HAND CLINICS

Hand Clin 19 (2003) 371–378

Outcome evaluation of the hand and wrist according to the International Classification of Functioning, Disability, and Health

Olivier Barbier, MD[a,*], Massimo Penta, PhD[b], Jean-Louis Thonnard, PhD[b]

[a]*Department of Orthopaedic Surgery, University of Louvain and St-Luc Hospital, Avenue Hippocrate, 10, 1200 Brussels, Belgium*

[b]*Rehabilitation and Physical Medicine Unit, University of Louvain and St-Luc Hospital, Avenue Hippocrate, 10, 1200 Brussels, Belgium*

Most hand surgeons are concerned about choosing instruments and scales to evaluate the outcome of their clinical practice. Several tools currently exist to quantify outcome in hand surgery at the impairment level (eg, mobility, hand strength, cutaneous sensation, dexterity). According to the World Health Organization's paradigm [1], however, activity limitations and participation restrictions are also clinically relevant. Because these domains cannot be inferred from the underlying impairments, they must be measured with specific and appropriate scales. The authors have recently built a measure of an upper limb-impaired individual's ability to manage manual activities in daily life. This manual ability test, ABILHAND, was developed using the Rasch measurement model, which allows ordinal scores to be converted into linear measures located on a unidimensional scale [2]. Participation and quality of life are difficult outcomes to measure because they are multidimensional and depend on such factors as functional abilities, general physical health, financial security, and stability of the social and familial environment.

International Classification of Functioning, Disability, and Health

The International Classification of Functioning, Disability, and Health, known as ICF [1], envisions the relations between a disease and three different levels of disablement: body functions, activities, and participation (Fig. 1).

Body functions are the physiologic or psychologic functions of body systems (eg, mental, sensory, neuromusculoskeletal, and movement-related functions). Body structures are anatomic parts of the body such as organs, limbs, and their components. Impairments generally refer to deviations in body function or structure from certain generally accepted population standards (eg, control normative data or the unimpaired limb in unilateral disorders). Once an impairment is localized, it can be scaled in severity according to its extent and magnitude.

An activity is the execution of a task or action by an individual. This domain deals with all aspects of daily life, envisioning human activities as the purposeful, integrated use of body functions (eg, activities of moving around and self-care activities, and communication, domestic, or interpersonal activities). Activity limitations are difficulties an individual may have in the performance of these daily activities, whatever the extent and magnitude of the underlying impairments. Hence, contextual factors such as the use of assisting devices or another person's help does

* Corresponding author.

E-mail address: Olivier.Barbier@orto.ucl.ac.be (O. Barbier).

0749-0712/03/$ - see front matter
doi:10.1016/S0749-0712(02)00150-6

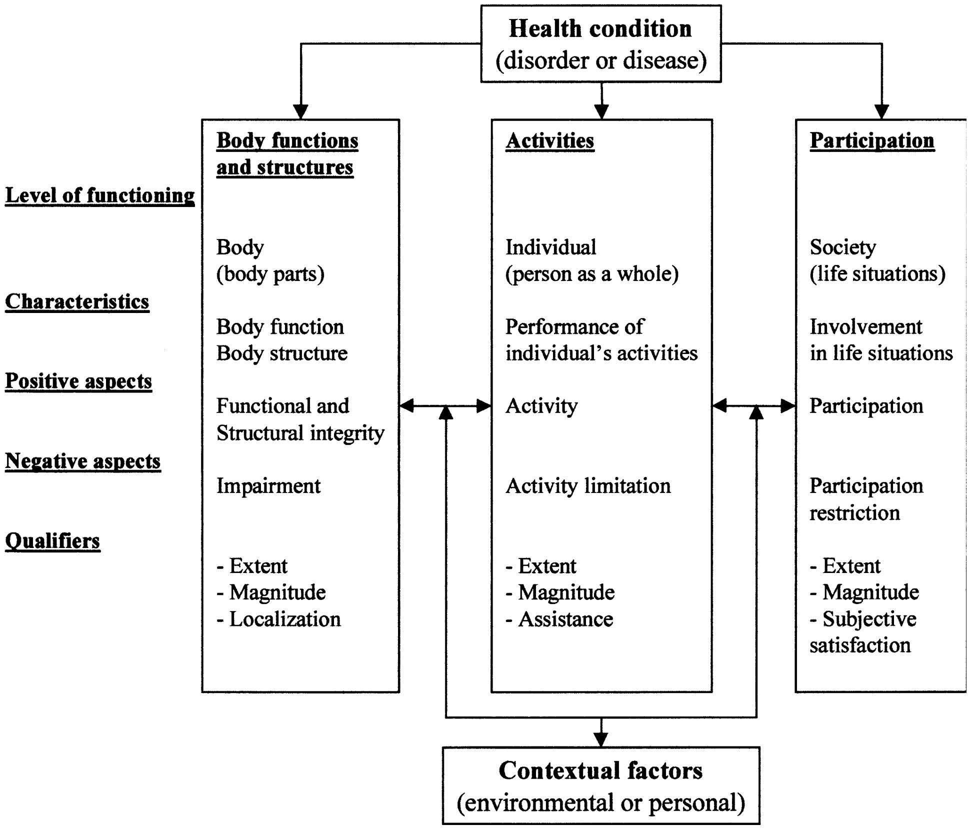

Fig. 1. Overview of dimensions of the ICF (modified from WHO [1]).

not eliminate the impairment, but might reduce limitations on activities in specific domains. Similarly, limitations sometimes can be overcome by executing activities in alternate manners, for example, substituting touch for vision in Braille reading or using two hands for an activity that usually requires just one.

Participation is defined as an individual's involvement in life situations in relation to health conditions, body functions and structures, activities, and contextual factors. It refers to the experience of people in the actual context in which they live (eg, participation in personal maintenance, home life, mobility, exchange of information, social relationships, employment, and economic life) and also includes society's response to the individual's level of functioning. Clearly this domain is not restricted simply to upper limb function or manual activities.

The relationship between the three domains is influenced by contextual factors representing the complete background of an individual's life, including environmental and personal factors. Environmental factors make up the physical, social, and attitudinal environment that may interfere with the individual's health condition or functional states. They are classified as (1) individual factors, the immediate personal environment, and direct personal contacts (eg, products for personal use in daily life, immediate family, colleagues, and care providers), (2) social structures and services (eg, health services, social security, housing), and (3) systems factors (eg, health, social security and employment systems, and policies). On the other hand, personal factors are the background of an individual's life that are not part of a health condition, such as age, race, gender, educational background, personality, specific aptitudes, lifestyle, and other personal features.

Evaluation of impairment

Impairment is the direct neurophysiologic consequences of an underlying pathology. Examples of common impairments for the hand and wrist are a reduction in range of motion (ROM) at a joint, weakness, decrease in cutaneous sensitivity, and loss of dexterity. Each of these domains can be explored in an upper limb test battery according to the following procedure.

Range of motion

Active range of motion is the maximal motion obtained at the joints without help. Lack of motivation and protective reflexes caused by pain may limit the active ROM scores. Passive ROM measurements require the examiner to forcibly mobilize the joint to overcome soft tissue resistance. The level of force applied, however, can alter the measurement considerably [3]. Published guidelines specify the starting position of the joint being evaluated, the correct placement of the goniometer relative to the joint, and the steps to follow for a valid measurement of active and passive ROM [4].

Boone et al [5] established that a single set of ROM measurements is as reliable as averaging several sets. Successive measurements of a joint should be done by the same examiner because intra-tester reliability is greater than inter-tester reliability [5]. Given these constraints, measurements of the ROM are accepted as valid, accurate, reliable, and interpretable with reference to published norms [6].

Muscle strength testing

Manual testing and isometric dynamometers are the most frequently used methods for assessing muscle strength and endurance at the hand and wrist. Manual testing is only useful for evaluating gross muscle deficits; muscle strength is reported using the Medical Research Council's 0–5 scale.

A dynamometer is used when more precise measurements of muscle strength are required. Inter- and intra-patient muscle strength measurements can be best compared if the standard procedures recommended by Mathiowetz [7] are followed: (1) grip strength is measured with the standard Jamar dynamometer, (2) the dynamometer is calibrated periodically to ensure accuracy, (3) the same dynamometer is used in pre- and post-testing of patients, (4) the body is in a standard position for each hand strength test, (5) the average of three trials is used as the patient's score, and (6) normative data are used with consideration of the patient's age and sex.

Quantitative sensory evaluation

Although a wide variety of clinical measures of sensory performance have been described for use in hand surgery, this remains an area in need of further scientific investigation. The most commonly used tests that reflect the extent of reinnervation measure pressure or vibration perception, although other recommended tests measure spatial discrimination, temperature sensation, or object shape recognition.

The Semmes-Weinstein monofilament test [8] is used to assess cutaneous touch–pressure thresholds. This test consists of 20 nylon monofilaments of different stiffness. The monofilaments are applied perpendicularly to the skin and exert a pressure ranging from 0.0045–448 g when slightly flexed [8]. Visual clues must be kept to a minimum and the hand should be supported to avoid the stimulus of proprioception.

Vibration perception thresholds are determined with a vibrometer and are presented as a function of frequency (a vibrogram) or as a sensitivity index (SI) as proposed by Lundborg et al [9].

The two-point spatial discrimination test has been widely used in the past. The use of this test remains controversial, however, because of its lack of reproducibility [10,11] and sensitivity to spurious nonspatial cues [12]. Consequently the Grating Orientation Task (GOT) seems to be a more adequate test to assess tactile spatial resolution [12,13]. This test is composed of plastic dome gratings with equidistant bar and groove widths measuring 0.35, 0.50, 0.75, 1.00, 1.25, 1.50, 2.00, and 3.00 mm (JVP Domes, Stoelting Co., Wood Dale, IL). For each trial, a grating is applied for approximately 1.5 seconds with a comfortable force of 0.65–0.95 N (corresponding to a displacement of approximately 1.0–1.5 mm at the skin) perpendicularly to the surface of the distal pad [13]. The bars and grooves are aligned in one of two orthogonal directions (ie, along or perpendicular to the long axis of the finger). Patients are required to identify the stimulus orientation (two-alternative forced choice paradigm) verbally before the stimulus is removed.

Cold and warm perception thresholds can be measured according to the methodology proposed by Yarnitsky and Sprecher [14]. In this study the authors present normative data for reaction-time–inclusive and –exclusive measurement algorithms.

The Moberg pickup test [15] and tests for object recognition, shape, or dimension identification also can be useful in assessing functional sensitivity. Nevertheless because of the lack of standardization of these testing techniques, one must be careful in interpreting their scores.

Dexterity

Dexterity must be evaluated because of its bearing on upper limb performance and on individual functional independence. Dexterity has been defined by Poirier et al [16] as "a manual skill requiring rapid coordination of fine and gross movements based on a certain number of capacities developed through learning, training, and experience." Speed and precision are the criteria used to measure this skill. There are two main types of dexterity: finger dexterity and manual dexterity.

Finger dexterity is defined as the ability to make rapid, skillful, controlled, manipulative movements of small objects in which the fingers are primarily involved [17]. The well known Purdue Pegboard Tests [18] of finger dexterity are composed of a board containing two rows of 25 holes. Cups containing pins, collars, and washers are placed at the top of the board. The test consists of four subtests. In the first three, the subject is required to take as many pins as possible within a 30-second period out of a cup and place each one into a hole in the board: first with the preferred hand, then with the other hand, and finally with both hands simultaneously. In the last subtest the subject uses alternate hands to assemble a pin, collar, and two washers in a hole in the board as many times as possible within a 60-second period. Several reliability and validity studies of the Purdue Pegboard Test have been done with children and adults, and normal values are available for all subtests [17–19]. The test–retest reliability coefficient for three trials varies from 0.82–0.91, according to the subtest.

Manual dexterity is defined as the ability to make skillful, controlled, arm–hand manipulations of larger objects under speed conditions [17]. The Box and Block Test [20] measures unilateral gross manual dexterity. It consists of moving, one by one, the maximum number of blocks from one compartment of a box to another of equal size within 60 seconds. A 6-month test–retest reliability study was done ($p = 0.98$ for the right hand and 0.94 for the left), and concomitant validity was measured with the Minnesota Rate Manipulation Test with the result $r = 0.91$ [21]. Mathiowetz et al [20] have standardized the measurement procedure for this test, studied its inter-rater reliability ($p = 1$ for both hands), and developed norms for adults.

ABILHAND: A Rasch-built measure of manual ability

Evaluating the execution of everyday life activities with the hand and wrist requires a test that is specific to these types of activities [22]. Moreover, the relationship between impairments and activity limitations is not straightforward. A patient may adopt either intrinsic or adaptive recovery mechanisms, the latter depending on the integrity of the unaffected organs/segments and on a complex interaction between psychosocial (eg, motivation), cognitive (eg, memory, attention, space perception), and sensorimotor skills. Manual ability may be defined as the capacity to manage daily activities requiring the use of the upper limbs, whatever the strategies involved. The authors have built the first measure specifically focused on manual ability and calibrated the scale in a sample of rheumatoid arthritic patients [22] and in a sample of chronic stroke patients [23].

Manual ability belongs to the domain of latent variables concealed within the person, such as pain, depression, intelligence, and the like. The "amount" of manual ability can be inferred from observed activities or a patient's perceived difficulty in performing activities as determined by questionnaires. Questionnaires, however, provide raw ordinal scores that might be misused as measures, when in fact they are merely ranks unsuited to conventional arithmetic [24]. To compare inter- or intra-patient manual ability, it must be expressed on a single unidimensional scale. A linear measure of manual ability can be properly estimated only from raw scores according to measurement models [25], the most promising being the Rasch model [2]. This model formulates the criteria for invariant comparisons between patient measures on a disability scale defined by a set of items conformed to the scale.

Instrument

The 56 items composing the ABILHAND questionnaire include unimanual and bimanual activities that explore a wide variety of manual activities. Some items were selected from existing scales, whereas other items were devised to extend the range of activities explored by the questionnaire. To have the greatest resolution in the measure of manual ability, the authors selected items for which manual skills are vital for successful performance. The questionnaire has been translated into French, English, Italian, and Swedish.

Procedures

The test is administrated as a self-reported anamnestic questionnaire. Patients are asked to rate the ease/difficulty in performing each activity, without any help, on the following three-level rating scale: impossible (0), difficult (1), and easy (2). Activities not attempted in the last 3 months are not scored and are entered as missing responses. Patients are asked to use the whole scale range to respond, and five training items are presented before the actual test. The 56 items are randomly presented to the patient. For unimanual and bimanual activities, patients are asked to give their response irrespective of the limbs they would actually use to perform the task.

ABILHAND calibration

The questionnaire was applied in a sample of rheumatoid arthritic (RA) patients [22] and in a sample of chronic stroke patients (CS) [23]. In the RA study, 46 items including uni- and bimanual activities were found to define a unidimensional manual ability scale. These results confirmed that the Rasch methodology was successful in producing a useful scale of manual ability. In the CS patient study, a larger sample of patients showed that unimanual activities (usually realized with one hand) were too easy for the patients. So a subset of 23 bimanual activities (usually realized with two hands) seemed to be more discriminating in CS patients. The manual ability measures in CS patients were significantly correlated with grip strength, motricity, dexterity, and depression.

The definition and use of the ABILHAND manual ability measure in a chronic stroke patient is illustrated in Fig. 2. Each line in the item map presents the most probable response to an item as a function of the patient's manual ability, with colons indicating expected half-score points between each successive levels of a response. Activities are listed in decreasing order of difficulty; note that the observed response (circles in Fig. 2) decreases with the item difficulty. The patient's total score is obtained by summing the individual item scores. The S-shaped score versus measure relationship presented in the bottom of Fig. 2 allows the total score to be converted into linear logit units. The logit is a probabilistic unit that expresses the odds ratio of success to failure on any item. The scale is centered on the average item difficulty; a 1-logit difference between two patients indicates that their odds of successful achievement of any activity are 2.7:1 ($e^{1/1}$), 2 logits results in a 7.4:1 odds. The score versus measure relationship also provides the 95% confidence interval of the manual ability measure.

Once the manual ability measure is obtained, it is necessary to verify the response pattern coherence. Unexpected responses are identified by comparing the observed patient's response with an item with the response expected to the item given the overall patient's manual ability measure. Unexpected responses are those in which the observed response lies outside the 95% confidence interval for the manual ability measure of the patient. Those activities were perceived as too easy or too difficult by the patient, given the patient's overall manual ability measure. For instance, the patient shown in Fig. 2 overestimated the easiness of "Tearing open a pack of chips," given his overall manual ability.

The invariant hierarchy of item difficulties across demographic and clinical subgroups in CS patients further supports the clinical applicability of the ABILHAND scale [23]. Despite the fact the patients presented different levels of impairments in grip strength, tactile sensitivity, and manual dexterity, the resulting hierarchy of item difficulty was invariant. This indicates that the manual ability definition provided by ABILHAND can be used equally to measure patients with various types of impairment and opens the way for further comparisons across diagnoses.

Participation and health-related quality of life

Participation restrictions are difficult to measure because there is no absolute reference against which to judge them [1]. Participation restrictions are judged with reference to the expectations of the specific individual patient, with some reference to the expectations of the limited group of people occupying the same cultural, social, economic, and physical environment as the subject. Participation restrictions also can be considered as referring to the change in a patient's quality of life. Health-related quality of life (HRQL) is the multivectorial term that embraces physical, functional, emotional, social, financial, and spiritual factors that affect quality of life. Because of the complexity of defining quality of life, there is no uniformity of opinion regarding an optimal evaluation

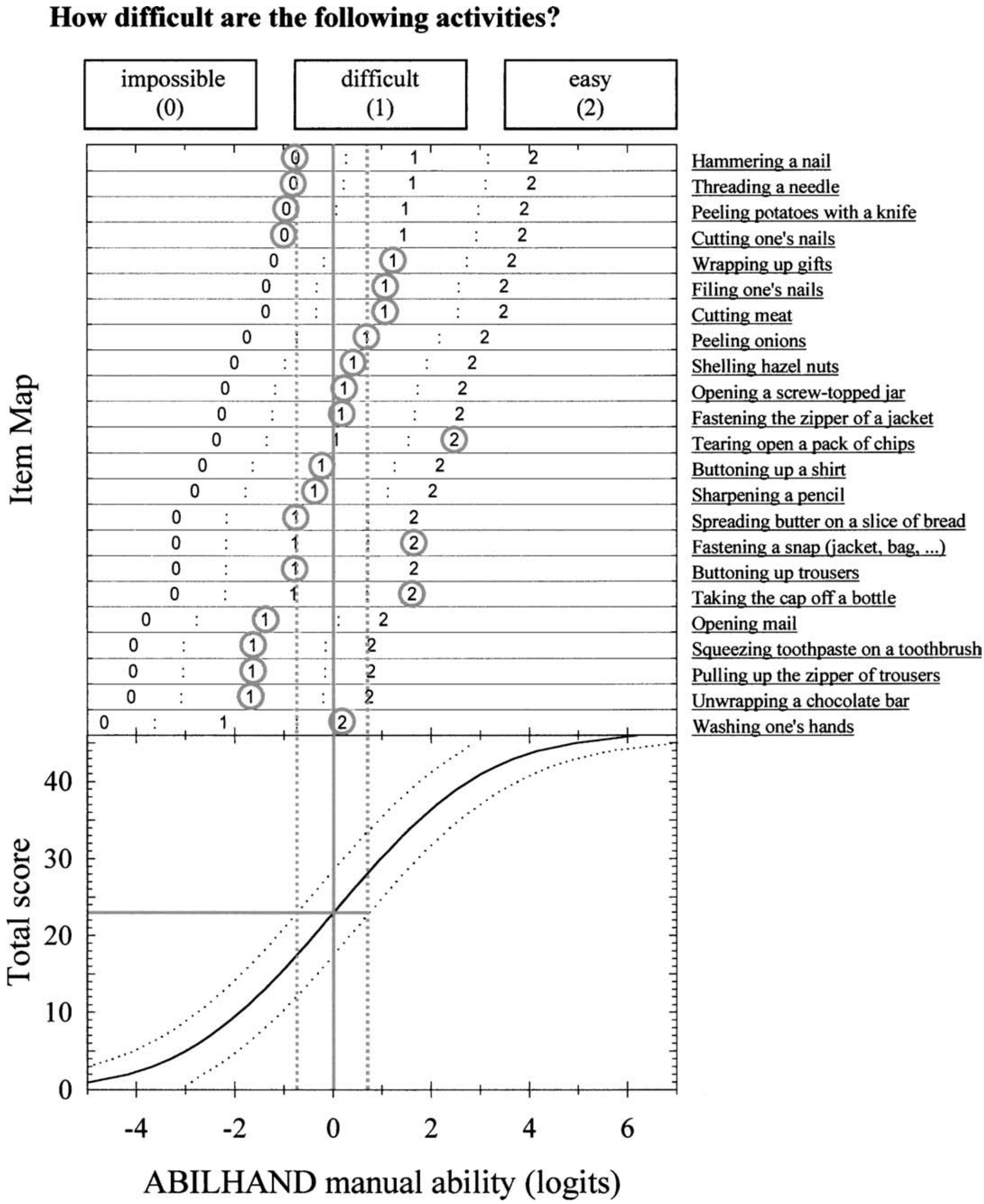

Fig. 2. A sample scoring form showing the structure of the ABILHAND measure by way of the item map (top panel) and the relationship between raw scores and manual ability measure (bottom panel). The item map provides a patient's expected score to each item (0 = impossible; 1 = any difficulty; 2 = easy) as a function of the patient's ability. For each item, the placement of the numeric labels indicates the manual ability required for a given expected score; the colons indicate expected half-score points. Circled labels indicate the answer provided by the CS patient to each question. The S-shaped score versus measure relationship (solid line) and its 95% confidence interval (dotted lines) demonstrate the nonlinearity of total raw scores, especially at the extremes of the scoring range. The figure allows scoring the test, locating the patient's measure (solid gray line) according to the 95% confidence interval (dotted gray lines), and analyzing the response coherence.

instrument [26]. In selecting a suitable measure, there is a tradeoff between the level of detail provided and ease of use in patient and staff burden.

One of the most widely used instruments to measure HRQL is the SF-36, a general health status measure that generates a profile of eight scales and summarizes physical and mental health measures [27]. The 36 specific questions are aggregated to provide scores on scales called physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health. The eight subscales are summarized into a mental and a physical summary measure. The first three scales contribute substantially to the overall physical health summary measure. The last three correlate most highly with the mental health summary measure. Vitality, general health, and social functioning correlate with both summary measures. The SF-36 can be administered by the subjects themselves or by an interviewer in person or over the phone and generally takes 5–10 minutes. It has been adapted to many cultures and languages and has undergone extensive psychometric testing. Perhaps the strongest criticism that can be leveled against the SF-36 is that, whereas it covers eight concepts generally well accepted as important to HRQL, it omits several others that include sleep adequacy, cognitive function, sexual function, health distress, family function, self-esteem, eating, recreation, and communication.

References

[1] World Health Organization. International classification of functioning, disability and health (ICF). Geneva: WHO; 2001.
[2] Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago, IL: Mesa Press; 1980.
[3] Amis AA, Miller JH. The elbow—measurement of joint movement. Clin Rheum Dis 1982;8:571–93.
[4] Norkin CC, White DJ. Measurement of joint motion: a guide to goniometry. 6th edition. Philadelphia: The FA Davis Co.; 1988.
[5] Boone DC, Azen SP, Lim CM. Reliability of goniometric measurements. Phys Ther 1978;58: 1355–90.
[6] American Academy of Orthopedic Surgeons. Joint motion: method of measuring and recording. Chicago, IL: The Academy; 1965.
[7] Mathiowetz V. Grip and pinch strength measurements. In: Amundsen LR, editor. Muscle strength testing. Instrumented and non-instrumented systems. 1st edition. New York: Churchill Livingstone Inc.; 1990. p. 163–77.
[8] Bell-Krotoski JA. Light touch-deep pressure testing using Semmes-Weinstein monofilaments. In: Hunter JM, Schneider LH, Mackin EJ, Callahan AD. Rehabilitation of the hand: surgery and therapy. 3rd edition. St. Louis: The CV Mosby Co.; 1990.
[9] Lundborg G, Dahlin LB, Lundström R, Necking LE, Strömberg T. Vibrotactile function of the hand in compression and vibration-induced neuropathy. Sensibility index: a new measure. Scand J Plast Reconstr Surg 1992;26(3):275–81.
[10] Szabo RM, Gelberman RH, Dimick MP. Sensibility testing in patients with carpal tunnel syndrome. J Bone Joint Surg 1984;66A:60–4.
[11] Bell-Krotoski JA, Buford WJ. The force/time relationship of clinically used sensory testing instrument. J Hand Ther 1988;1:76–85.
[12] Johnson KO, Van Boven RW, Hsiao SS. The perception of two points is not the spatial resolution threshold. In: Boivie J, Hansson P, Lindblomu, editors. Touch, temperature, and pain in health and disease: mechanisms and assessments. Progress in pain research and management. Vol. 3. Seattle: IASP Press; 1994. p. 389–404.
[13] Van Boven RW, Johnson KO. The limit of tactile spatial resolution in humans: grating orientation discrimination at the lip, tongue, and finger. Neurol 1994;44:2361–6.
[14] Yarnitsky D, Sprecher E. Thermal testing: normative data and repeatability for various test algorithms. J Neurol Sci 1994;125:39–45.
[15] Moberg E. Objective methods for determining the functional value of the sensibility of the hand. J Bone Surg 1958;40B:454–76.
[16] Poirier F. Dexterity as a valid measure of hand function: a pilot study. Occup Ther Health Care 1987;4:69–83.
[17] Fleishman EA, Ellison GD. A factor analysis of fine manipulative tests. J Appl Psychol 1962;46:96–105.
[18] Tiffin J, Asher EJ. The Purdue pegboard: norms and studies of reliability and validity. J Appl Psychol 1948;32:234–47.
[19] Reddon JR, Gill DM, Gank SE, Maerz MD. Purdue pegboard: test-retest estimates. Percept Mot Skills 1988;66:503–6.
[20] Mathiowetz V, Volland G, Kasman N, Weber K. Adult norms for the box and block test of manual dexterity. Am J Occup Ther 1985;39:386–91.
[21] Cromwell FS. Occupational therapists manual for basic skills assessment: primary prevocational evaluation. Pasadena, CA: Fair Oaks Printing; 1965. p. 29–31.
[22] Penta M, Thonnard J-L, Tesio L. ABILHAND: a Rasch-built measure of manual ability. Arch Phys Med Rehabil 1998;79:1038–42.
[23] Penta M, Tesio L, Arnould C, Zancan A, Thonnard J-L. The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients. Rasch-based validation and relationship to upper limb impairment. Stroke 2001;32:1627–34.

[24] Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Arch Phys Med Rehabil 1989;70:308–12.
[25] Wright BD, Linacre JM. Observations are always ordinal: measurement, however, must be interval. Arch Phys Med Rehabil 1989;70:857–60.
[26] Wade DT. Measurement in neurological rehabilitation. New York: Oxford University Press Inc.; 1992.
[27] Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston: Health Institute, New England Medical Center; 1993.