Validity and Interobserver Reliability of Visual Observation to Assess Partial Weight-Bearing

Henri L. Hurkmans, PhD, Johannes B. Bussmann, PhD, Eric Benda, PT


Objective: To determine the validity and interobserver reliability of visual observation to assess partial weight-bearing.

Design: Validation and interobserver reliability study.

Setting: University medical center.

Participants: Patients (N=10) with a total hip arthroplasty operated 1 to 12 months prior to the study referred by 10 physical therapists (5 experienced and 5 inexperienced in training patients in partial weight-bearing).

Interventions: Not applicable.

Main Outcome Measures: The amount of weight-bearing assessed by visual estimation (visual analog scale score) in percentage body weight (BW). Actual weight-bearing (percentage BW) as measured with the Pedar Mobile system. The mean difference (systematic error) between visual estimation and the Pedar system and the SD of the differences (random error) were determined by the limits of agreement (LOA) method with multiple observations per subject. The intraclass correlation coefficient (ICC) was calculated as a measure for the interobserver reliability.

Results: The mean difference ± SD between visual observation and the reference method was −9.5 ± 20.1 percentage BW (95% confidence interval, −24.0 to 5.0 percentage BW) with LOA ranging from −49.8 to 30.8 percentage BW. The ICC was .57. The therapists’ experience in partial weight-bearing training had no effect on the mean difference (P=.349) between the 2 methods.

Conclusions: Visual observation is not a valid and reliable method to assess partial weight-bearing.

Key Words: Rehabilitation; Reproducibility of results; Weight-bearing.
© 2009 by the American Congress of Rehabilitation Medicine

Restriction of lower-limb loading is frequently instructed by the surgeon or orthopedic surgeon after lower-limb surgery to ensure proper healing of a fracture, osteotomy, or hip arthroplasty.1–6 The general concept behind PWB is to decrease the forces at the healing site by reducing the external load on the operated leg. For fracture healing (to induce bone growth) and for cementless implant fixation (for osseointegration), limited micromotion is necessary. However, too much micromotion can lead to delayed fracture healing or nonunion and less strong fixation of uncemented implants.7 Therefore, it is standard care for the physical therapist to ensure proper limb loading of the operated leg during rehabilitation.

In clinical practice, visual observation is one of the most common methods used by the physical therapist to estimate the amount of loading under the patient’s foot. Other clinical techniques used to assess weight-bearing—for example, palpation and a bathroom scale—have previously been evaluated.7–11 Palpation by placing the hand under the foot of the patient was found to be subjective guesswork at best, and bathroom scales are able to assess weight-bearing only during standing.10–12 Unfortunately, no studies are available to support the validity and/or reliability of visual observation to assess PWB. Previous studies found that visual estimation is an inaccurate method to determine BW.13–15 This may indicate that visual assessment of the amount of weight placed on the leg during standing is also an inaccurate method. However, this provides no information on the accuracy of weight-bearing assessment during walking because the amount of load placed on the legs during walking is caused not only by body mass and acceleration of gravity but also by additional (eg, forward) accelerations of body mass. Hurkmans et al16 found that when physical therapists used visual observation to train and control PWB, 55% of the patients did not load the leg at the prescribed target load. Therefore, visual observation does not seem to be a valid method to assess PWB and can result in inaccurate limb loading, and consequently may lead to complications. At our department, PWB training is mostly performed by a group of physical therapists who have several years of experience in PWB training. Occasionally, during the weekends, PWB is performed by physical therapists who do not have this kind of experience in PWB training. Therefore, we wanted to determine whether inexperienced therapists can visually estimate weight-bearing as well as experienced therapists.

The present study aimed to determine the validity and interobserver reliability of visual observation to assess the amount of weight-bearing on the operated leg of patients with a THA during walking. We also investigated whether the therapist’s amount of experience in PWB training had an effect on the systematic error of visual observation to determine the amount of weight-bearing.
METHODS

Patient Population
We performed a prospective study in which a convenience sample of 10 patients with a THA who were operated 1 to 12 months prior to the study was selected. Patients were included if they gave informed consent, were between 40 and 80 years of age, and received PWB postoperatively using elbow crutches. Patients with neuromuscular diseases, foot orthosis, and foot deformities that needed special footwear were excluded. This study was approved by the institutional review board.

Observers
Visual observation was performed by a group of 10 physical therapists (5 experienced and 5 inexperienced in PWB training). The therapists who were experienced in PWB training provided physical therapy in the department of orthopedics and had at least 5 years of experience in PWB training. The therapists who were inexperienced in PWB training provided physical therapy in the department of neurology, the department of lung diseases, and the department of internal medicine and provided PWB training occasionally. The experienced observers had clinical experience ranging from 8 to 23 (mean 17.8 ± 9.4) years, of which at least 5 years of experience was in PWB training. The clinical experience of the observers without experience in PWB training ranged from 14 to 33 (mean ± SD, 27.8 ± 8.2) years.

A power analysis was conducted and determined that with an α level of 5%, a sample size of 5 was necessary for a t test to have at least 80% power to detect a difference of 10% BW.17

Instrumentation
The actual amount of weight-bearing was measured with an insole pressure device, the Pedar Mobile system,18 which was used as the reference method.18,19 The Pedar Mobile system is a portable device with matrix insoles (2mm thick), each containing 99 capacitive sensors.

A VAS was used by the physical therapists to rate the amount of weight-bearing by visual observation in percentage BW. The VAS used was a horizontal line 100mm in length with 11 equidistantly spaced vertical lines. The descriptions 0% and 100% of weight-bearing by visual observation in percentage BW. The Pedar system was turned on, and a 0 setting was performed.18 Then the patients walked 3 times at their own speed for 15m at each instructed weight-bearing level. The 3 weight-bearing levels were placed in random order. The patients were observed from the operated side (lateral, anterior, and posterior view) by the physical therapists simultaneously, during each of the 3 walking trials. At the end of each of the 3 walking trials, the physical therapists scored an average weight-bearing load on the VAS. The physical therapists were unaware of the 10%, 50%, or 100% BW target weight-bearing levels, nor the weight-bearing level at which the patients were walking during each trial.

Data Analysis
Pedar-m Expert version 8.2 software7 was used to calculate the vertical force data from the Pedar system. From each walking trial, the first and last 2 steps were excluded from analysis. For each step, the maximum peak load was determined, and from these maximum peak loads, the mean ± SD peak load (percentage BW) was calculated. Normative distribution for the mean peak loads and VAS scores was tested using the Kolmogorov-Smirnov test. For comparison of the visual observation method with the reference method, we used the 95% LOA method with multiple observations per subject described by Bland and Altman.21 The mean difference between the 2 methods was calculated, which represents the systematic error (bias) between the measurements. The SD of this difference represents the random error. The LOA were given by the mean difference ±2 × SD, indicating the total error—that is, systematic and random error together.22–24 The 95% CI for the mean difference and LOA was calculated using, respectively, the SEs \( \sqrt{\frac{s^2}{n}} \) and \( \sqrt{\frac{3s^2}{n}} \) with t equal to 2.262 (df=9, α=0.05).23 An assumption for using the 95% LOA is that there is no significant relationship between the difference and the mean—that is, the mean and SD are constant throughout the range of measurements.22 This was determined by plotting the difference against the average and calculating the correlation coefficient.

Because no criteria exist for clinically acceptable LOA for visual observation to assess PWB, we defined a priori that a difference of ±10% BW would be acceptable. Therefore, the percentage of agreement within the 10% BW was calculated.

To determine the interobserver reliability, we calculated the ICC. The ICC was calculated as the ratio of the variance between patients (ie, variance of interest) to the total variance (ie, variance of interest and error variance). If \( \text{Var}_p \) is the variance between patients, \( \text{Var}_r \) the variance between observers, and \( \text{Var}_{pe} \) the variance attributed to the interaction between patient and observers, the ICC is calculated as \( \text{Var}_p / (\text{Var}_p + \text{Var}_r + \text{Var}_{pe}) \).25,26

The LOA and ICC were calculated for the total group of physical therapists and for the experienced and inexperienced group separately. The t test was used to determine a possible effect of experience on the mean difference between visual observation and the reference method. All statistic analyses were performed with SPSS7 for Windows. The level of significance for all tests was set at 5%.

Table 1: Patient Characteristics

<table>
<thead>
<tr>
<th>Patient No.</th>
<th>Sex</th>
<th>Age (y)</th>
<th>BW (kg)</th>
<th>Operated Leg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>57</td>
<td>92</td>
<td>R</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>63</td>
<td>80</td>
<td>R</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>63</td>
<td>84</td>
<td>L</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>58</td>
<td>80</td>
<td>L</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>63</td>
<td>78</td>
<td>L</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>71</td>
<td>82</td>
<td>R</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>78</td>
<td>57</td>
<td>R</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>79</td>
<td>57</td>
<td>R</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>64</td>
<td>90</td>
<td>R</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>60</td>
<td>48</td>
<td>R</td>
</tr>
</tbody>
</table>

Abbreviations: F, female; L, left; M, male; R, right.
The aim of our study was to determine the validity and interobserver reliability of visual observation to assess partial weight-bearing. The results showed no significant systematic error for visual observation to estimate the amount of weight-bearing. However, expressed by a large random error, a considerable lack of agreement existed between visual observation and the reference method. The mean difference between visual observation and the reference was 46.7. There was no significant relationship between the difference and the average of the visual observation and the reference method (r = -0.139; P = 0.464). Therapists’ experience in PWB training had no effect on the mean difference between visual observation and the reference method (P = 0.349).

The ICC was 0.57 for the total group of physical therapists. For the experienced and inexperienced group of physical therapists, the ICCs were 0.47 and 0.59, respectively (see Table 2).

**DISCUSSION**

The patient characteristics and the raw weight-bearing data of the visual estimation VAS scores are presented in Table 1 and Figure 1. After the instruction to load the leg at 3 weight-bearing levels (low, mid, high), we observed that most of the patients did not load their legs on these prescribed target loads.

The mean difference between visual observation and the reference method was -9.5% BW; this difference was not significant (95% CI, -24.0 to 5.0 percentage BW; Table 2, Figure 2). The upper limit of agreement was 30.8% BW, and the lower limit was -49.8% BW. The percentage of agreement within ±10% BW was 46.7. There was no significant relationship between the difference and the average of the visual observation and the reference method (r = -0.139; P = 0.464). Therapists’ experience in PWB training had no effect on the mean difference between visual observation and the reference method (P = 0.349).

The ICC was 0.57 for the total group of physical therapists. For the experienced and inexperienced group of physical therapists, the ICCs were 0.47 and 0.59, respectively (see Table 2).

**RESULTS**

The patient characteristics and the raw weight-bearing data of the visual estimation VAS scores are presented in Table 1 and Figure 1. After the instruction to load the leg at 3 weight-bearing levels (low, mid, high), we observed that most of the patients did not load their legs on these prescribed target loads.

The mean difference between visual observation and the reference method was -9.5% BW; this difference was not significant (95% CI, -24.0 to 5.0 percentage BW; Table 2, Figure 2). The upper limit of agreement was 30.8% BW, and the lower limit was -49.8% BW. The percentage of agreement within ±10% BW was 46.7. There was no significant relationship between the difference and the average of the visual observation and the reference method (r = -0.139; P = 0.464). Therapists’ experience in PWB training had no effect on the mean difference between visual observation and the reference method (P = 0.349). The ICC was 0.57 for the total group of physical therapists. For the experienced and inexperienced group of physical therapists, the ICCs were 0.47 and 0.59, respectively (see Table 2).

The choice of a measuring technique for PWB depends largely on how accurately one needs to measure the amount of weight-bearing in the clinic to avoid complications. Unfortunately, the amount of error in weight-bearing estimation that is acceptable in clinical practice is unknown. In the present study, we used our clinically acceptable LOA of ±10% BW, which resulted in 53% of the visual weight-bearing scores being inaccurate. However, because these clinically acceptable limits may be too strict, clinical studies are needed to assess the relationship between the amount of weight-bearing and post-operative complications. More research is needed on which local forces are harmful for the healing site (eg, hip joint), which forces occur at the healing site during PWB, and what

**Table 2: Validity and Interobserver Reliability of Visual Observation of PWB**

<table>
<thead>
<tr>
<th>Physical Therapists</th>
<th>Systematic Error Mean ± SD 95% CI</th>
<th>Total Error - LOA - LL 95% CI UL 95% CI</th>
<th>Agreement (%) Within 10% BW</th>
<th>Reliability ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (n=10)</td>
<td>-9.5 ± 20.1 -24.0 to 5.0</td>
<td>-49.8 -74.7 to -24.9 30.8 5.8 to 55.7</td>
<td>46.7</td>
<td>.57</td>
</tr>
<tr>
<td>Experienced (n=5)</td>
<td>-8.6 ± 20.5 -26.4 to 9.2</td>
<td>-49.6 -75.0 to -24.2 32.4 7.0 to 57.8</td>
<td>53.3</td>
<td>.47</td>
</tr>
<tr>
<td>Inexperienced (n=5)</td>
<td>-10.1 ± 20.9 -28.4 to 8.2</td>
<td>-51.9 -77.8 to -26.0 31.7 5.8 to 57.8</td>
<td>40.0</td>
<td>.59</td>
</tr>
</tbody>
</table>

Abbreviations: LL, lower limit; UL, upper limit.
the relationship is between local forces and the forces under the foot (ie, the vertical ground reaction forces). Until then, currently prescribed target loads (eg, 10% BW, 15% BW, 50% BW, 8kg) with clinically defined upper and lower limits of 10% BW will be used. For this, visual observation is too crude a method to ensure correct weight-bearing.

In clinical practice, visual observation is also used in combination with palpation of upper-arm muscles and/or the hand-under-the-foot technique to estimate limb loading; to our knowledge, no studies have evaluated palpation separately (or a combination of these 3 methods) to assess PWB. Only 1 study has evaluated the hand-under-the-foot technique, and the authors concluded that the amount of weight placed on the therapist’s hand was subjective guesswork at best. Therefore, more research is needed that evaluates these techniques individually but also in combination to assess the amount of weight-bearing. A promising technique to train patients to load the leg more accurately is audio feedback. However, only 2 studies have evaluated audio feedback for weight-bearing training, and no information was given about the instrument’s validity to measure the limb load. These 2 studies evaluated audio feedback during the patient’s hospital stay, but additionally, because of reduced hospital duration of care, it is important to know whether patients perform proper weight-bearing at home. Therefore, more knowledge is needed regarding the effect of PWB training with audio feedback on the amount of weight-bearing during the entire postoperative recovery—that is, in and outside the hospital.

Study Limitations

The patients with THA participating in this study differed slightly from the patients with THA seen by the physical therapist directly postoperatively. Our patients could load the leg at 100% BW, had no pain, and were not afraid to load the operated leg. Factors like pain and anxiety are known to influence the patient’s walking pattern directly postoperatively and could lead to lower leg loading. Therefore, these factors might help the physical therapist to estimate the amount of weight-bearing by visual observation because of an altered walking pattern. However, in our experience, patients with THA recover relatively quickly with regard to their postoperative pain and anxiety, and often have to be instructed not to walk too fast.

CONCLUSIONS

Visual observation alone is not a valid and reliable method to assess the loading of the leg during partial weight-bearing. Experience of the physical therapist in PWB training did not influence the outcome. Although PWB is frequently performed in the clinic, limited information is available on the validity and reliability of techniques (visual observation, palpation of upper arm muscles, hand-under-the-foot) used by the physical therapist to assess the amount of weight-bearing. Therefore, more studies are needed that evaluate these techniques individually but also in combination to assess the amount of weight-bearing.

Acknowledgments. We thank S. Groenendijk and M. van der Velde for their support in data collection.

References


**Suppliers**

a. Novel GmbH, Ismaninger St 51, 81675 Munich, Germany.
c. Version 12; SPSS Inc, 233 S Wacker Dr, 11th Fl, Chicago, IL 60606.

Arch Phys Med Rehabil Vol 90, February 2009