

# Establishing a Standardized Clinical Assessment Tool of Pathologic and Prosthetic Hand Function: Normative Data, Reliability, and Validity

Colin M. Light, PhD, Paul H. Chappell, PhD, Peter J. Kyberd, PhD

**ABSTRACT.** Light CM, Chappell PH, Kyberd PJ. Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: normative data, reliability, and validity. *Arch Phys Med Rehabil* 2002;83:776-83.

**Objective:** To develop a new assessment procedure, the Southampton Hand Assessment Procedure (SHAP), that allows contextual results of hand function to be obtained readily in a clinical environment.

**Design:** Reliability (test-retest, interrater) and validity (criterion, content) of new assessment procedure against standard medical outcome measure techniques.

**Setting:** Normative data collected in a university laboratory.

**Participants:** Twenty-four volunteers selected on the basis of optimum hand function using these criteria: age (range, 18-25y), and no adverse hand trauma, neurologic condition, or disabling effects of the upper limb.

**Interventions:** Not applicable.

**Main Outcome Measures:** The normative control group was assessed for variability, and the procedure measured in terms of interrater and test-retest reliability. The absence of a direct comparison prevents a criterion standard from being established; however, content validity was shown by expert peer review.

**Results:** The control group data were shown to be multivariate gaussian; test-retest and interrater reliability were demonstrated at the 95% confidence level. The content validity was demonstrated by peer panel approval.

**Conclusions:** Results of the control group established the statistical integrity of SHAP. Clinical trials are underway, although more extensive use of the procedure is advocated in primary care and rehabilitation centers where physiotherapy and occupational therapy are actively used in hand rehabilitation.

**Key Words:** Hand injuries; Rehabilitation; Treatment outcome.

© 2002 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

**F**UNCTIONALITY IS A BROAD term that has application in a wide variety of disciplines. Its use in relation to hand evaluation techniques encompasses dexterity; gross manipula-

tive ability; and, more generally, the ability to perform tasks encountered during everyday living. Considering the scope of this terminology, we suggest a generic definition of functionality, suitability to the task. According to this definition, the level of functionality at which the hand or upper limb performs must be a measure of how adaptable, or suitable, it is to performing the required tasks.

Despite the specific nature of quantifying upper-limb function, the underlying principles of evaluation stem from the desire to analyze medical outcome measures. There are several reasons to use clinical measurement for monitoring treatment effectiveness (which also embodies medical devices such as prostheses). The overall objective is to improve the quality of medical care by assessing the performance of existing techniques.<sup>1</sup> However, the more immediate benefits lie in the discrimination between individuals (or groups) on an underlying health issue, such as functional performance. If the magnitude of these differences can be quantified, researchers can establish group or population trends. Outcome measures are also frequently warranted for evaluating and monitoring longitudinal change in individuals, especially during a course of treatment, rehabilitation process, or for assessing pre- and postoperative performance.<sup>2</sup> These needs have led to the development of the Southampton Hand Assessment Procedure (SHAP).

This assessment tool is designed to be a standardized and objective method of evaluating pathologic or prosthetic hand function, and its reliability and validity are established in this article. Furthermore, the average assessment time of 20 minutes and its portable nature show the clinical focus of this procedure.

## Natural Hand Assessment

Although the hand contributes about 90% of the function to the upper limb,<sup>3</sup> the upper extremity is an entire system, the coordinated movements of which create overall mobility and dexterity. To distinguish the functionality of the hand, it must be assessed as an isolated manipulator (ie, decoupled from the rest of the upper limb) wherever possible.

Hand assessment procedures also warrant the identification of prehensile patterns as a further means of validation. The distinct functional positions of the hand (fig 1) are arguably vital to the evaluation by ensuring assessment of the complete range of grip structures. Although there is little conformity to specific classifications of prehensile patterns,<sup>4-9</sup> the general characteristic remains largely consistent<sup>10</sup> (fig 1): tripod pinch, tip pinch, lateral pinch, power grip, spherical (or flexion) grip, and extension grip.

Although many hand assessment procedures have been devised and implemented, there is little or no uniformity among them, which precludes any form of meta-analysis. The reader is referred to a comprehensive critical review of existing techniques.<sup>11</sup> The review reveals little or no conformity to a standardized procedure for the assessment of hands, both natural and prosthetic. Conventionally, measurement of hand function by these procedures has been by time limit (quantity completed

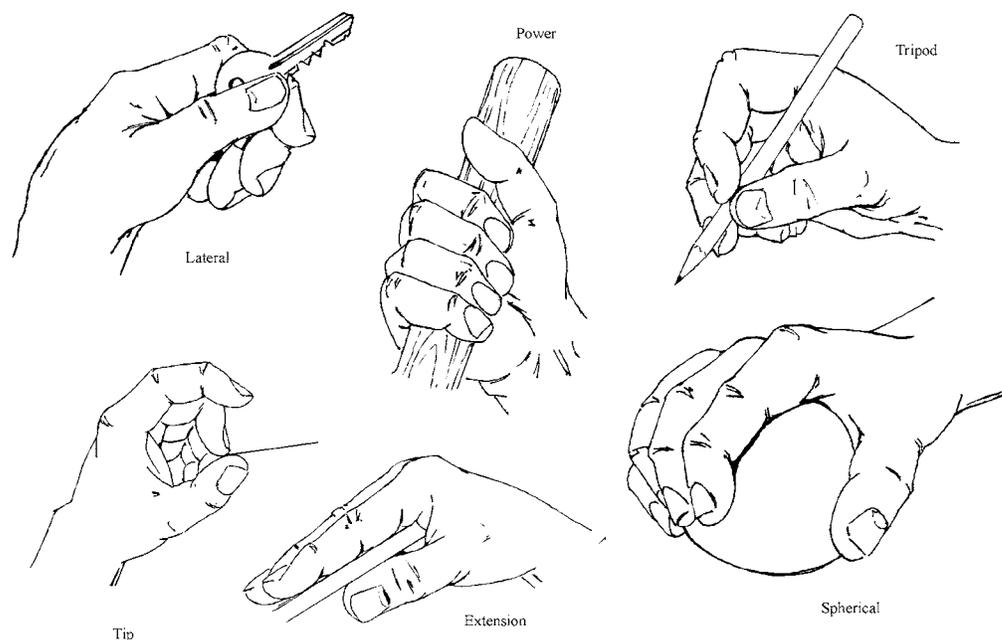
From the Department of Electronics and Computer Science, University of Southampton, Southampton (Light, Chappell); and Oxford Orthopaedic Engineering Centre, Nuffield Orthopaedic Centre, Oxford (Kyberd), England.

Accepted in revised form November 7, 2001.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

Reprint requests to Paul H. Chappell, PhD, Dept of Electronics and Computer Science, University of Southampton, Highfield, Southampton, SO17 1BJ, England. 0003-9993/02/8306-6827\$35.00/0

doi:10.1053/apmr.2002.32737



**Fig 1. Prehensile pattern classifications.** Courtesy of Colin M. Light, PhD © 1998. Reprinted with permission.

Artist: R. O'Neill

in a finite time period), work limit (time to complete task), qualitative scoring (the way in which the object is handled), or assessor's opinion (eg, level of difficulty for the subject). Time is an easy parameter to measure and manipulate statistically, but it is not necessarily the most valid measure of hand function.<sup>12-14</sup> Despite this limitation, without the development of more complex, specialist apparatus, investigators have little alternative but to quantify results by time.<sup>15,16</sup>

Assessments that require subjective ratings<sup>17-23</sup> cannot easily lead to the production of a standardized evaluation procedure, because opinion cannot be considered conformist when data collation may be from global and international sources. This caveat is particularly valid when subjectively assessing grip stability or difficulty of prehension. Consequently, a demand exists for a standardized and reliable procedure that can be accomplished with minimal training of the therapist or the user.

The conclusions of the review<sup>11</sup> highlight that what most procedures have in common is an absence of statistically evident standardization, reliability, and validity. Few procedures are based on the assessment of prehensile patterns. These procedures, therefore, rarely ensure a true reflection of the range (and proportional use) of hand function during everyday living. Moreover, the clinical norm of subjective measurement frequently relies on coarse grading scales that, while demonstrably repeatable, are too inaccurate to be considered a true metric of disability. Instead, they simply provide a hierarchic ranking. Although procedures such as the Test Évaluant la Performance des Membres supérieurs des Personnes Agées<sup>24</sup> and the Action Research Armtest<sup>25</sup> provide a comprehensive approach to the functional evaluation of the upper limb, they lack the specific focus of assessing hand impairment.

### Prosthetic Hand Assessment

Evaluations of hand prostheses are even harder to assimilate than the procedures that relate solely to the natural hand. Existing solutions have used either superficial techniques (loosely based on hand assessment tests), or have focused on engineering evaluations (ie, range of motion, strength, system

reliability). Since the advent of commercial myoelectric prostheses, these tests often have been used to evaluate the comparative benefit, or disadvantage, of these devices over body-powered split hooks,<sup>26,27</sup> rather than striving to produce an independent assessment of functionality.<sup>21,22,28</sup> This approach has arisen from the historical need to evaluate new prostheses against the benchmark of the split hook device.<sup>14,29</sup>

Although these measures adhere to medical outcome measurement standards (ie, standardization and statistical evidencing of a procedure's efficacy), researchers have made little attempt to objectively quantify the level of functionality achieved by a hand prosthesis. Even fewer investigators have attempted to compare such a rating against those achievable by the natural hand (whether dysfunctional or healthy). Establishing a context of hand performance, whether pathologic or prosthetic, permits ready identification of an individual's functionality.

The assessment of prosthesis users warrants specific criteria. Prosthesis users require coordinated movement of the upper limb and therefore do not exhibit the separable functions of hand shaping and arm movement seen in natural upper-limb subjects.<sup>30</sup> Hence, conventional assessment of the user does not isolate hand function. Unilateral prosthesis wearers rarely use the device for reaching and grasping of objects, and it mainly fulfills a stabilizing role for the natural hand in bimanual tasks.<sup>31</sup> The wearer's functional ability depends a wide range of factors relating to prosthetic use (eg, time between amputation and prosthesis fitting user's age, weight of the device).

The purpose of the SHAP is not to establish a prosthesis wearer's overall function, which is actually a measure of their adaptation to their disability. Instead, it is to determine the effectiveness of a terminal device and controller by focusing the evaluation on the unilateral performance of the user. This process should highlight functional differences between devices and suitable control schemes<sup>32</sup> for the wearer, while still providing a contextual measure of hand function relative to that of a norm.

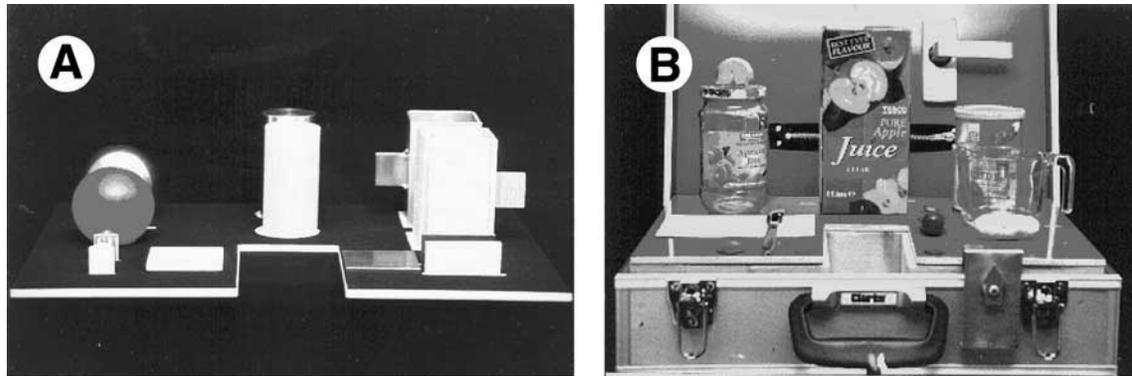


Fig 2. (A) Abstract objects and form board and (B) SHAP ADLs.

The SHAP has broad research implications for the investigation of specific patient cohort groups, such as the evaluation of hand function in osteoarthritis sufferers, stroke patients, individuals using functional electric stimulation, burn victims, and persons with hand trauma injuries.

### METHOD

During the development of the assessment procedure, a number of tasks were considered and their relative merits evaluated. The review of existing techniques<sup>11</sup> provided a forum for emphasizing potentially useful activities as well as highlighting those that may be inappropriate. The following section outlines the methodology for inclusion or exclusion of these activities that ultimately result in the 26 timed tasks that comprise the SHAP (12 abstract tasks, 14 activities of daily living [ADLs]).

#### Abstract Object Tasks

The stability of grip or deviation of prehension from a norm has consistently been evaluated by subjective assessor opinion. This approach causes difficulty in reconciling the standardization of the procedure. Hand assessment techniques are typically composed of a series of ADLs that are used to evaluate functionality; however, other methods do exist. Form-board tasks<sup>3,19</sup> have been used in different formats for both natural and prosthetic hand assessment. This abstract test evaluates prehension without the complication of tools or equipment used during ADLs that often cause intermediate grip patterns or adverse evaluation effects. Because the shape and form of an ADL task is likely to be known to the subject, a psychological prejudice may exist as to his/her ability to perform the task. The abstract objects remove such an effect to a limited extent.

The adaptation of the form board test for SHAP has led to a procedure in which specific grip patterns (determined by object shape) are used to account for the functional range of the hand. In timed tasks, subjects move each object individually. Any subject unable to perform the natural grip for a specific object is expected to take longer by implementing an abnormal grip pattern in order to move the object with an optimum balance of grip pressure and stability. Subjects with impaired hand function (pathologic or prosthetic) frequently perform compensatory movements that enable them to carry out the task more quickly than if their natural pattern had been implemented. Although they are encouraged to achieve natural prehension, the result is, nevertheless, scored because it remains a reflection of the subject's functional ability.

The form board objects are produced in 2 sets for use in the SHAP procedure (fig 2A). The first set is made of noncompliant dense materials (ie, heavyweight abstracts), and the second is from marginally compliant, low-density materials (ie, lightweight abstracts) to produce a difference in both weight and yield. This arrangement provides a means of implicitly evaluating the method of grip implementation and also mitigates for subjects with poor grip pressure.

The Pigeon-Hole Test has been used to evaluate the functionality of prostheses during upper-limb movement.<sup>19,28</sup> This procedure requires the subject to move different objects through several height levels (mouth, chest, waist, knee), and focuses on the individual's ability to grasp and release objects at differing heights. However, this test requires coordinated tracking, trajectory, and grasping motions (a notably different process in prosthesis users than in normals<sup>33</sup>) and is questionable as a method of assessing pure hand function. This omission will fail to highlight the poor performance of body-powered devices in this area of upper-limb movement. However, the assessment procedure is designed to allow an evaluation of all forms of hand, natural and prosthetic, and therefore precludes, or at least minimizes, the assessment of arm function wherever possible. All SHAP tasks focus on prehensile ability and involve minimal transport effects, thereby limiting the assessment of gross upper-limb function.

#### Activities of Daily Living

The compilation of ADLs was sourced<sup>11</sup> from the most significant and reliable hand or prosthesis assessment procedures (table 1, fig 2B). Although not reliability tested, the essential tasks identified by McWilliam<sup>34</sup> were also included in

Table 1: ADLs and Their Grip Classifications

No.	Task	Natural Grip Classification
1	Pick up coins	Tip
2	Undo buttons	Tip/tripod
3	Simulate food cutting	Tripod/power
4	Simulate page turning	Extension
5	Remove jar lid	Spherical
6	Pour water from jug	Lateral
7	Pour water from carton	Spherical
8	Move empty tin	Power
9	Move full jar	Power

the selection criteria to ensure that the broadest range of ADLs were considered.

Any activities that required subjective assessment or that were likely to cause a large variability in timing were omitted. For example, the task of writing, although an important everyday activity, was excluded on the basis of large variability in the performance of writing skills with little relevance to hand functionality (ie, the speed of writing is not necessarily correlated to hand dexterity).

Although some SHAP tasks appear to be bimanual (eg, the button board), the procedure requires these to be carried out unilaterally. For these tasks, the opposing hand acts only as a stabilizer, thereby ensuring functional assessment of the impaired hand.

Sollerman and Ejekkar<sup>35</sup> estimated the percentage use of the 8 types of grip pattern during everyday living. These results were incorporated within the 6 SHAP prehensile categories described previously, thereby obtaining estimates of grip pattern use during habitual activities.

Each selected activity was assigned to the most appropriate prehensile pattern classification(s). The SHAP ADLs were then compiled in approximate proportion to the Sollerman and Ejekkar<sup>35</sup> percentage use values. Consequently, in the SHAP ADLs, a spherical grip is required for 10% of the tasks, a tripod grip for 10%, a power grip for 25%, a lateral grip for 20%, a tip grip for 20%, and an extension grip for 10%. This configuration ensures that the full range of natural grips has been evaluated. It also enables the measure of overall functionality to reflect everyday hand performance without the need for any weighting or adjustment of the prehensile pattern data. By using these criteria, we selected the ADLs as shown in table 1.

The SHAP avoids the traditional subjective method of assessment by the use of a self-timed technique. To prevent a subject taking too long (or being unable) to complete a task, a boundary condition must be introduced. Other procedures have imposed boundary times and conditions without consideration of the individual nature of the task.<sup>12,35,36</sup> For example, Jebsen et al<sup>12</sup> limited subjects to a time of 80 seconds for each task, which ranged from writing (normal mean time, 12s) to the moving of large light objects (normal mean time, 3s). A single boundary condition is inappropriate, considering the large variability in task times.

This time limit can be viewed as the point of minimal function. Myoprostheses users take approximately 6 times as long to complete a task as do subjects with natural hand function and twice as long as persons using a body-powered prosthesis.<sup>37</sup> As one of the slowest functional groups, myoprostheses wearers have little function but nevertheless require classification on the functional scale. Given consideration of previous assessment procedures, as well as the average expected performance of those with severely impaired hand function a boundary condition of 8 times that of the norm is imposed for each SHAP task.

### Creating a Scoring System: The Index of Functionality

The goal of the hand assessment procedure is to produce a tangible and meaningful metric of a subject's function compared with that of a peer norm. Reaction-time effects from both assessor and subject have caused adverse standardization effects in other hand-assessment procedures.<sup>38</sup> However, an overall functionality measure must reflect the prehensile groups that are integral to the procedure rather than being a summation of individual task times (as is the case with the majority of timed assessment procedures to date).

Because each prehensile pattern, or profile, contributes individually (but to differing extents) to the subject's function, a

multivariate approach is necessary to establish the 6 prehensile profiles, and the resultant overall Index of Functionality (IOF). This goal was achieved by using the Euclidean squared distance, which is a measure between samples in an  $i$ -dimensional problem, where  $l = 1, 2, \dots, 6$  (prehensile patterns) in this case.<sup>39</sup> This distance  $d$  is determined by using the  $z$  value (see equation 1) for each of the prehensile patterns, thereby giving a multivariate metric from the norm in each case (see equation 2).

$$z_i = \frac{x_i - \bar{x}_i}{s_i} \quad (1)$$

where  $x_i$  is the subject's time for prehensile pattern  $i$ ,  $\bar{x}_i$  is the mean time for pattern  $i$  in the normative sample, and  $s_i$  is the standard deviation of times for pattern  $i$  in the normative sample.

$$d = \sqrt{\sum_{i=1}^6 (z_i)^2} \quad (2)$$

The distance measure  $d$  is rescaled to a value of 100 when  $x_i$  is equal to the corresponding  $\bar{x}_i$ , and diminishes to 0 for a subject who reaches the boundary condition for each task (and hence is deemed to have minimal function). This measure is known as the subject's IOF and provides an original metric capable of distinguishing between levels of function attributable to specific prehensile patterns.

In a similar manner, an individual IOF may be obtained for each of the prehensile patterns, which themselves are comprised of multiple tasks. Each of these individual indices is based on the Euclidean distance in  $t$ -dimensional space, where  $t$  is the number of tasks in the patterns, ranging from 4 to 7 depending on the prehensile pattern. To automate the calculation of the IOF and therefore aid subject assessment in the clinical environment, a graphics user interface and database was created.

### Establishing Standardization

Medical outcome measures must adhere to a set of procedural requirements to ensure their efficacy. The SHAP has been evaluated according to the methodology (described below) to ensure conformity to these standards.

A standardized procedure is an essential component in any evaluation process to ensure that the study is both repeatable and reliable. The traditional approach to assessment, especially in the health care environment, has focused on the consultation of experts either individually or as a consensus panel.<sup>40</sup> This form of global subjective measure not only is divergent from the goal of standardization but also highlights the invalidity of comparative studies, especially when data are taken from sources outside the original group.

The procedure for executing a SHAP test is documented in detail<sup>41</sup> and includes specification of the subject's initial position relative to the assessment board and the exact layout of each task. It also specifies the prehensile pattern to be used during the demonstration of each activity and the subsequent instructions that must be given to the subject. Hence, a standardized procedure has been established that ensures, as much as possible, that the process is both repeatable and reliable.

Task times are an absolute measure: they fit a ratio scale (with a known zero). The outcome of the SHAP avoids subjective assessment and ensures that results possess a magnitude effect rather than the relative ranking used in alternative scales.<sup>2</sup> However, speed is not synonymous with hand function.<sup>42,43</sup> Other factors, such as grip strength, inherent stability

of the hand, and proprioception, contribute to overall functionality. Because essential hand function and the speed of task performance are determined by the hand's ability to form a natural and optimal prehensile structure, the time taken to execute a task will correlated strongly with function. Although this execution time undoubtedly varies among normal subjects, pathologic hand function will almost certainly cause that speed to differ from that of the norm to an appreciable degree. To comprehensively assess all factors that affect impairment would require a procedure so complex and expensive that it would be precluded from clinical use. Founding the SHAP measure on prehensile patterns should result in a more valid evaluation of hand function than is possible from existing timed procedures.

### Establishing Normative Data

To show statistically that a procedure adheres to medical outcome measure requirements, it is necessary to establish a normative control group while adhering to ethical guidelines. Ethics approval for the present study was obtained from the Southampton and South West Hampshire Joint Research Ethics Committee. The control group was comprised of undergraduates between the ages of 18 and 25 years who had experienced no adverse hand trauma, neurologic condition, or disabling effects of the upper limb. We hypothesized that these subjects would possess near optimum hand function because they were within a prime health age group and had no specialist occupation that would adversely affect performance.

Initially, a group of 18 subjects (9 men, 9 women; designated group A) were assessed, and each evaluation was replicated 3 times. A single assessor was used throughout these control studies, and a minimum period of 24 hours was allowed between replicate assessments (thereby attenuating a direct learning effect by the subjects).

Interrater reliability was evaluated by assessing 3 more subjects (group B). And an additional 3 normative studies (group C) were performed by a single assessor, thereby producing an overall group of 24 samples (each with 3 replicates). Their results formed the basis from which all subsequent assessments of pathologic (or prosthetic) hand function were referenced.

### Variability

The variability of the control group must be shown to possess a normal distribution in order to meet outcome measurement criteria.<sup>2</sup> A common method of verifying normality is to form a normal quantile plot.<sup>44</sup> If the data are normally distributed, all points in the plot will map close to a straight line.

### Reliability

Reliability is perhaps the most crucial measure of the assessment's efficacy in demonstrating that the procedure is truly objective. The test must produce consistent results among a control group with both a single assessor (test-retest reliability) and multiple assessors (interrater reliability).

A single rater assessed control group A with 3 replicate evaluations for each subject. To establish test-retest reliability, it is necessary to show minimal variance in the interaction of the subject and the tasks (thereby indicating that there is no statistically significant effect in the replicate trials).

The most appropriate method to determine whether the data have test-retest reliability is by an analysis of variance (ANOVA) based on the null hypothesis that no significant difference exists between replicates. This hypothesis was tested in the present study at an  $\alpha$  level of .05 (the significance level related to the probability of having a type I error, ie, rejecting

a true hypothesis). The F value obtained from the ANOVA must exceed an  $F_{crit}$  value, which is based on this 95% confidence interval (CI) to prove that the null hypothesis should be rejected and therefore that there is a statistically significant difference between replicates.<sup>45</sup>

The disadvantage of this statistical technique is that the result depends not only on the degree of departure from the null hypothesis, but also on sample size.<sup>46</sup> Consequently, the result of performing an ANOVA test on all 18 subjects, covering all 26 tasks and 3 replicate trials (a total of 1403 degrees of freedom) means that small effects can readily achieve extreme significance levels.

### Validity

Reliability is not a singularly sufficient measure for the procedure's effectiveness because validity must also be ensured wherever possible. However, validity can be extremely difficult to quantify during the development of any new procedure because validation process is usually by comparison with a criterion standard. Existing techniques either fail to comprehensively adhere to medical outcome measurement criteria, or evaluate gross upper-limb function. If the objectives of criterion standards do not adhere to those of the procedure under validation, then the result is largely irrelevant. Hence validation of SHAP compared with a clinical criterion standard procedure was rejected as unfeasible.

An alternative approach would be to establish consistency between clinicians' subjective ratings of disability and the SHAP IOF scale. However, validating an objective ratio scale relative to a subjective ordinal scale would be inappropriate. As with other new clinical outcome measures, it is believed that the traditional criterion validity cannot be applied to the SHAP IOF scale.

## RESULTS

### Establishing Normative Data

Figure 3 shows the mean task times obtained from the normative group. Error bars indicate twice the standard error of the mean, thereby depicting a 95% CI. Although all subjects were provided with an IOF relative to that of the optimum norm, it would be advisable to establish additional normative data for different age groups. Because function is expected to diminish with age,<sup>47</sup> it would be sensible to provide clinicians with an age-matched benchmark of normal function.

### Variability

To ensure comprehensive normality, we tested each task from the normative database, as well as the multivariate IOF. A linear fit trendline to the normal quantile plots for each task revealed  $R^2$  values ranging from .88 (heavyweight extension, abstract object) to .99 (pouring water from a carton, ADL), with an overall average of .93. The control data for each task are normally distributed, as shown by the near straight-line relationship in each plot.

To use a multidimensional metric for the index and profiles of functionality, it is necessary to establish the multivariate normality of the control group. If the data are multivariate normal, then the Mahalanobis distance<sup>39,48</sup> from an individual will have an F distribution when rescaled appropriately. We calculated the scaled Mahalanobis distance values for each subject in the normative group, sorted them into increasing order, and plotted them against the quantiles of the  $F_{6,18}$  distribution. The resultant plot has an  $R^2$  value of .93 to a linear trendline, thereby indicating that the data appear to be near multivariate gaussian.

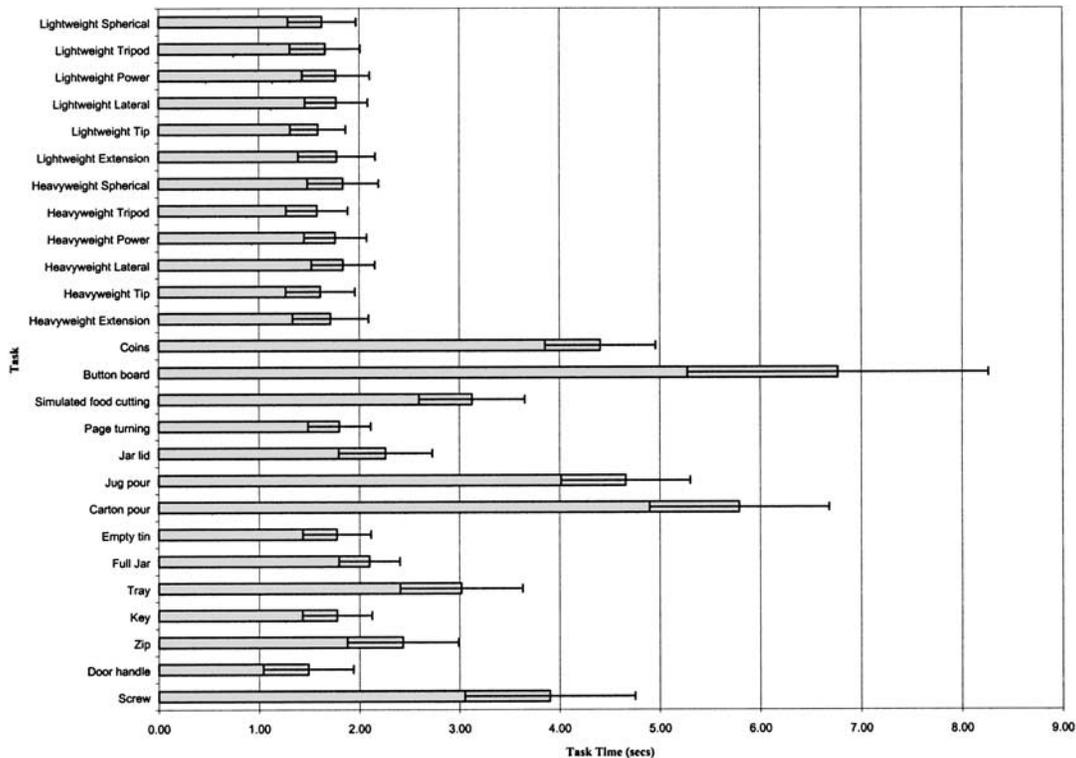


Fig 3. Mean norm task times and variation.

### Reliability

An ANOVA test was performed for all subjects and replicates on a task-by-task basis. If a procedure is repeatable at an elementary level, then the overall assessment should also adhere to this level of reliability. The result indicates that the  $F_{crit}$  value was not exceeded for any of the 26 tasks, however, some tasks did appear to show a departure from the null hypothesis (with  $F$  values exceeding unity). This suggests therefore that some tasks are less repeatable than others based on the normative data available, although not to a statistically significant extent.

To verify this effect, the Euclidean multivariate metrics for all subjects and replicates were also tested for reliability. The ANOVA test revealed an  $F$  value of .39 ( $F_{crit}=3.28$ ) and  $P$  value of .68, thereby indicating that there is no statistically adverse effect in the repeatability of the assessment procedure. Because the subjects were undoubtedly affected by external factors (eg, the trials being performed at different times of the day or in some cases in different locations), this result indicates the procedure's test-retest reliability.

To establish consistency between assessors, we constructed an experiment using normative group B and 3 different raters. All evaluations were performed consecutively for each subject, and the experiment was performed in 1 time period to minimize external effects from influencing subject performance. A Latin square design approach was adopted<sup>46</sup> that also eliminates the order of assessment (or learning) effects.

The rater was found to be statistically insignificant ( $F=2.65$ ,  $F_{crit}=3.09$ ,  $P=.075$ ) at the 95% CI level, as was the interaction between the rater and the subject ( $F=2.12$ ,  $F_{crit}=2.46$ ,  $P=.084$ ), and the interaction between the rater and the task ( $F=.75$ ,  $F_{crit}=1.48$ ,  $P=.87$ ). Consequently, the assessor appears to have statistically little effect on either the execution of

the assessment procedure or the subject's performance thereby indicating interrater reliability.

Internal consistency is a reliability effect that establishes consistency between measures. However, in the case of SHAP, these measures are independent rather than homogeneous assessments of a subject's specific area of disability. Consequently, any internal consistency effect is meaningless in this context.

### Validity

The content validity of the assessment procedure can be shown by the methodology of development. The critical review of existing techniques<sup>11</sup> highlighted specific areas of weakness as well as extracting topics considered important by consensus opinion. The results of the present study were used to form the SHAP and were subsequently presented to a panel of hand therapists. The panel consisted of occupational therapists and physiotherapists from the University of Southampton and the Wessex Rehabilitation Centre at Salisbury District Hospital, UK. Their recommendations were implemented, and the panel's subsequent approval established the content validity of the procedure as far as possible.

Clinical trials have been underway for the past 24 months and continue at the Wessex Rehabilitation Centre, Oxford Orthopaedic Engineering Centre (UK), and the Institute of Biomedical Engineering (University of New Brunswick, NB, Canada). Because detailed procedural instructions are provided, a short training period (usually <1h) is sufficient for the therapists to feel competent to perform independent assessments (each typically taking 20–30min). Preliminary studies showed significantly impaired hand function among prosthesis users, as expected. These subjects' IOF scores ranged from 37 to 48 (compared with an IOF rating of 100 in normals). The

precision grip types of tripod and tip grips proved notably difficult for them, and their lateral prehension was also below their overall IOF scores (unless an active wrist unit was fitted to the prosthesis). Other assessments included subjects with partial hand loss, extensor tendon grafts, and fractured wrists. The IOF scores revealed diminished functionality when these subjects executed the prehensile patterns that are typically difficult for this associated injury. Treatment effectiveness shown in some cases: multiple assessments produced improvements in IOF scores ranging from 4% to 23%.

### DISCUSSION

Adherence to outcome measurement guidelines is crucial to the efficacy of any assessment procedure. Establishing a standardized protocol, forming a normative database, and evaluating the reliability and validity of the procedure are fundamental to this process and yet are absent from most clinical evaluations of functionality. The SHAP fulfills these characteristics in the following manner:

1. It consists of 12 abstract object tasks and 14 ADLs, each of which is founded on 1 (or more) of 6 prehensile patterns. The expected proportion of everyday implementation of these prehensile groups is reflected throughout the procedure.
2. The self-timed nature of SHAP eliminates the need for subjective opinion by the assessor.
3. A standardized procedure ensures that assessments are consistent.<sup>41</sup> The data from the control group have been shown statistically to be normally distributed.
4. The SHAP has been shown to be reliable by statistically insignificant differences between subjects' performance during replicate assessments or with various assessors. Internal consistency measures are inappropriate in this instance.
5. Criterion validity cannot be established because of the lack of a benchmark, however, critical review and expert consensus opinion indicate the content validity of the procedure.
6. The procedure is able to evaluate the functionality of hand prostheses, whether they are passive, mechanical, or myoelectrically controlled and is unbiased to the type of terminal device used.
7. It takes approximately 20 minutes to complete and is a self-contained portable unit ideally suited to use in a clinical environment.

### CONCLUSION

There is little or no conformity to a standardized and objective procedure for the assessment of pathologic and prosthetic hand function.<sup>11</sup> Existing procedures frequently fail to adhere to medical outcome measurement design criteria or are unable to comprehensively cover the prehensile range of the hand.

The SHAP has been designed to account for these shortcomings and to permit evaluation of hand function in the clinical setting. The SHAP's outcome measure is a contextual rating of functionality (relative to that of normal hand function) that enables the clinician to initially determine the subject's disability, and subsequently monitor the client's performance throughout a course of treatment or rehabilitation.

The SHAP has broad implications for the assessment of hand function, ranging from clinical groups (such as burn victims or stroke patients) to the research arena (eg, the investigation of impairment in wrist fracture patients). It also enables clinicians to quantify, compare, and monitor the unilateral functional performance of hand prostheses and controllers worn by their patients.

A normative database (N=24; age range, 18–25y) has been established as the benchmark of normal hand function. This control group has also been used to establish the statistical integrity of the hand assessment procedure and shows an effective outcome measure is obtained from SHAP.

**Acknowledgments:** We gratefully acknowledge the contributions of Bridget Ellis, School of Occupational Therapy and Physiotherapy, at the University of Southampton, UK. Particular thanks and credit is due to Carol Bexon, OT, Megan Greenstock, PT, and all therapists at the Wessex Rehabilitation Centre, Salisbury District Hospital, UK, who provided valuable clinical feedback. We also thank Dinah Stocker, OT, of the Institute of Biomedical Engineering, University of New Brunswick, NB, Canada, for contributions of prosthetic therapy and assessment expertise.

### References

1. Lawrence M, Schofield T, editors. Medical audit in primary health care. Oxford: Oxford Univ Pr; 1993.
2. Wilkin D, Hallam L, Doggett MA. Measures of need and outcome for primary health care. Oxford: Oxford Univ Pr; 1992.
3. Magee DJ. Orthopedic physical assessment. Philadelphia: WB Saunders; 1997.
4. Kapandji IA. The physiology of the joints: upper limb. Vol 1. Edinburgh: Churchill Livingstone; 1982.
5. Landsmeer JM. Power and precision handling. *Ann Rheumat Dis* 1962;21:164-70.
6. MacKenzie CL, Iberall T. The grasping hand. Amsterdam: North-Holland; 1994.
7. Napier JR. The prehensile movements of the human hand. *J Bone Joint Surg Br* 1956;38:902-13.
8. Jacobsen C, Sperling L. Classification of the hand grip: a preliminary study. *J Occup Med* 1976;18:395-8.
9. Tylor CL, Schwarz RJ. The anatomy and mechanics of the human hand. *Artif Limbs* 1955;2:49-62.
10. Kamakura N, Matsuo M, Harumi I, Mitsuboshi F, Miura Y. Patterns of static prehension in normal hands. *Am J Occup Ther* 1980;34:437-45.
11. Light CM, Chappell PH, Kyberd PJ, Ellis BS. A critical review of functionality assessment in natural and prosthetic hands. *Br J Occup Ther* 1999;62:7-12.
12. Jebsen RH, Taylor N, Trieschmann RB, Trotter MJ, Howard LA. An objective and standardized test of hand function. *Arch Phys Med Rehabil* 1969;50:311-9.
13. Moberg E. Criticism and study of methods for examining sensibility in the hand. *Neurology* 1962;12:8-19.
14. Edelstein JE, Berger N. Performance comparison among children fitted with myoelectric and body-powered hands. *Arch Phys Med Rehabil* 1993;74:376-80.
15. Walker PS, Davidson W, Erkman MJ. An apparatus to assess function of the hand. *J Hand Surg [Am]* 1978;3:189-93.
16. Memberg WD, Crago PE. A grasp force and position sensor for the quantitative evaluation of neuroprosthetic hand grasp systems. *IEEE Trans Rehabil Eng* 1995;3:175-81.
17. Carroll D. A quantitative test of upper extremity function. *J Chronic Dis* 1965;18:479-91.
18. Smith HB. Smith hand function test. *Am J Occup Ther* 1973;27:244-51.
19. Kay HW. An evaluation plan for the Beograd hand. In: *Proceedings of the 3rd Symposium on Advances in External Control of Human Extremities*; 1969 Aug; Dubrovnik (Yugoslavia). p 129-37.
20. Sanderson ER, Scott RN. UNB test of prosthetics function: a test for unilateral upper extremity amputees, ages 2-13. Fredericton (NB): Univ New Brunswick, Bio-engineering Institute; 1985.
21. Codd RD. Development and evaluation of adaptive control for a hand prosthesis [dissertation]. Southampton (UK): Univ Southampton; 1975.
22. Bergman K, Ornholmer L. Functional benefit of an adaptive myoelectric prosthetic hand compared to a conventional myo hand. *Prosthet Orthot Int* 1992;16:32-7.
23. O'Neill G. The development of a standardised assessment of hand function. *Br J Occup Ther* 1995;58:477-80.

24. Desrosiers J, Hébert R, Dutil E. Upper Extremity Performance Test for the Elderly (TEMPE): normative data and correlates with sensorimotor parameters. *Arch Phys Med Rehabil* 1995;76:1125-9.
25. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *Int J Rehabil Res* 1981;4:483-92.
26. Agnew PJ. Functional effectiveness of a myoelectric prosthesis compared with a functional split-hook prosthesis: a single subject experiment. *Prosthet Orthot Int* 1981;5:92-6.
27. van Lunteren A, van Lunteren-Gerritsen GH, Stassen HG, Zuithoff MJ. A field evaluation of arm prostheses for unilateral amputees. *Prosthet Orthot Int* 1983;7:141-51.
28. Kay HW, Kajganic M, Ivancevic N. Medical evaluation of the Belgrade electronic hand. In: *Proceedings of the 3rd Symposium on Advances in External Control of Human Extremities*; 1969 Aug; Dubrovnik (Yugoslavia). p 129-37.
29. Mendez MA. Evaluation of a myoelectric hand prosthesis for children with a below-elbow absence. *Prosthet Orthot Int* 1985;9:137-40.
30. Wing AM, Turton A, Fraser C. Grasp size and accuracy of approach in reaching. *J Mot Behav* 1986;18:245-60.
31. Fraser CM. An evaluation of the use made of cosmetic and functional prostheses by unilateral upper limb amputees. *Prosthet Orthot Int* 1998;22:216-23.
32. Kurtz I, Heim W, Bauer-Hume H, Hubbard S, Ramdial S. Programmable control: technical aspects. In: *Proceedings of the Myoelectric Controls Symposium*; 1999 Aug; Fredericton (NB). p 130-3.
33. Wing AM, Fraser C. The contribution of the thumb to reaching movements. *Q J Exp Psychol A* 1983;35 Pt 2:297-309.
34. McWilliam RP. A list of everyday tasks for use in prosthesis design and development. *Bull Prosthet Res* 1970;10(13):135-64.
35. Sollerman C, Ejeskar A. Sollerman hand function test. A standardised method and its use in tetraplegic patients. *Scand J Plast Reconstr Surg Hand Surg* 1995;29:167-76.
36. Krebs D. Prehension assessment: prosthetic therapy for the upper limb child amputee. Thorofare (NJ): Slack; 1987.
37. Stein RB, Walley M. Functional comparison of upper extremity amputees using myoelectric and conventional prostheses. *Arch Phys Med Rehabil* 1983;64:243-8.
38. Potvin AR, Tourtellotte WW, Daily JS, et al. Simulated activities of daily living examination. *Arch Phys Med Rehabil* 1972;53:476-86.
39. Gnanadesikan R. *Methods for statistical analysis of multivariate observations*. New York: Wiley; 1997.
40. Eddy DM. *A manual for assessing health practices and designing practice policies*. Philadelphia: American College of Physicians; 1992.
41. Light CM. *An intelligent hand prosthesis and evaluation of pathological and prosthetic hand function [dissertation]*. Southampton (UK): Univ Southampton; 2000. p 179-91.
42. Jones LA. The assessment of hand function: a critical review of techniques. *J Hand Surg [Am]* 1989;14(2 Pt 1):221-8.
43. McPhee, SD. Functional hand evaluations: a review. *Am J Occup Ther* 1987;41:158-63.
44. Moore DS, McCabe GP. *Introduction to the practice of statistics*. New York: WH Freeman; 1989.
45. Armitage P. *Statistical methods in medical research*. New York: Wiley; 1971.
46. Abelson RP. *Statistics as principled argument*. Hillsdale (NJ): Lawrence Erlbaum Associates; 1995.
47. Mathiowetz V, Volland G, Kashman N. Adult norms for the box and block test of manual dexterity. *Am J Occup Ther* 1985;39:386-91.
48. Morrison DF. *Multivariate statistical methods*. New York: McGraw-Hill; 1990.