Validity and Reliability of the Ocular Motor Nerve Palsy Scale

Ling-Yun Zhou*, Chang Su, Tie-Juan Liu, Xue-Mei Li
Ocular Motility Disorder Treatment & Rehabilitation Center, Department of Acupuncture, the First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang Province, China

Funding: This study was supported by the National Natural Science Foundation of China, No. 81674052.

Abstract
Objective and accurate assessment of the degree of ocular motor nerve palsy is helpful not only in the evaluation of prognosis, but also for the screening of treatment methods. However, there is currently no comprehensive measure of its severity. In this study, we designed the Ocular Motor Nerve Palsy Scale and investigated its validity and reliability. Six experts were invited to grade and evaluate the scale. The study recruited 106 patients with a definite diagnosis of unilateral isolated ocular motor nerve palsy. Three physicians evaluated the patients using the scale. One of the three physicians evaluated the patients again after 24 hours. The content validity index (CVI) and factor analysis were used to analyze the scale’s construct validity. The intraclass correlation coefficient and Cronbach’s alpha were used to evaluate the inter-rater and test-retest reliability and the internal consistency. The CVI results (I-CVI = 1.0, S-CVI = 0.9, Pc = 0.016, K* = 1) indicated good content validity. Factor analysis extracted two common factors that accounted for 85.2% of the variance. Furthermore, the load value of each component was above 0.8, indicating good construct validity. The Ocular Motor Nerve Palsy Scale was found to be highly reliable, with an inter-rater reliability intraclass correlation coefficient of 0.965 (P < 0.01), a test-retest reliability intraclass correlation coefficient of 0.976 (P < 0.01), and Cronbach’s alpha values of 0.63–0.70. In conclusion, the Ocular Motor Nerve Palsy Scale with good validity and reliability can be used to quantify the severity of ocular motor nerve palsy.

This study was registered at Chinese Clinical Trial Registry (registration number: ChiCTR-OOC-17010702).

Key Words: nerve regeneration; Ocular Motor Nerve Palsy Scale; oculomotor nerve paralysis; abducens nerve paralysis; trochlear nerve paralysis; validity; reliability; internal consistency; diplopia; eyeball movement disorder; ptosis; neural regeneration

Introduction
The ocular motor nerves (cranial nerves III, IV, and VI) innervate different extraocular muscles. Damage to these nerves results in complex clinical symptoms and signs, and it is difficult to rate the severity of damage. At present, the severity of damage to these nerves is usually demonstrated by describing a single symptom. For example, the Hirischberg cover-uncover test with prim bars (Rucker et al., 2011) and the Maddox rod and red glass test are used to detect ocular deviation (Newman-Toker and Rizzo, 2010). The Hess screen test and Lancaster screen test are used to test diplopia (Christoff and Guyton, 2006; Roper-Hall, 2006). Scott and Kraft (1985) developed a grading method to describe eye movement limitations. However, there is no system or scale to synthetically evaluate the severity of ocular motor nerve palsy (OMNP). To address the urgent need for a useful clinical tool, this study designed a scale based on the symptoms and signs of OMNP. The Ocular Motor Nerve Palsy Scale (OMNPS) evaluates severity by assessing diplopia at 1 meter, eye movement, palpebral fissure, pupil size, and light reflex.
The reliability and validity of the scale were investigated to verify its practicability and feasibility.

**Participants and Methods**

**Participants**

One hundred six patients with unilateral OMNP were enrolled through advertisement and clinical practice at The First Affiliated Hospital of Harbin Medical University, Department of Ocular Motility Disorder Treatment & Rehabilitation Center, China, from March 2016 to February 2017.

**Inclusion criteria**

Patients presenting with all of the following criteria were considered for inclusion in the study: voluntary participation in the evaluation and diagnosis of unilateral isolated CN-III IV VI paralysis, a binocular visual acuity difference of less than 0.1, and able to complete a computerized diplopia test independently.

All patients gave signed informed consent, and the study was approved by the Ethics Committee of The First Affiliated Hospital of Harbin Medical University of China (approval number: 201452).

**Exclusion criteria**

Patients with one or more of the following conditions were excluded from the study: color blindness, color weakness, abnormal retinal correspondence, or any other severe condition.

**Scale design**

**Item selection**

The five main items, diplopia at 1 meter, eye movement condition, palpebral fissure, pupil size, and light reflex, were assessed according to the clinical symptoms and signs of OMNP (Danchaivijitr, 2004; Brazis, 2009; Prasad and Volpe, 2010; Galtrey et al., 2015). The scale is displayed in Table 1.

**Testing methods (\(\text{\textit{\textdollar}}\))**

Diplopia at 1 meter (Zhou et al., 2016, 2017): After head fixation, patients completed a computerized diplopia test at 1 meter wearing red and blue glasses. The direction and maximum angle of deviation of diplopia were measured and recorded automatically by the computer.

Eye movement (Corbett, 2003): After head fixation, patients were asked to move the eyeball toward the left and right, and straight up and down. The functional eye movement in four directions was observed and graded according to the scoring criteria.

The palpebral fissure (Greenberg et al., 2012): The frontal muscle above the eyebrow was fixed and did not affect the activity of the upper eyelid. Patients were asked to open their eyelids as wide as possible. The distance between the midpoint of the upper and lower eyelids (accurate to the millimeter) was measured.

Pupil size (Jacobson, 1998): Pupil diameter was measured using a ruler under stationary indoor lights and all testing rooms had the same level of light.

Light reflex (Cornblath, 2014): The physician directly tested the patient’s light reflex with a flash light.

**OMNPS scoring criteria**

The scale is divided into three sub-scales that reflect the type of cranial nerve damage. PART CN-IV is designed for trochlear nerve paralysis patients, and includes diplopia at 1 meter only. PART CN-VI is designed for abducens nerve paralysis, and includes diplopia at 1 meter and eye movement. PART CN-III is designed for oculomotor nerve paralysis patients and includes all five items. (1) Diplopia at 1 meter: the maximum angle of diplopia is graded on five levels, with each level representing 5°. If the maximum deviation angle is horizontal, it only needs grading; if it is vertical, the angle is multiplied by 3 before grading. (2) Eye movement: This test consists of two parts. Horizontal movement is assessed on five levels with reference to the midline and two quarter lines. Vertical movement is assessed on four levels with reference to the line of the inner and temporal canthus. (3–4) Palpebral fissure and pupil size: affected eyes are compared with healthy eyes and there are four levels for palpebral fissure and four levels for pupil size. (5) Light reflex: The direct light reflex and grading are observed. Higher scores represent more severe symptoms. The details of the scoring criteria are listed in Table 1.

**OMNPS grading by six experts:**

Our scale was sent to six experts by email. The experts were invited to grade the scale in relation to OMNP. The scale is shown in Table 2. The experts were asked to choose 1 (not relevant) to 4 (highly relevant) according to the item’s relevance to OMNP. Content validity indices were calculated from the experts’ scores, including item-level content validity index (I-CVI = number of experts who gave a grade of 3 or 4/number of experts), scale-level content validity index/universal agreement (S-CVI/UA = number of items graded as 3 or 4/number of items), and S-CVI/Ave = average I-CVI scores of five items.

**Clinical examination procedure:**

**Physicians**

The physicians familiarized themselves with the scale before administering the tests. In a pilot study, the physicians administered the test to five patients (not including the 106 patients recruited for this study) until they could run the test smoothly and accurately. The physicians were blind to the patients’ medical histories. The patients received the examination in random order.

**Testing process**

Three physicians tested the 106 consecutive patients. The computerized diplopia test was performed under the physicians’ instruction, and data were automatically generated by computer and then graded by the physicians. To prevent bias, we applied the following procedures. The testing procedure was performed in three independent clinical consulting rooms. Communication between physicians was forbidden.
during the process, and the results were compared by physicians who were not involved with the research. One of the three physicians measured the patients again after 24 hours without any intervention. The data were independently cross-checked by the other two researchers.

Statistical analysis
All normally distributed data are expressed as the mean ± SD. The data were analyzed using Statistical 19.0 software (IBM SPSS Statistics, IBM Co., Somers, NY, USA). The CVI was used to assess content validity. Factor analysis was utilized to evaluate the construct validity of PART CN-III. The intraclass correlation coefficient (ICC) with a 95% confidence interval was calculated to assess the inter-rater reliability and test-retest reliability. After comparison of the total score, the scores for each item were compared except for diplopia at 1 meter. Cronbach’s Alpha was used to evaluate internal consistency.

Results
General information
The 106 patients with OMNP consisted of 67 males and 39 females ranging in age from 11–79 (54.72 ± 14.42) years. Table 3 presents patients’ scores grouped according to the affected cranial nerve.

Validity analysis
Content validity
As displayed in Table 4, the I-CVI was 0.83–1, demonstrating that the five items were strongly correlated with OMNP (Pc = 0.016, K* = 0.82–1) (Polit et al., 2007). The S-CVI/AU was 0.8 (> 0.8 indicates good content validity; Davis, 1992). S-CVI/Ave was 0.966 (> 0.9 indicates good content validity; Waltz et al., 2005). The results indicate that the content validity of the OMNPS is good.

Construct validity
First, the Kaiser-Meyer-Olkin test and Bartlett’s spherical test were performed. The Kaiser-Meyer-Olkin value was 0.792 and Bartlett’s spherical test was significant at P < 0.01, confirming that the data were suitable for factor analysis. Principal component analysis was performed. Two common factors were extracted and the contribution rate was 85.2%. After matrix rotation, eye movement condition, palpebral fissure, pupil size, and light reflex loaded on factor 1 with values of 0.867, 0.887, 0.927, and 0.899, respectively. Maximum angle of diplopia at 1 meter loaded on factor 2 with a value of 0.990. The variance, contribution rate, and cumulative contribution rate are listed in Table 5. Two common factors were extracted, demonstrating that all five items reflected symptoms and signs of OMNP. This was in agreement with the theoretical design of the scale. All of the items

Table 2 Content validity grading form

Do you think the weather scales have correlation with ocular motor nerve palsy, and how much is it?

Max angle of diplopia
Eye movement condition
Palpebral fissure
Pupil size
Light reflex

1= not relevant 2= somewhat relevant 3= quite relevant 4= highly relevant

Table 4 Results of content validity

<table>
<thead>
<tr>
<th>Item</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>Number of experts who scored 3 or 4</th>
<th>I-CVI</th>
<th>Pc</th>
<th>K*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eyeball movement</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>1.0</td>
<td>0.016</td>
<td>1.0</td>
</tr>
<tr>
<td>Diplopia in 1 m</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>1.0</td>
<td>0.016</td>
<td>1.0</td>
</tr>
<tr>
<td>Palpebral fissure</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1.0</td>
<td>0.016</td>
<td>1.0</td>
</tr>
<tr>
<td>Pupil size</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>1.0</td>
<td>0.016</td>
<td>1.0</td>
</tr>
<tr>
<td>Light reflex</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>0.83</td>
<td>0.094</td>
<td>0.81</td>
<td></td>
</tr>
</tbody>
</table>

1 means not relevant, 2 means somewhat relevant, 3 means quite relevant, and 4 means highly relevant. The content validity index includes I-CVI, S-CVI/UA, and S-CVI/Ave. The item-level content validity index (I-CVI, 0.83–1) is the number of experts who gave a grade of 3 or 4/total number of experts. The scale-level content validity index/universal agreement (S-CVI/UA) = 0.8 (number of items with experts grading 3 or 4, n = 4/number of items, n = 5) S-CVI/Ave = 0.966 (average I-CVI for the 5 items). Pc (probability of a chance occurrence) was computed using the formula for a binomial random variable, with one specific outcome: $P_c = \left[\frac{N!}{A!(N-A)!}\right]0.5^N$, where N represents the number of experts and A represents the number who agreed that the item has high relevance; $K^*$ = kappa, indicating agreement on relevance: $K^* = (I-CVI - pc)/(1-pc)$.

Table 5 Results of construct validity (variance, cumulative contribution rate and load values of main components)

<table>
<thead>
<tr>
<th>Common factor</th>
<th>Cumulative contribution rate (%)</th>
<th>Load values on Mean±SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diplopia in 1 m</td>
<td>1.013</td>
<td>85.221</td>
</tr>
<tr>
<td>Eyeball movement</td>
<td>3.248</td>
<td>64.953</td>
</tr>
<tr>
<td>Palpebral fissure</td>
<td>0.332</td>
<td>91.866</td>
</tr>
<tr>
<td>Pupil size</td>
<td>0.292</td>
<td>97.700</td>
</tr>
<tr>
<td>Light reflex</td>
<td>0.115</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3 Ocular motor nerve palsy score in 106 patients

<table>
<thead>
<tr>
<th>Affected Cranial nerve</th>
<th>n</th>
<th>Highest score</th>
<th>Lowest score</th>
<th>Mean±SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>III</td>
<td>41</td>
<td>22</td>
<td>3</td>
<td>9.29±4.03</td>
</tr>
<tr>
<td>IV</td>
<td>30</td>
<td>4</td>
<td>1</td>
<td>2.83±1.34</td>
</tr>
<tr>
<td>VI</td>
<td>35</td>
<td>8</td>
<td>1</td>
<td>5.11±2.27</td>
</tr>
</tbody>
</table>

III, IV and VI represent the affected cranial nerve.

Table 6 Intraclass correlation coefficient (ICC) results of each item

<table>
<thead>
<tr>
<th></th>
<th>Eyeball movement</th>
<th>Palpebral fissure</th>
<th>Pupil size</th>
<th>Light reflex</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-retest</td>
<td>0.962</td>
<td>0.913</td>
<td>0.930</td>
<td>0.968</td>
<td>0.976</td>
</tr>
<tr>
<td>Inter-rater</td>
<td>0.928</td>
<td>0.941</td>
<td>0.844</td>
<td>0.941</td>
<td>0.965</td>
</tr>
</tbody>
</table>

Test-retest represents the ICC of test-retest reliability. Inter-rater represents the ICC of inter-rater reliability. ICC>=0.75 indicates good reliability. All coefficients are significant ($P < 0.01$).

Reliability analysis

The results of the reliability analysis are shown in Table 6. In terms of the test-retest reliability, the ICC of each individual item was 0.913–0.968 ($P < 0.01$); and the ICC of the total score was 0.976 ($P < 0.01$). The ICC for inter-rater reliability was 0.844–0.941 ($P < 0.01$), and the total score ICC was 0.965 ($P < 0.01$). A 95% confidence interval was used for all ICs. ICC values less than 0.4 indicate poor reliability, and values greater than 0.75 indicate good reliability (Pan and Ni, 1999).

Internal consistency measured by Cronbach’s alpha was 0.636–0.701 ($P < 0.01$), indicating high reliability.

Discussion

Rating scales are commonly used to evaluate the severity of disease and efficacy of treatment, especially in neurological diseases and dyskinesia (Hobart et al., 2001; Middleton, 2012; Giang et al., 2013). OMNP is a common type of dyskinesia that affects patients’ quality of life (Kuo et al., 2010). However, due to the complex manifestations of OMNP, it is difficult to assess its severity. To date, no study or scale has examined this issue and no system or scale is available to evaluate the severity of OMNP in neurology and ophthalmology, or to document the progression and degree of recovery. Therefore, to address this clinical need, this study designed a measurement scale and evaluated its reliability and validity.
Content validity
In the present study, I-CVI and S-CVI were used to assess the content validity. After correcting the random consistency (Pc = 0.016, K* = 1), the results still indicate good content validity (I-CVI = 0.83–1 S-CVI/UA = 0.83 S-CVI/ Ave = 0.966) according to previous studies (Davis, 1992; Waltz et al., 2005). Good content validity not only demonstrates the strong association between scale scores and OMNP, but also indicates that the score reflects the severity of OMNP in some respects. All five items measure the symptoms and signs of OMNP and are strongly associated with its pathophysiology (Brazis, 2009; Cornblath, 2014). The testing method generally follows the accepted method of neuro-ophthalmology examination (Danchaiyijitr, 2004; Brazis, 2009; Prasad and Volpe, 2010; Galtrey et al., 2015). We made some corrections and modifications to the scale to make it suitable for daily practice. For the eye movement measure, we graded horizontal movement on five levels and vertical movement on four levels according to the horizontal and vertical palpebral fissure size in healthy subjects. The reference values of the computerized diplopia test were horizontal 2.55° and vertical 0.76° (Zhou et al., 2016, 2017). The horizontal value is about three times the vertical value, which is why we multiplied the vertical measure by three before grading.

Construct validity
Factor analysis is generally used to assess construct validity. In this study, two common factors were extracted and their total contribution was 85.2%, which is significant factor analysis (Nia et al., 2017; Oh and Kim, 2017). After matrix rotation, eye movement, palpebral fissure, pupil size, and light reflex loaded on factor 1 and maximum angle of diplopia at 1 meter loaded on factor 2. We concluded that factor one measures the signs of OMNP and factor 2 reflects the symptoms of OMNP. The extraction of these two factors demonstrates that all five items reflect the symptoms and signs of OMNP. All items had loadings above 0.8 on their own factor, indicating agreement with the intended theoretical design of the scoring and confirming that the scale has good construct validity.

Reliability
The ICCs of inter-rater reliability and test-retest reliability were all above 0.75 (P < 0.01), indicating very good reliability (Pan and Ni, 1999). The inter-rater and test-retest reliability results are similar to those of other scale development studies (Ardali et al., 2017; Cavalini et al., 2017; De Roo et al., 2017). Cronbach’s alpha values were in the acceptable range, from 0.636–0.701 (P < 0.01). Overall, the OMNPS shows good reliability. We excluded patients’ scores for the angle of diplopia from the reliability and test-retest analysis; it would be meaningless for the patients to repeat the computerized diplopia test three times a day, because the data were automatically generated by computer and the error was very small. The OMNPS includes three sub-scales for different types of OMNP (Newman and Bruce, 2007; Cornblath, 2014). For instance, abducens nerve paralysis only involves eye movement disorder and diplopia, and the other three items are irrelevant for these patients. Hence, we developed three sub-scales for different situations. PART CN-IV is scored from 0–4, PART CN-VI 0–8, and PART CN-III 0–27; higher scores indicate greater severity.

Bias
To minimize bias, we used various control processes (Kottner et al., 2011). The test order was randomized, and the tests were conducted in the consulting room and resembled the daily situation as far as possible. Communication between physicians was forbidden during the process (Stockkendahl et al., 2006). To avoid the Hawthorne effect, physicians who were not involved in the research compared the results (Wickstrom and Bendix, 2000). The single blind method was used for the physicians (Kottner et al., 2009). The data were cross-checked and consolidated by two other researchers.

Limitations
We did not report criterion validity, because there was no suitable standard method for comparison. The scale’s application is limited to patients with unilateral mononeuropathy, because three items compare the affected eye with the healthy eye. The scale requires further modification to enable a wide range of application in the clinic. Rotation diplopia could not be quantified because the computerized diplopia test was based on the Hess principle. The Hess test does not assess rotation diplopia, but we believe that computerized diplopia testing will undergo further development. The sensitivity and responsiveness of the scale need to be investigated in the future.

Conclusion
The OMNPS shows high content validity, construct validity, inter-rater reliability, test-retest reliability, and internal consistency. It is therefore suitable for use as a clinical scale to evaluate the severity of OMNP.


References


(Copiedited by Wang J, Li CH, Qiu Y, Song LP, Zhao M)