The Physical Performance Test and the evaluation of functional status in patients with cerebral aneurysms

JOSEPH T. KING JR., M.D., M.S.C.E., JOEL TSEVAT, M.D., M.P.H., AND MARK S. ROBERTS, M.D., M.P.P.

Section of Neurosurgery, Veterans Affairs Connecticut Healthcare System, West Haven; Department of Neurosurgery, Yale University, New Haven, Connecticut; Section of Outcomes Research, Division of General Internal Medicine, Department of Internal Medicine; Center for Clinical Effectiveness, Institute for Health Policy and Health Services Research, University of Cincinnati Medical Center, Cincinnati; Veterans Affairs Medical Center, Cincinnati, Ohio; and Center for Research on Health Care and Section of Decision Sciences and Clinical Systems Modeling, Division of General Internal Medicine, Department of Medicine, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

Cerebral aneurysms can adversely affect the functional status of patients who harbor them through rupture, mass effect, or treatment sequela. The authors compared the Physical Performance Test (PPT)—an observer-administered test requiring the completion of seven tasks—with three functional status measures based on patient self-reports: the Glasgow Outcome Scale (GOS), the modified Rankin Scale (mRS), and the modified Barthel Index (mBI).

Methods. Data were collected from 144 patients with aneurysms who had been recruited from a neurosurgery clinic. The patients completed the PPT and were assigned GOS scores, mRS scores, and mBIs based on their responses during a structured interview. The validity of the PPT was assessed by examining the relationship between the results of the PPT and the values assigned to each patient by using the GOS, mRS, and mBI by applying rank-order methods; the reliability of the PPT was assessed using the Cronbach α coefficient.

The mean age of the patients was 52.8 years; 72% were women and 54% had survived a subarachnoid hemorrhage. The mean ± standard deviation PPT score was 24.0 ± 3.7. The PPT generated a broad distribution of scores, whereas GOS and mRS scores and mBIs displayed significant ceiling effects, that is, 75% of the patients were clustered in the highest categories. The PPT scores were validated by their strong association with GOS scores, mRS scores, and mBIs (for all, \( p < 0.001 \)), and the reliability of the PPT was demonstrated by a Cronbach α value of 0.77.

Conclusions. The PPT provides an objective measure of functional status in patients with cerebral aneurysms. The results suggest that the PPT may differentiate among patients better than the GOS, mRS, or mBI. The PPT is a valid and reliable instrument for measuring functional status in patients with cerebral aneurysms.

KEY WORDS • cerebral aneurysm • Physical Performance Test • functional outcome
vational study performed in a cohort of patients with cerebral aneurysms. We assessed performance by comparing PPT scores with values assigned to patients by using several common functional outcomes scales (GOS, mRS, and mBI). The validity and reliability of an outcomes instrument should be determined before widespread use of the instrument. Validity is the extent to which a test measures what it is intended to measure. Reliability refers to the stability and reproducibility of measures of the same concept over time, across methods of gathering data, or within the same instrument.\textsuperscript{1,9,23} We used GOS scores, mRS scores, and mBIs to assess the validity of the PPT, and we evaluated the interitem reliability of the PPT.

\section*{Clinical Material and Methods}

\subsection*{Patient Population}

Our study focused on patients with cerebral aneurysms at the University of Pittsburgh Medical Center neurosurgery clinics, who were recruited between June 2001 and February 2004. After we had obtained their informed consent, the patients underwent a structured interview and testing administered by trained research assistants. Additional data were abstracted from paper and electronic medical records. The protocol was approved by the institutional review boards of the University of Pittsburgh and Yale University. Patients received $25 after completing the interview.

\subsection*{Physical Performance Test}

We used the seven-item version of the PPT,\textsuperscript{27} which requires each patient to write a sentence, simulate eating, lift a book onto a shelf, put on and remove a jacket, pick up a penny from the floor, walk 50 ft, and turn 360°. All tasks are assigned a score on a scale ranging from 0 to 4. Scores for the first six tasks are based on time to completion, and the score for the turn task is based on steadiness and continuity. The task scores were summed to generate a composite score ranging from 0 to 28, with higher scores demonstrating better functioning for both the individual tasks and the composite score.

\subsection*{Glasgow Outcome Scale, mRS, and mBI}

Patient interview responses and a computerized scoring algorithm were used to classify patients using the GOS\textsuperscript{15} and a modification of the RS\textsuperscript{2} and to calculate a modification of the RS\textsuperscript{2} (mRS). The GOS is used to classify patients into one of five categories: good recovery, resumption of normal life despite deficits (GOS Score 5); moderate disability, disabled but independent (GOS Score 4); severe impairment, conscious but disabled (GOS Score 3); persistent vegetative state (GOS Score 2); and death (GOS Score 1). The mRS is used to categorizes patient into six classes: no neurological symptoms or disabilities (mRS Score 0); no disabilities despite neurological symptoms (mRS Score 1); slight disability, unable to perform all ADL (mRS Score 2); moderate disability, walks without assistance (mRS Score 3); severe disability, unable to walk without assistance (mRS Score 4); severe disability, bedridden, incontinent, lives in a nursing home (mRS Score 5); and death (mRS Score 6). The mBI combines responses from 10 multiple-choice questions used to assess ADL to produce a score ranging from 0 to 20, with higher scores representing better functioning.

\begin{table}
\centering
\begin{tabular}{lc}
\hline
Characteristic & Value* \\
\hline
age (yrs) & \\
mean ± SD & 52.8 ± 12.4 \\
range & 25–83 \\
sex & \\
female & 104 (72) \\
race & \\
Caucasian & 136 (94) \\
African-American & 8 (6) \\
education & \\
high school or technical school graduate & 132 (92) \\
unimpaired (MMSE Score 24–30) & 92 (64) \\
unknown & 8 (5) \\
cognitive functioning & \\
impaired (MMSE Score 0–23) & 44 (31) \\
unimpaired (MMSE Score 24–30) & 92 (64) \\
no. of aneurysms & \\
1 & 106 (74) \\
2 & 19 (13) \\
3 & 14 (10) \\
4 & 7 (2) \\
5 & 1 (1) \\
7 & 1 (1) \\
aneurysm location & \\
aneurysm location & 181 (86) \\
aneurysm location & 29 (14) \\
unsecured aneurysm(s) & 76 (53) \\
endovascular embolization & & \\
endovascular embolization & 20 (14) \\
both & 7 (5) \\
none & 43 (30) \\
\hline
\end{tabular}
\caption{Summary of characteristics in 144 patients with cerebral aneurysms}
\end{table}

\textsuperscript{*} Values are expressed as number of patients (%) unless otherwise noted.

\subsection*{Mini-Mental State Examination}

Patients completed the MMSE, which was administered by trained examiners.\textsuperscript{18} The MMSE consists of seven tasks designed to measure orientation, memory, attention, and naming, as well as the ability to follow verbal and written commands, write a sentence spontaneously, and copy a complex polygon. The tasks are scored individually, and the scores are summed to yield a composite score ranging from 0 to 30. Lower scores represent worse cognitive functioning, and scores lower than 24 are considered indicative of cognitive impairment.

\subsection*{Statistical Analysis}

Categorical variables were tabulated, and means and SDs were calculated. The characteristics of patients in the study population who could be evaluated were compared with those of patients in whom data collection was incomplete by using the Mann–Whitney U-test.\textsuperscript{18} We compared the PPT scores of subgroups within our study population (those with a history of SAH, the presence of an unsecured aneurysm, or previous aneurysm endovascular treatment) by using the Mann–Whitney U-test.\textsuperscript{18} The relationships between PPT scores and functional scores measured using categorical scales (GOS and mRS) were assessed by creating box plots and performing the Cuzick nonparametric test for
The association between PPT scores and the continuous scale mBI was assessed using scatter plots and the Spearman rank correlation, a nonparametric test of association. The Cronbach’s coefficient was used to assess the interitem reliability of the PPT by using the Nunally criterion of 0.7. The Spearman rank correlation was used to assess the relationship between PPT scores and cognitive functioning as measured with the MMSE. Statistical significance was defined by a two-tailed probability value less than 0.05.

Results

Patient Population

Two hundred seventeen eligible patients consented to participate in the study, 144 of whom completed the PPT and answered enough questions to allow the determination of functional status measured using the GOS, mRS, and mBI. Incomplete data collection was caused by errors in survey completion, research staffing issues, and patient time constraints. The mean patient age was 52.8 ± 12.4 years (mean ± SD); 72% of the patients were women and 94% were Caucasian (Table 1). Fifty-four percent of the patients had suffered an SAH, 26% had multiple aneurysms, and 70% had undergone surgical or endovascular treatment of an aneurysm. Eighty-six percent of the aneurysms were located in the anterior circulation. There were no differences between the study population and the group of excluded patients with respect to age, sex, race, education, cognitive functioning, cigarette smoking, hypertension, history of SAH, number of aneurysms, presence of an unsecured aneurysm, aneurysm location, or prior aneurysm treatment (for all, p > 0.061).

Physical Performance Test

The mean PPT score was 24 ± 3.7, and 15 patients (10%) attained the maximum score of 28. The PPT provided a broad distribution of scores across the study population (Fig. 1). Ninety-five percent of patients were able to complete all seven test items (Table 2). A history of SAH, the presence of an unsecured aneurysm, or previous surgical or endovascular treatment did not affect the PPT score (for all, p ≥ 0.118). The value of the Cronbach’s coefficient for the PPT was 0.78, which exceeds the Nunally criterion of greater than 0.70 as a threshold for interitem reliability.

Glasgow Outcome Scale, mRS, and mBI

Patients were assigned scores on the GOS, mRS, and mBI based on their interview responses. All the scales showed a marked “ceiling effect,” whereby most patients
were clustered at the very top of the scale (Fig. 1). Eighty-one percent of the patients had a good recovery (GOS Score 5), 76% had no significant disability and/or deficits (mRS Score 0 or 1), and 81% had no difficulty with ADL (mBI 20). Note that the study cohort was composed of patients well enough to travel to a neurosurgery clinic, and thus patients with the lowest levels of functioning (persistent vegetative state [GOS Score 2] or death [GOS Score 1]; severe disability, bedridden [mRS Score 5] or death [mRS Score 6]) were not included in the analyses.

Mini-Mental State Examination

One hundred thirty-six patients (95%) completed the MMSE. The mean score was 24 ± 2.2, and 44 patients (31%) had scores less than 24, values consistent with cognitive impairment (Table 1). There was no significant correlation between PPT scores and cognitive functioning as measured using the MMSE (p = 0.13, p = 0.130).

Comparison of the PPT and Self-Report Measures

Scores on the PPT were significantly higher in patients with better functioning assessed using the GOS (p < 0.001), mRS (p < 0.001), and mBI (p < 0.001) (Fig. 2). There were rare discrepancies between PPT scores obtained from objective testing and GOS scores, mRS scores, and mBIs derived from self-reports. Four patients (3%) were unable to perform at least one of the seven PPT tasks, were included in the highest category of the GOS, mRS, or mBI based on their self-reported abilities. For example, four patients were unable to pick up a book and place it on a shelf during administration of the PPT, yet based on their interview responses, one (25%) was assigned GOS Score 5 (good recovery, normal activities with or without deficits), two (50%) were assigned mRS Score 0 (no significant disability, no symptoms), and three (75%) received an mBI of 20 (no deficits in ADL).

Discussion

Cerebral aneurysms and their treatment can adversely affect functional status. We used two types of instruments to assess functional status in a population of patients with cerebral aneurysms: the PPT, an objective instrument involving standardized tasks, and the GOS, mRS, and mBI, whose scores are based on the patient’s self-report. The PPT has been used to assess functional status in residents of long-term care facilities; patients with chronic obstructive pulmonary disease; patients who have undergone Achilles tendon surgery or transmetatarsal amputation; and elderly persons involved in exercise programs. In our study population, the PPT produced a wide distribution of scores, whereas the GOS, mRS, and mBI all exhibited clustering of scores at the highest levels of functioning (that is, the ceiling effect). The strong association between PPT scores and the GOS scores, mRS scores, and mBIs served to validate the PPT, and the Cronbach α coefficient demonstrated the reliability of the PPT. Marked discrepancies between observer-based functional status (PPT score) and self-reported abilities (GOS and mRS scores and the mBI) only occurred in 3% of patients.

One of the goals of an outcomes measure is to discriminate among the study population by generating a range of scores with few patients at the extremes of the scale. Some instruments exhibit floor and ceiling effects (in which patients are clustered at the lowest or highest portions of the scale, respectively), which limits their usefulness. The GOS scores, mRS scores, and mBIs demonstrate significant ceiling effects in our study population. These instruments do not provide fine gradations for the assessment of patients at middle and high levels of functioning, and thus may not be well suited to evaluating patients harboring aneurysms who are in an outpatient setting. The GOS, mRS, and mBI tend to “lump” together patients with one or more minor deficits into a single-outcome category. In contrast, the PPT has a wide distribution of scores and a minimal ceiling effect, and appears to be more sensitive to limitations in physical functioning when compared with self-reported measures.

Observer-based functional status measures offer the advantage of uniform data collection and scoring, whereas the scoring algorithms of self-reported measures yield more variable results. As originally described, the GOS did not provide a scoring algorithm; patients were assigned to one of five categories based on the closest match to the descriptions of each category. This led to considerable interobserver variability in scoring and subsequent efforts to improve consistency through a structured interview. Similar to the GOS scores, mRS scores are determined by matching patients’ capabilities to category descriptions. In the current study, we used responses to a structured interview and a computerized scoring algorithm to standardize the GOS and mRS category assignments. In contrast to the GOS and mRS, the original BI and subsequent modified versions rely on responses to a multiple-choice survey to calculate a score. Some response options to Barthel questions are subjective, which may influence final scores. For example, responses to the toileting question are “dependent,” “needs help but can do some things alone,” and “independent.” The PPT involves standardized scoring of timed tasks to generate a score, eliminating subjective respondent effects and minimizing observer bias.

The GOS, mRS, and mBI values are all derived from patient or proxy responses and are therefore subject to reporting errors. The PPT scores are determined by patient performance during a series of standardized tasks, and this performance is quantified by a trained observer. Based on the experiences of other investigators, we expected to find

### TABLE 2

**Detailed results in 144 patients with cerebral aneurysms for the seven-item PPT**

<table>
<thead>
<tr>
<th>PPT Task</th>
<th>No. of Patients Able to Complete (%)</th>
<th>Duration of Task (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>write a sentence</td>
<td>144 (100)</td>
<td>12.5 ± 4.6, 7.0–32.5</td>
</tr>
<tr>
<td>simulate eating</td>
<td>144 (100)</td>
<td>11.0 ± 3.0, 5.5–24.0</td>
</tr>
<tr>
<td>place book on shelf</td>
<td>140 (97)</td>
<td>2.2 ± 1.1, 1.0–7.5</td>
</tr>
<tr>
<td>put on &amp; remove jacket</td>
<td>143 (99)</td>
<td>11.4 ± 5.6, 5.0–40.5</td>
</tr>
<tr>
<td>turn 360°</td>
<td>140 (97)</td>
<td>—</td>
</tr>
<tr>
<td>pick up penny</td>
<td>143 (99)</td>
<td>2.9 ± 1.8, 1.0–12.0</td>
</tr>
<tr>
<td>walk 50 ft</td>
<td>143 (99)</td>
<td>14.2 ± 4.0, 8.5–33.5</td>
</tr>
</tbody>
</table>

* The 360° turn is not a timed task and thus there are no time data to report. This task is scored by whether or not the patient is able to use continuous steps while turning 360°. — = not applicable.
Physical Performance Test in patients with aneurysms

Fig. 2. Left and Center: Box plots showing the distribution of PPT scores for each functional status category of the GOS (left) and mRS (center). Within each category, the central horizontal line represents the median value, the box the interquartile range containing 50% of the responses, the whiskers adjacent values, and the circles outliers. Right: Scatterplot demonstrating the distribution of PPT scores compared with the continuous mBIs. The PPT scores were significantly higher in patients with better functioning assessed using the GOS, mRS, and mBI.

numerous discrepancies between the observer-based assessment (the PPT) and the self-reported assessments (GOS, mRS, and mBI). Reuben and colleagues found poor correlations between self-administered, interview-administered, and objective performance-based measures of physical function. Frequent disagreements in functional status ratings between caregivers and patients have been documented using the PPT in a population of patients with mild dementia. Nevertheless, we found marked differences between PPT scores and self-reported assessments in only 3% of our patients. Our failure to replicate earlier findings may be related to differences in the ages of the patient populations. The mean age of our patient population was 52.8 years, whereas most other studies in which the PPT has been used involved older patients, who may be more susceptible to overreporting of their functional statuses.

The creators of the PPT have suggested that the scale might be useful for screening functional impairment, monitoring functional status changes (such as response to therapy), and possibly predicting future functional decline. These tasks are all relevant to patients with cerebral aneurysms. These lesions can produce a myriad of functional deficits that the PPT could aid in identifying. After SAH or aneurysm treatment, patients may receive physical or cognitive therapy, and the PPT could be used to track improvements. Patients, families, and healthcare providers are often interested in prognosis, and the PPT may provide useful information about what the future holds. Finally, the PPT may be a useful tool for comparing outcomes of various aneurysm treatment strategies.

Because functional status has a cognitive component, we expected to see an association between PPT and MMSE scores. However, we found no association between functional status as measured with the PPT and cognitive functioning as measured with the MMSE. Our measurement of cognitive functioning was limited to the MMSE; time constraints prevented us from administering a full neurocognitive test battery to our patients. The MMSE, while widely used as a brief screening tool for cognitive functioning, is not a substitute for formal neurocognitive testing. Perhaps more extensive neurocognitive testing may have shown an association between functional status and cognitive status. Alternatively, except for writing a sentence, tasks of the PPT are not particularly mentally taxing, and thus this test may not be sensitive to mild or moderate decrements in cognitive functioning.

Several factors may limit the generalizability of our findings. Our study sample was drawn from patients with aneurysms who were receiving care at a single academic institution. Not all eligible patients at our institution enrolled in the study, and one third of the enrollees could not be included in the analysis because they did not complete the PPT, GOS, mRS, and/or mBI. Although the sex and age distributions of our cohort were similar to those reported in many other aneurysm studies, Caucasians were overrepresented in our study relative to the general population. Finally, our findings on the performance, validity, and reliability of the PPT need to be corroborated by other investigators in other series of patients with aneurysms.

Conclusions

The PPT appears to offer several advantages over self-reported measures of functional status such as the GOS, mRS, and mBI. 1) The PPT provides better discrimination between patients by distributing the study population across a broader range of scores. 2) The PPT is not subject to ceiling effects—only 10% of our patients attained the maximum PPT score, whereas greater than 75% of patients achieved the highest scores on the GOS, mRS, and mBI. 3) The PPT is more accurate than the other outcome measures, because it avoids the overreporting bias inherent in many self-reported measures of functional status. These benefits come at a price, however. The PPT requires trained observers, props, and patient cooperation. In addition, the PPT data must be obtained via direct interaction, and cannot be obtained from medical records, telephone interviews, or written surveys. Investigators studying functional outcomes in patients with cerebral aneurysms who require improved discrimination among patients and who can marshal the increased resources needed for data collection should consider using the PPT.

References

18. Mann HB, Whitney DR: On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statistics 18:50–60, 1947

Manuscript received August 30, 2004. Accepted in final form November 1, 2005.

Dr. King is supported by a Mentored Patient-Oriented Research Career Award from the National Institute of Neurological Disorders and Stroke (Grant No. 1K23 NS02169-06).

Address reprint requests to: Joseph T. King Jr., M.D., M.S.C.E., Section of Neurosurgery, Veterans Affairs Connecticut Healthcare System/112, 950 Campbell Avenue, West Haven, Connecticut 06516. email: Joseph.KingJr@med.va.gov.