = 0.026). Consequently measurement errors were calculated, suggesting that, after an intervention, the pre-treatment DHI-score should at least decrease with 12 points (lower bound 99% confidence interval for a true change (Table 3)) before

the intervention could be said to be effective for an individual patient.

Test-retest reliability scores for the DHI total score and the emotional, physical and functional sub-scores are presented in Table 4 and were all excellent and significant. Intraclass Correlation Coefficients ranged between 0.94 and 0.99.

As can be seen in Table 5, most weighted kappa values (π W) exceeded 0.80 indicating substantial item per item test-retest reliability. A moderate, but still sufficient, test-retest reliability was established for items 8 (π W: 0.74), Table 2

Characteristics of patients involved in the DHI reliability study. Test and retest DHIscores are included

	Mean	SD	SEM	Range
Age all (n = 106)	53.6	11.5	1.11	26.7 - 78.9
Age men $(n = 54)$	54.0	10.8	1.47	32.5 - 76.7
Age women $(n = 52)$	53.2	12.2	1.69	26.7 - 78.9
DHI-physical (all)	11.8	8.4	0.81	0 - 28
DHI-emotional (all)	8.8	8.9	0.87	0 - 34
DHI-functional (all)	13.4	10.8	1.05	0 - 36
DHI-total (all)	34.0	25.7	2.50	0 – 96
DHI-total (men)	29.0	22.7	3.09	0 - 82
DHI-total (women)	39.2	27.7	3.85	0 – 96
Retest DHI-physical (all)	11.5	8.5	0.82	0-28
Retest DHI-emotional (all)	8.4	9.1	0.88	0 - 36
Retest DHI-functional (all)	13.1	10.5	1.02	0 - 36
Retest DHI-total (all)	32.9	25.5	2.48	0 – 96
Retest DHI-total (men)	27.6	22.4	3.05	0 - 86
Retest DHI-total (women)	38.4	27.5	3.82	0 – 96
			1	

DHI: Dizziness Handicap Inventory; SD: Standard Deviation; SEM: Standard Error of the Mean.

Table 3

Mean differences (diff) between Dizziness Handicap Inventory (DHI) test- and retestscores, p-value (paired samples *t* test) and lower and upper bounds of 95% and 99% prediction intervals for the total sample (n = 106) and the vestibular schwannoma (VS) subsample (n = 59)

subsumple (i = 57)							
Total sample	Mean diff	SD diff	p-value	Lower 95	Upper 95	Lower 99	Upper 99
DHI total	-1.11	4.21	0.008	-9	7	-12	10
DHI emotional	-0.47	2.15	0.026	-5	4	-6	5
DHI physical	-0.38	2.86	0.177	-6	5	-8	7
DHI functional	-0.26	2.23	0.225	-5	4	-6	6
VS subsample	Mean diff	SD diff	p-value	Lower 95	Upper 95	Lower 99	Upper 99
DHI total	-0.78	3.68	0.109	-8	7	-11	9
DHI emotional	-0.37	1.68	0.094	-4	3	-5	4
DHI physical	-0.71	2.70	0.047	-6	5	-8	6
DHI functional	0.31	1.89	0.219	-3	4	-5	5

SD: Standard Deviation.

Table 4

Intraclass Correlation Coefficients (ICC) determining test-retest reliability of the Dizziness Handicap Inventory (DHI)

	ICC value (95% lower confidence limit)	p-value
DHI – total score	0.99 (0.98)	p < 0.001
DHI – functional subscore	0.98 (0.97)	p < 0.001
DHI – emotional subscore	0.97 (0.96)	p < 0.001
DHI – physical subscore	0.94 (0.92)	p < 0.001

10 (κ W: 0.79) and 11 (κ W: 0.79). Over 80% agreement was noted for all items except for item 8 (74%).

Discussion

Except for Jacobson and Newman's original study, the mean DHI total score in our population was lower (34 ± 26) , when compared to other studies.57,9,14-17 This may be due to the large number of patients (n = 20) with an as-yet-non-operated vestibular schwannoma in our patient population. Gradual dysfunction of the vestibular nerve caused by a vestibular schwannoma is usually accompanied by vestibular compensation due to central nervous system plasticity thus minimizing the resultant symptoms.

In Table 6 the test-retest reliability scores cited by other investigators are shown. Our study confirms the data in Jacobson and Newman's original test-retest reliability study, where, as in our study both tests were completed the same day.⁵ In their study, Pearson product-moment correlations were ranging from 0.92 to 0.97. Our data are also very close to the data gathered by Enloe and Shields¹⁷, where a 24 to 48 hours interval was used (ICC's: 0.79 -0.95). Nyabenda et al.6 obtained high values using the French translation in a normative sample with a test-retest period ranging from 8 to 10 weeks (ICC's: 0.96 -0.98). In 71 patients, a high degree of test-retest reliability (1 week) of the Chinese DHI was found in the emotional subscore (ICC: 0.83), functional subscore (ICC: 0.84) and the total score (ICC: 0.87), but a lower level of reliability in the physical subscore (ICC: 0.64).8 Weighted kappa was 0.69

Table 5

Weighted kappa and percentage of agreement for test-retest agreement presented for all subjects (n = 106) (DHI: Dizziness Handicap Inventory; P: physical; F: functional; E: emotional)

L. chiotohal)							
DHI - item	Weighted x ^a	99% Lcl ^b	99% Ucl ^c	% ^d			
P 1	0.83	0.70	0.96	85			
E 2	0.85	0.73	0.97	87			
F 3	0.97	0.93	1.00	94			
P 4	0.93	0.85	1.00	91			
F 5	0.93	0.84	1.00	95			
F 6	0.85	0.74	0.97	85			
F 7	0.85	0.71	1.00	92			
P 8	0.74	0.60	0.89	74			
E 9	0.97	0.89	1.00	98			
E 10	0.79	0.61	0.98	88			
P 11	0.79	0.64	0.93	81			
F 12	0.94	0.86	1.00	94			
P 13	0.82	0.64	1.00	91			
F 14	0.85	0.75	0.96	84			
E 15	0.96	0.92	1.00	95			
F 16	0.87	0.74	0.99	90			
P 17	0.86	0.77	0.95	82			
E 18	0.84	0.72	0.96	85			
F 19	0.86	0.74	0.97	87			
E 20	0.81	0.56	1.00	94			
E 21	0.84	0.73	0.95	82			
E 22	0.96	0.89	1.00	97			
E 23	0.93	0.86	1.00	93			
F 24	0.91	0.85	0.97	87			
P 25	0.84	0.72	0.97	85			

^a value of the calculated weighted kappa (\varkappa)

^b lower value of the 99% confidence limit of the weighted kappa

[°] upper value of the 99% confidence limit of the weighted kappa

^d percentage of agreement.

(1 week test-retest interval) for the Swedish version.⁷

As can be seen in Table 3, a significant 1 point averaged improvement of the total DHI-scores was observed in our population. This improvement, when compared with the calculated measurement error, being 12 points, is clinically irrelevant. However we cannot exclude that the vestibular manipulation in between sessions could have influenced the patients perception of emotional and physical aspects of their dizziness and balance problem. The measurement error calculated from our data was smaller than the 18 points (95% confidence interval for a true change) suggested by Jacobson and Newman, and was in agreement with Enloe and Shields who, using a 95% confidence interval, suggested a minimal decrease of 9.22 points on the DHI total score to document improvement in selfperceived handicap in the individual patient.5,17

Conclusion

The Dutch version of the DHI showed to be a highly reliable instrument and a decrease with at

Study	n	Statistics	Test-retest period	Reliability DHI-T	Reliability DHI-P	Reliability DHI-F	Reliability DHI-E
Jacobson and Newman (1990)	14	Pearson	Same day	r = 0.97	r = 0.92	r = 0.94	r = 0.97
Enloe and Shields (1997)	24	Pearson	24-48 hours	r = 0.96	r = 0.82	r = 0.94	r = 0.95
Enloe and Shields (1997)	24	ICC	24-48 hours	r = 0.94	r = 0.79	r = 0.95	r = 0.95
Jarlsäter and Mattson (2003)	15	жW	1 week	$\varkappa W = 0.69$			
Poon <i>et al.</i> (2004)	49	ICC	1 week	r = 0.87	r = 0.64	r = 0.84	r = 0.83
Nyabenda et al. (2004)*	47	ICC	8-10 weeks	r = 0.98	r = 0.97	r = 0.97	r = 0.96
Present study	106	ICC	Same day	r = 0.99	r = 0.94	r = 0.98	r = 0.97

 Table 6

 Previously published test-retest reliability values for the Dizziness Handicap Inventory (DHI)

ICC: Intraclass Correlation Coefficient; xW: weighted kappa; DHI-T: DHI total score; DHI-P: DHI physical subscore; DHI-F: DHI functional subscore; DHI-E: DHI emotional subscore.

*: normative study.

least 12 points is needed to demonstrate a significant change in an individual patient's selfperceived dizziness handicap.

Acknowledgements

Anja Van der Stappen (MD), Kris Van de Poel (PhD), Kevin McGuire and Josie Meaney for their help in translating the questionnaire.

References

- Yardley L, Donovan-Hall M, Smith HE, Walsh BM, Mullee M, Bronstein AM. Effectiveness of primary care-based vestibular rehabilitation for chronic dizziness. *Ann Intern Med.* 2004;141:598-605.
- Topuz O, Topuz B, Ardiç FN, Sarhus M, Ögmen G, Ardiç F. Efficacy of vestibular rehabilitation on chronic unilateral vestibular dysfunction. *Clin Rehabil.* 2004;18: 76-83.
- 3. Corna S, Nardone A, Prestinari A, Galante M, Grasso M, Schieppati M. Comparison of Cawthorne-Cooksey exercises and sinusoidal support surface translations to improve balance in patients with unilateral vestibular deficit. *Arch Phys Med Rehabil.* 2003;84:1173-1184.
- Badke MB, Pyle GM, Shea T, Miedaner J. Outcomes in vestibular ablative procedures. *Otol Neurotol.* 2002;23:504-509.
- 5. Jacobson GP, Newman CW. The development of the Dizziness Handicap Inventory. *Arch Otolaryngol Head Neck Surg.* 1990;116:424-427.
- 6. Nyabenda A, Briart C, Deggouj, Gersdorff M. Normative study and reliability of French version of the dizziness handicap inventory [in French]. *Ann Readapt Med Phys.* 2004;47:105-113.
- 7. Jarlsäter S, Mattson E. Test of reliability of the Dizziness Handicap Inventory and the Activities-Specific Balance Confidence Scale for use in

Sweden. Adv Physiother. 2003;5:137-144.

- 8. Poon DM, Chow LC, Au DK, Hui Y, Leung MC. 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;113:1006-1011.
- 9. Perez N, Garmendia I, Garcia-Granero M, Martin E, Garcia-Tapia R. Factor analysis and correlation between Dizziness Handicap Inventory and Dizziness Characteristics and Impact on Quality of life scales. *Acta Otolaryngol Suppl.* 2001;545:145-154.
- Asmundson GJ, Stein MB, Ireland D. A factor analytic study of the dizziness handicap inventory: does it assess phobic avoidance in vestibular referrals? J Vestib Res. 1999;9:63-68.
- Jacobson GP, Calder JH. A screening version of the Dizziness Handicap Inventory (DHI-S). *Am J Otol.* 1998; 19:804-808.
- 12. 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;78:233-241.
- Jacobson GP, Newman CW, Hunter L, Balzer GK. Balance function test correlates of the Dizziness Handicap Inventory. J Am Acad Audiol. 1991; 2:253-260.
- Robertson DD, Ireland DJ. Dizziness Handicap Inventory correlates of computerized dynamic posturography. J Otolaryngol. 1995;24:118-124.
- Mann GC, Whitney SL, Redfern MS, Borello-France DF, Furman JM. Functional reach and single leg stance in patients with peripheral vestibular disorders. *J Vestib Res.* 1996;6:343-353.
- 16. Lin SI, Tsai TT, Lee IH, Wu YN. Perception of unsteadiness in patients with dizziness: association with handicap and imbalance. *J Biomed Sci.* 2002;9:428-435.
- 17. Enloe LJ, Shields RK. Evaluation of health-related quality of life in indi-

viduals with vestibular disease using disease-specific and general outcome measures. *Phys Ther.* 1997;77:890-903.

- Fielder H, Denholm SW, Lyons RA, Fielder CP. Measurement of health status in patients with vertigo. *Clin Otolaryngol Allied Sci.* 1996;21:124-126.
- Gill-Body KM, Beninato M, Krebs DE. Relationship among balance impairments, functional performance, and disability in people with peripheral vestibular hypofunction. *Phys Ther.* 2000;80:748-758.
- Murray K, Carroll S, Hill K. Relation between change in balance and selfreported handicap after vestibular rehabilitation therapy. *Physiother Res Int.* 2001;6:251-263.
- 21. Whitney SL, Wrisley DM, Brown KE, Furman JM. Is perception of handicap related to functional performance in persons with vestibular dysfunction? *Otol Neurotol.* 2004;25:139-143.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25:3186-3191.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33: 159-174.
- Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ; 1988.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-310.

Luc Vereeck

University College Antwerpen, Department of Health Care Division of Neuro- and Psychomotor Physical Therapy Van Aertselaerstraat 31 B-2170 Merksem, Belgium Tel.: +32 3 641 82 65 Fax: +32 3 641 82 70 E-mail: l.vereeck@ha.be