Brief Report

Inter- and Intra-Rater Reliability of the Visual Vertical in Subacute Stroke

Celine Piscicelli, MS; Julien Barra, PhD; Patrice Davoine, MD; Anne Chrispin, MD; Sylvie Nadeau, PhD; Dominic Pérennou, MD, PhD

Background and Purpose—Visual vertical (VV) has been increasingly used as a routine clinical assessment to identify alteration of verticality perception as a possible cause of postural disorders after stroke. This study aims to determine whether the reliability of VV is sufficient to support a wide clinical use in neurorehabilitation for monitoring of patients with stroke.

Methods—Twenty patients with subacute stroke in neurorehabilitation unit were tested after a first and unique hemispheric stroke. To evaluate the inter-rater reliability, VV was assessed the same day by 2 examiners whose degrees of expertise differed. The second examiner repeated the test the next day to investigate intrarater reliability. VV orientation (mean, primary criterion) and uncertainty (SD, secondary criterion) were calculated for 10 trials. Their reliability was quantified by the intraclass correlation coefficient, Bland–Altman plots, and the minimal detectable change. The concordance between 2 examiners was quantified by Cohen’s κ coefficients (κ).

Results—About VV orientation, inter- and intrarater reliability were excellent (intraclass correlation coefficient, 0.979 and 0.982). The Bland–Altman plots and the minimal detectable change revealed a difference inferior to 2° between 2 tests. The concordance between 2 assessments for the diagnosis of abnormal VV orientation was absolute for the same examiner (κ=1; P<0.05) and excellent between 2 examiners (κ=0.92; P<0.05). As for VV uncertainty the intrarater reliability was satisfactory (intraclass correlation coefficient, 0.836) but the inter-rater reliability was poor (intraclass correlation coefficient, 0.211).

Conclusions—The orientation of the VV is a highly reliable criterion, which may be used both in research and in routine clinical practice. (Stroke. 2015;46:00-00. DOI: 10.1161/STROKEAHA.115.009610.)

Key Words: posture ■ rehabilitation ■ reliability ■ stroke

In stroke rehabilitation, verticality perception has been progressively integrated into routine assessments as a basic spatial cognition measurement to complement other tests. Assessing verticality perception is considered clinically relevant to guide poststroke rehabilitation and monitor postural recovery because it informs the causes of the body’s misorientation with respect to vertical (lateropulsion). The visual vertical (VV) is the most commonly used test to assess verticality perception both in research and in clinical practice. It is a simple test that consists of adjusting a luminous rod to the vertical in darkness. VV perception is estimated by 2 measures: VV orientation, which reflects the direction perceived as vertical by individuals, and VV uncertainty, which reflects the robustness of the internal model of verticality.

Although VV has been progressively generalized for post-stroke assessment, with a bewildering array of paradigms, its psychometric properties remain to be established. Thus, the aim of this study was to determine whether VV measurements have enough reliability to support the use both in research and in routine clinical practice. We hypothesized that using a previously validated methodological approach for VV assessment, VV orientation and VV uncertainty would be reliable measures with higher intrarater than inter-rater reliability.

Materials and Methods

Patients
We assessed 20 consecutive patients with subacute stroke (64±15 years; 8 women, 12 men; 8 left- and 12 right-sided lesions) admitted in a neurorehabilitation unit for a single recent (5.3±2.6 weeks) hemorrhagic (7/20) or ischemic (13/20) hemispheric stroke. Demographic and clinical data gathered dealt with stroke features, functional outcome, postural behavior, spatial neglect, aphasia, and...
depression. Individual and mean data are shown in Supplement Data I in the online-only Data Supplement (Figure I in the online-only Data Supplement and Table I in the online-only Data Supplement). Patients were excluded if they had brain stem or cerebellar strokes involving vestibulo-ocular symptoms because in those patients a different mechanism is involved in abnormal VV perception. We also excluded patients with neuropathy, psychiatric disorders, or major comprehension problems because of aphasia or dementia and those with unstable medical problems. Participants gave their informed consent according to the guidelines of the institutions’ ethics committees.

Setup for VV Measurement

VV perception was assessed in darkness by 10 adjustments of the direction of a bright line (15 cm long, 2 mm wide) presented on a computer screen, at eye level, using a previously validated methodological approach for VV assessment. Patients were seated with their head and trunk maintained upright and were asked to verbally adjust the line to the vertical. There was no time limit or feedback from the examiner. The measures started after 2 minutes of darkness. For each patient and each session, we calculated 2 criteria for the 10 trials: the mean VV orientation as a primary criterion and the mean VV uncertainty (within-subject SD) as a secondary criterion. According to standards, after sign transformation according to the lesion side, a negative value corresponded to a contralesional tilt. Patients were categorized as having normal VV perception (from −2.5° to 2.5°), contralesional (−2.5° to 1.9°), or ipsilesional (>2.5°) bias according to normal ranges validated in a previous study.

Procedure

VV was assessed 3 times. In the first session, 2 examiners (A and B), whose degrees of expertise differed (a novice versus an expert in VV assessment), performed independent successive measurements (A1 and B1) with a 5-minute resting period in between. During the resting period, patients kept their eyes open in a lit environment and could adopt a spontaneous sitting position. Each examiner had to install the patients before each measurement recording. The order of assessment by examiners A and B was counterbalanced between patients. The examiner was blinded to the patients’ VV estimates measured by the other examiner. In the second session, the next day at the same time, examiner B repeated the entire procedure (B2 measurement).

Statistical Analysis

The distribution of VV orientation and VV uncertainty tested using Kolmogorov–Smirnov test was normal allowing the use of parametric statistics.

Inter-rater reliability (examiners A1 versus B1) and intrarater reliability (same examiner B1 versus B2) were determined by calculating the intraclass correlation coefficient (ICC). Absolute reliability was assessed with the Bland–Altman method with 95% confidence intervals, and with the SE of measurement (SEM=SD/√n−1). SEM was then used to estimate the minimal detectable change (MDCse=1. 96xSEM√2) for 95% degree of confidence on VV, the amount by which a patient’s score needs to change to ensure that the change is greater than measurement error (see details in Supplementary Data II).

In addition, we used Cohen’s κ coefficient (κ) to calculate the percentage agreement between the 2 examiners for the classification of VV perception for patients (contralesional bias, ipsilesional bias, no bias) in regards to normative data. A κ or ICC was considered in term of agreement: under 0.6, questionable; 0.6 to 0.79, moderate; and >0.8, high.

Statistical tests were performed with SPSS 21 (IBM Corporation, Armonk, NY). The α risk was set at P≤0.05. Unless other indication, data are presented in the form mean±SD.

Results

The average VV orientation was −3° and the average VV uncertainty was 1.6° (first assessment regardless of examiner). VV perception was normal for 8 patients (40%) and was abnormally tilted to the contralesional side for 9 (45%; −7.2±2.8°) and toward the ipsilesional side for 3 (15%; 2.8±0.1°). Individual data for each examiner and each measurement are shown in the Table (Supplementary Data I in the online-only Data Supplement).

VV Orientation

Inter-rater Reliability

The inter-rater reliability was excellent, with an ICC of 0.979 (Table; Figure 1A). The Bland–Altman plots (Figure 1B) showed that all of the differences but one were within the 95% limits of agreement (−2.3° to 2.6°). The mean difference between A1 and B1 measurements was small (0.1±1.2°), meaning that differences between the 2 examiners were acceptable for VV orientation. The MDCse indicated that a change of ≥1.85° was required to be 95% certain that the change was not because of measurement error.

The agreement was excellent between the 2 examiners in discriminating normal from abnormal VV perception (κ=0.89; P<0.05) and more precisely in discriminating contralesional, ipsilesional, and no VV bias (κ=0.92; P<0.05).

Intrarater Reliability

The intrarater reliability was excellent, with an ICC of 0.982 (Figure 1C). The Bland–Altman plots (Figure 1D) showed that all the differences were within the 95% limits of agreement (−2° to 1.9°) with a low mean difference between B1 and B2 measurements (−0.05±1°), meaning that when retested, the VV orientation measure was similar. The MDC95 was 1.74°.

The agreement between the 2 measurements from a same examiner (B1 and B2) was absolute in discriminating normal from abnormal VV perception (κ=1; P<0.05) and more precisely in discriminating contralesional, ipsilesional, and no VV bias (κ=1; P<0.05).

VV Uncertainty

Inter-rater Reliability

Although the inter-rater reliability (Table; Figure 2A) was poor as determined by an ICC of 0.21, the Bland–Altman

| Table. Inter- and Intrarater Reliability Values for VV Orientation and VV Uncertainty |
|-----------------------------------|---------|--------|--------|--------|
| W orientation                      | ICC     | SEM    | MDC 95%| MDC 90%|
| Inter-rater reliability (A1 vs B1) | 0.979   | 0.67a  | 1.85a  | 1.56a  |
| Intrarater reliability (B1 vs B2)  | 0.982   | 0.63a  | 1.74a  | 1.46a  |
| VV uncertainty                     |         |        |        |        |
| Inter-rater reliability (A1 vs B1) | 0.211   | 0.94a  | 2.62a  | 2.2a   |
| Intrarater reliability (B1 vs B2)  | 0.836   | 0.55a  | 1.54a  | 1.3a   |

ICC indicates intraclass correlation coefficient; MDC, minimal detectable change; SEM, standard error of measurement; and W, visual vertical.
plots (Figure 2B) showed that all of the differences but one were within the 95% limits of agreement (−2.6° to 1.6°). The mean difference between A1 and B1 uncertainty measures was small (−0.5±1.08°) meaning that differences between 2 examiners were clinically acceptable for VV uncertainty. The MDC was 2.62°.

Intrarater Reliability
Intrarater reliability (Figure 2C) was excellent (ICC=0.836). The Bland–Altman plots (Figure 2D) showed that most differences were within the 95% limits of agreement (−1.03° to 1.33°) with a small mean difference between B1 and B2 uncertainty measures (0.15±0.6°). This meant that when retested, the VV uncertainty was similar. The MDC was 1.54°.

Discussion
The clinical characteristics of our patients and the frequency and magnitude of VV perception were representative of those previously reported in the literature for a large cohort of patients with stroke.5,6,10

Our study showed that VV orientation is a reliable measure for patients with subacute stroke and has excellent inter- and intrarater reliability (ICCs >0.98) and acceptable measurement error (<0.7°). We found an MDC <2°, which means that a change ≥2° can be interpreted with 95% confidence as a real change. This threshold is critical for clinical rehabilitation settings in the follow-up of patients with stroke with a tilted perception of verticality but also for clinical stroke trials to analyze the responsiveness to a specific rehabilitation program.

For VV uncertainty, our results revealed a poor relative reliability about ICC values, especially for inter-rater reliability. Yet, absolute reliability (Bland–Altman analysis13 and SEM14) remained clinically satisfactory. In the present study, the 2 examiners had different degrees of expertise, which might explain their different VV uncertainties and affect the reliability of this measure. This suggests that clear guidelines for patients’ installation, instructions, and handling, as well as training would benefit examiners. Lastly, the lowest ICCs for the VV uncertainty is also affected by the lower intersubject variance (to 0.6° from 3.9°) as compared with VV orientation (to −12.8° from 4.2°). ICC values are the ratio of the variance between subjects and the total variance and thus coefficients are systematically lower in homogeneous samples.14 This also explained why the outsider had more influence on VV uncertainty than VV orientation.
Limitations

These findings are valid for the conditions under which VV was assessed in our study, that is, patients with subacute hemispheric stroke seated with trunk and head maintained upright. Moreover, in our study, patients verbally adjusted the visual line to the vertical. In some studies, patients estimated VV by manipulating a remote control. Similar to others we preferred using the verbal modality to circumvent possible biases because of handedness, apraxia, learning, or executive dysfunctions. The verbal modality does not preclude testing aphasic patients, because a code answer can easily be set up for these patients.

Conclusions

Our study reveals that inter- and intrarater reliability of the VV orientation is excellent as a measure of verticality perception in patients with subacute hemispheric stroke.

Acknowledgments

We thank the neuropsychologist student and clinicians of the neurorehabilitation unit of the University Hospital of Grenoble.

Sources of Funding

This work has been supported by a grant from La Société Francophone Posture Equilibre Locomotion.

Disclosures

None.

References


Inter- and Intra-Rater Reliability of the Visual Vertical in Subacute Stroke
Celine Piscicelli, Julien Barra, Patrice Davoine, Anne Chrispin, Sylvie Nadeau and Dominic Pérennou

Stroke, published online May 28, 2015;
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/early/2015/05/28/STROKEAHA.115.009610

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2015/06/22/STROKEAHA.115.009610.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/
Supplemental Material

Supplementary Data I

Demographic and clinical data gathered dealt with stroke features, functional outcome, postural behaviour, spatial neglect, aphasia and depression (Figure and Table). The mean score of disability for the 20 stroke patients was $3.5 \pm 1.05$ as assessed with the modified Rankin Scale, ranging from 0 (no symptoms) to 5 (severe disability)(1). Postural behavior was assessed with the Postural Assessment Scale for Stroke Patients (PASS),(2) ranging from 0 (no postural control) to 36 (good postural control); the mean score was 30±6.5. Gait, assessed with the Lindmark scale,(3) ranging from 0 (no gait) to 6 (independent gait), categorizing the patients as able to walk with help (mean score: 3.25±2.5). Spatial neglect was diagnosed when scores for at least two tests were abnormal among the three following tests(4): the bells cancellation test, the line bisection test and the Catherine Bergego behavioral Scale. According to this criterion, 7/20 patients (35%) had clinical signs of spatial neglect. Depression was assessed with the Aphasic Depression Rating Scale for stroke patients in subacute stage,(5) ranging from 0 (no depression) to 32 (severe depression); the mean score for the 20 stroke patients was 9.8±4.8. All measures, including the VV, were performed under conditions of the clinical daily routine.

For each patient, the rehabilitation therapy lasted about 3-4 hours a day, with exercises graduated according to their impairments and recovery and mainly focused on language, spatial neglect, cognitive impairment, swallowing, balance, gait and prehension. No patient had undergone any type of sensory manipulation or stimulation and their sense of verticality was never assessed prior to the study.
Supplemental Figure I. Overlapping lesion plots for the 12 right-hemisphere and 8 left-hemisphere stroke patients. The number of overlapping lesions is illustrated by different color-codes with increasing frequencies ranging from violet (n = 1) to yellow (n = max).
### Supplemental Table 1: Individual clinical data

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age (yrs)</th>
<th>Gender</th>
<th>Lesion side</th>
<th>I/H</th>
<th>Delay (wks)</th>
<th>Rankin 0-5</th>
<th>PASS 0-36</th>
<th>GaitLH 0-6</th>
<th>Neglect</th>
<th>Aphasia</th>
<th>ADRS 0-32</th>
<th>VV Orientation</th>
<th>VV Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>44</td>
<td>M</td>
<td>Right</td>
<td>H</td>
<td>8</td>
<td>20</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>-</td>
<td>-9.7° -9.8°</td>
<td>2.8° +2.6°</td>
</tr>
<tr>
<td>19</td>
<td>83</td>
<td>F</td>
<td>Right</td>
<td>H</td>
<td>4</td>
<td>5</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>-1.0° -0.8°</td>
<td>-1.0° -0.7°</td>
<td>2.3° +3.6°</td>
</tr>
<tr>
<td>8</td>
<td>74</td>
<td>M</td>
<td>Right</td>
<td>I</td>
<td>4</td>
<td>2</td>
<td>36</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>-0.4° 0.04°</td>
<td>0.9°</td>
<td>1.8° +1.1°</td>
</tr>
<tr>
<td>17</td>
<td>35</td>
<td>M</td>
<td>Right</td>
<td>I</td>
<td>8</td>
<td>2</td>
<td>36</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>-5.6° -6.1°</td>
<td>-5.2° 0.6°</td>
</tr>
<tr>
<td>10</td>
<td>72</td>
<td>M</td>
<td>Right</td>
<td>I</td>
<td>4</td>
<td>3</td>
<td>32</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>-5.0° -5.0°</td>
<td>-6.8° 1.1°</td>
</tr>
<tr>
<td>18</td>
<td>64</td>
<td>F</td>
<td>Right</td>
<td>I</td>
<td>4</td>
<td>3</td>
<td>32</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td>-</td>
<td>-12.8° -</td>
</tr>
<tr>
<td>1</td>
<td>67</td>
<td>M</td>
<td>Right</td>
<td>I</td>
<td>4</td>
<td>4</td>
<td>32</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>-4.0°</td>
<td>-8.8° -6.8°</td>
</tr>
<tr>
<td>3</td>
<td>52</td>
<td>M</td>
<td>Right</td>
<td>I</td>
<td>8</td>
<td>4</td>
<td>29</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>18</td>
<td>-</td>
<td>-8.9°</td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td>F</td>
<td>Right</td>
<td>I</td>
<td>4.5</td>
<td>4</td>
<td>23</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>-6.4° -6.2°</td>
<td>-5.1° 1.2°</td>
</tr>
<tr>
<td>9</td>
<td>75</td>
<td>F</td>
<td>Right</td>
<td>I</td>
<td>12</td>
<td>4</td>
<td>21</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>3.0° 4.2°</td>
<td>3.2° 1.8°</td>
</tr>
<tr>
<td>15</td>
<td>82</td>
<td>F</td>
<td>Right</td>
<td>I</td>
<td>4</td>
<td>3</td>
<td>36</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0.2° -1.0°</td>
<td>-0.1° 2.2°</td>
</tr>
<tr>
<td>20</td>
<td>42</td>
<td>F</td>
<td>Right</td>
<td>I</td>
<td>12</td>
<td>5</td>
<td>29</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>-7.2° -7.7°</td>
<td>-7.2° 2.0°</td>
</tr>
<tr>
<td>6</td>
<td>63</td>
<td>M</td>
<td>Left</td>
<td>H</td>
<td>4</td>
<td>2</td>
<td>36</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>1.5° 1.9°</td>
<td>0.4° 1.5°</td>
</tr>
<tr>
<td>4</td>
<td>84</td>
<td>M</td>
<td>Left</td>
<td>H</td>
<td>4</td>
<td>3</td>
<td>35</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>-2.3° -2.3°</td>
<td>-2.4° 2.2°</td>
</tr>
<tr>
<td>12</td>
<td>69</td>
<td>M</td>
<td>Left</td>
<td>H</td>
<td>4</td>
<td>3</td>
<td>29</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>-0.8° -1.0°</td>
<td>0.2° 0.7°</td>
</tr>
<tr>
<td>13</td>
<td>73</td>
<td>M</td>
<td>Left</td>
<td>H</td>
<td>4</td>
<td>3</td>
<td>36</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.8° 2.0°</td>
<td>1.8° 1.3°</td>
</tr>
<tr>
<td>2</td>
<td>56</td>
<td>F</td>
<td>Left</td>
<td>H</td>
<td>4</td>
<td>4</td>
<td>30</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>2.6° 3.3°</td>
<td>3.0° 1.4°</td>
</tr>
<tr>
<td>16</td>
<td>61</td>
<td>M</td>
<td>Left</td>
<td>I</td>
<td>3</td>
<td>2</td>
<td>36</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>-0.8° 0.1°</td>
<td>-0.6° 1.4°</td>
</tr>
<tr>
<td>11</td>
<td>56</td>
<td>M</td>
<td>Left</td>
<td>I</td>
<td>4</td>
<td>3</td>
<td>36</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>-0.1° -0.2°</td>
<td>0.8° 1.0°</td>
</tr>
<tr>
<td>5</td>
<td>87</td>
<td>F</td>
<td>Left</td>
<td>I</td>
<td>4</td>
<td>5</td>
<td>14</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>-3.8° -3.8°</td>
<td>-4.0° 2.0°</td>
</tr>
</tbody>
</table>

Yrs= Years; M=Male; F=Female; I=Ischemia; H=Haemorragia; wks=Weeks; PASS=Postural Assessment Scale for Stroke; GaitLH= Lindmark Scale; ADRS=Aphasic Depression Rating Scale; A= rater A; B1= rater B session 1; B2=rater B session 2. Neglect and Aphasia were reported as 0= absence and 1 = presence. Patients were ordered according to first the side of the lesions; secondly the stroke etiology and finally the Rankin score. The patients’ number corresponded to the chronological order of inclusion.
Supplementary Data II: Analysis

**Bland-Altman plots** are constructed by plotting the between-session difference versus the mean value of the 2 sessions for each variable. These plots, along with the 95% confidence intervals calculated for the between-session differences, are then used to visualize systematic variations across the zero line\(^6\).

**Standard error of measurement** (SEM), assessing how precisely a test measures a subject’s true score, is obtained by:

\[ \text{SEM} = \text{SD}_X \times \sqrt{1 - \text{ICC}} \] \(^7\)

**Minimal detectable change (MDC)**, a threshold to determine true differences between successive measurements,\(^8\) is calculated for two degrees of confidence:

\[ \text{MDC}_{95} = 1.96 \times \text{SEM} \times \sqrt{2} \]
\[ \text{MDC}_{90} = 1.65 \times \text{SEM} \times \sqrt{2} \]
References:


