EURA MEDICOPHYS 2007;43:525-41

Functional assessment in physiotherapy

A literature review

J.-L. THONNARD, M. PENTA

Physical Medecine and Rehabilitation Unit
Catholic University of Louvain, Belgium

The present literature review on functional assessment in physiotherapy was carried out for the following reasons: 1) to identify the functional instruments used in the field of physiotherapy that were supported by published evidence of their psychometric qualities; 2) to investigate how these instruments relate to the International Classification of Functioning, Disability and Health (ICF); and 3) to investigate the use of functional instruments in the financing of physiotherapy. A search of Medline from 1990 to December 2005, in the domains of functional evaluation, psychometric qualities, functional classification, and health policy in relation to physiotherapy resulted in a list of 1,567 studies. Two reviewers examined the resulting references on the basis of their title and abstract, in order to select the studies that presented data on the psychometric qualities of functional evaluation tests, leading to a final selection of 44 such studies. A selection of functional tests was identified in four major diagnostic groups treated in community physiotherapy: musculoskeletal disorders (including lower back pain), stroke, the elderly, and traumatic brain injuries. The functional tests authors identified essentially cover the body and activities dimension of the ICF. The selected tests could be used as a basis for the standardisation of functional evaluation of the major diagnostic groups treated in community physiotherapy. This means that standards are available for reporting and following the evolution of patients both longitudinally and transversally. Nevertheless, in the current literature review no attempt at using functional outcomes as a rationale for financing physiotherapy could be found to date.

KEY WORDS: Physical therapy modalities - Rehabilitation - Physiotherapy.

Physiotherapy is based on an evaluation of the patient's functional health. This evaluation supplements the medical diagnosis, which is primarily concerned with pathology. For example, in addition to the medical diagnosis of osteoarthritis, the physical therapist needs information on pain, joint range of motion, muscle strength, etc. This information is used to select appropriate interventions, follow the patient's recovery and assess treatment outcome.

The International Classification of Functioning, Disability and Health (ICF, World Health Organisation 2001) has been developed to offer an excellent conceptual framework for envisioning the consequences of health condition or pathology on the function of individuals. It establishes a common language for describing the consequences of health that facilitates the comparison of data across countries, health care disciplines, services and time. This extensive system allows up to 1 424 ICF codes to be scored to describe an individual's functional health Three separate but related dimensions of functioning are defined: body functions and structures (body dimension) activity

Funding.—This study was funded by a grant from Centre Fédéral d'Expertise des Soins de Santé in Belgium.

Address reprint requests to: J.-L. Thonnard, Université Catholique de Louvain, Unité de Réadaptation, Tour Pasteur (5375), Avenue Mounier, 53, BE-1200 Bruxelles, Belgium.
E-mail: jean-louis.thonnard@uclouvain.be

(individual dimension), and participation (social dimension). 1) Body functions are the physiologic or psychologic functions of body systems (*e.g.*, mental, sensory, neuromusculoskeletal and movement related functions). Body structures are anatomical parts of the body such as organs, limbs and their components. Impairments refer to anomaly, defect, loss or other significant deviation in body functions and structures. 2) An activity is the execution of a task or action by the patient. This dimension deals with all aspects of daily life, envisioning human activities as the purposeful, integrated use of body functions (*e.g.*, activities related to moving about and self-care, and communication, domestic, or interpersonal activities). Activity limitations are difficulties the patient 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, alternative strategies or another person's help do not eliminate the impairment, but might reduce limitations on activities in specific domains. 3) Participation is defined as the patient's involvement in life situations in relation to health conditions, body functions and structures, activities and contextual factors. It refers to the experience of the patient in the actual context of life (*e.g.*, social relationships, employment and economic life) and it also includes society's response to the patient's level of functioning. Participation restrictions are the problems the patient may experience in their involvement in life situations.

Rehabilitation can be defined as a reiterative problem solving and educational process that focuses on disability (altered activities) and aims to maximise participation in society while minimising the stress on and distress of the patient and family.[1] "Community," is used to refer to the physical and social environment of the patient in the place where she or he lives. It is usually set in contrast to a hospital, or another institutional setting where the patient may go or reside for short periods while receiving a service. The community might include a nursing home or residential care home if it is the patient's long-term residence, but would not normally include a hospital or short-term stays in a nursing home. The economico-political profile of rehabilitation in the community has increased in the last decade. Three factors have contributed to this rise in interest: 1) the increasing awareness of disability; 2) the increasing evidence that active rehabilitation has a beneficial effect in terms of reducing dependence and care costs, and in terms of improving the patient's quality of life; and 3) the increasing pressure to shorten the time patients spend in hospital beds. Politicians and health-care funding agencies have seen an opportunity to control healthcare expenditure while also appearing to improve rehabilitation services to disabled people. They have started to champion community rehabilitation in the hope that this will contain or reduce expenditure by shortening the time patients spend in the hospital while also improving patient outcomes.

While the outcome of physiotherapy can be assessed with functional tests, the clinical purpose of measurement is an important issue that should be considered. For instance, a diagnostic test should evaluate patient functioning in relation to the diagnosis in question, and variation across repeated measurements should be minimized. In order to monitor the progress of patients, the test should provide reproducible results over time, but also a sufficient range of measurement with enough sensitivity to detect meaningful changes. Psychometric qualities of measures are encompassed under the umbrella terms of validity, reliability (including sensitivity and specificity) and responsiveness [2] to determine if they are meaningful for diagnosis or for the evaluation of progress.

Validity describes how well a functional test measures what it purports to measure. Ideally, validity is assessed against a gold standard, even though such a standard is not always available for tests measuring variables that were not previously defined. While there are many ways to address the issue of validity, it is generally appraised in terms of the content of the test or by examining the behaviour of the measurement scale in regard to the underlying theoretical construct. Reliability refers to the reproducibility of the measure, and is therefore dependent upon the amount of measurement error associated with the measure. A reliable test has items that are internally consistent (*i.e.,* related to one another) and minimises measurement errors in order to provide the same results (within confidence limits) regardless of the particular circumstances in which the measure was obtained (*e.g.,*rater, time of evaluation). Finally, measures intended to measure a patient's recovery or lack of regression should be sensitive to change (have the ability to detect change given the range of measurement and the measurement error), as well as show responsiveness (the ability to detect meaningful change).

The purpose of this literature review on function-

al assessment in physiotherapy is: 1) to identify the functional instruments used in the field of physiotherapy that are supported by published evidence of their psychometric qualities; 2) to investigate how these instruments relate to the ICF; and 3) to investigate the use of functional instruments in the financing of physiotherapy.

Methods

A literature search of Medline was undertaken between October and December 2005 in order to identify functional evaluation methods of potential utility for physiotherapy practice within the Belgian health care system. The literature search focused on four specific areas: 1) the identification of functional tests; 2) psychometric qualities of the tests; 3) the ICF dimensions covered by a functional test; and 4) the use of functional tests in health care policy. The following keywords were preliminarily identified in order to cover the scope of the study: physiotherapy, functional, evaluation, assessment, outcome, disability, impairment, handicap, participation, quality of life, life satisfaction, validation, reliability, psychometrics, health care financing, ICF and ICIDH. The keywords were then entered into the Medical Subject Heading (MeSH) in order to find their most relevant counterpart. The following MeSH terms were retained: physiotherapy, activities of daily living, disability evaluation, outcome assessment, quality of life, evaluation studies, psychometrics, reproducibility of results, validation studies, ICF, delivery of health care, economics, health planning guidelines, health policy, health services research, health services, national health programs and program evaluation. Each MeSH term was then combined with the keyword physiotherapy and with a fixed list of qualifiers (valid*, reliab*, reproducib*, repeatab*, responsiv*, sensitiv*, specificity, psychometr*) representing psychometric qualities, in order to limit the search to a manageable number of references. The complete literature search resulted in 1,567 studies matching these criteria. No systematic review was found: only papers reporting primary research were obtained in conducting this literature review.

Two reviewers examined the resulting references and selected the studies that presented data on the psychometric qualities of functional evaluation tests used in physiotherapy. The references were selected on the basis of their title and the abstract was consulted when necessary. Studies with a title or abstract including discussion of the psychometric aspects of functional tests or other data relating to functional assessment in community-based physiotherapy were included. If studies were published earlier than 1990, were not written in English, were specific to hospital rehabilitation, required the use of sophisticated devices (*e.g.,*goniometer, dynamometer), related to psychiatric or scar care or focused on country-specific adaptations of previously validated scales, they were excluded. This process allowed 60 references to be selected for the review.

Two independent readers reviewed the selected list of references in order to collect data from each study and encoded them in a standard way. The readers performed a preliminary assessment of 10 studies together in order to standardise the data collection. Functional tests were reviewed systematically in order to describe the clinical utility of each test, identify the ICF dimension covered by the test and extract the published data on the test's psychometric qualities, including its validity, reliability and responsiveness.

Results

Four types of studies were identified among the 60 references reviewed. Forty-four studies presented psychometric qualities of functional tests used in physiotherapy. Fifteen studies discussed clinical, methodological or organisational issues, not directly presenting evidence of functional test qualities. One study discussed issues related to patient satisfaction with physiotherapy services. The latter 16 studies were not formally used to review the psychometric qualities of functional tests (*e.g.,*clinical utility, metric properties); instead they were used to develop the arguments presented in the introduction and in the general discussion of the study.

Each of the 44 studies investigating functional tests had addressed the metric qualities of one or multiple tests. The tests were evaluated for validity, reliability and/or for responsiveness in the following diagnostic groups: patients with musculoskeletal disorders (N=24) including lower back pain, spine disorders, chronic pain, lower limb amputees, lower limb dysfunction and ankilosing spondylitis; stroke patients (N=16); elderly patients (N=4) and patients with brain injury (N=3). Evidence of clinical utility, ICF dimension cov-

ered and metric qualities of functional tests are described as they were found in the literature.

In the papers we selected, content validity was assessed using either experts' or patients' advice or through a literature review. Concurrent and construct validity was generally assessed by correlation analyses. The internal consistency was assessed with Chronbach's alpha.[2] A higher diversity of methods were used to assess test/re-test reliability: paired t-test, correlation analysis, percentage of agreement, intraclass correlation coefficient,[3] weighted or unweighted Cohen's kappa coefficient [4, 5] and Kendall's coefficient of concordance.[6] Inter-rater reliability was generally assessed by comparing blinded rater evaluations of videotaped patient performance. Multiple rater assessments were compared with the same methods used to address test/re-test reliability. Responsiveness was generally assessed with analysis of variance (with or without repeated measures), percentage change score, effect size,[7] standardised response mean (Liang et al 1990) or with the responsiveness index.[8]

A summary of the functional tests used in musculoskeletal disorders is presented in Table I. Among the 24 studies reviewed, 16 investigated patients with lower back pain, two with spine disorders non-specific to the lower back, two with chronic pain non- specific to the lower back, two with lower extremity dysfunction, one with lower limb amputation and one with ankylosing spondylitis. Three main types of functional tests were identified. Patient classification based on anatomo-pathological or treatment criteria were used by the physical therapist mostly for treatment planning. Clinical tests carried out by the physical therapist as part of the clinical examination (N=8) were used essentially to evaluate the body structures and functions. Patient-reported questionnaires, either self-administered or interview-based, were used essentially to evaluate the activities of the patient, though some of them also address the body or participation dimension.

The evidence table of the functional tests used in stroke patients is presented in Table II. A total of 16 functional tests were reviewed. Among them, 10 clinical tests carried out by the physical therapist addressed the body dimension. Five tests addressed the whole body functions after stroke, four were specific to balance function and one to gait function. The evaluation of activities was carried out either with clinical tests or with questionnaires (either self-administered or interview-based). Two test address activities of the whole body, two tests were specific to the activities of the upper limb, one to locomotor activity and one to mobility.

The table of evidence of the functional tests used in elderly patients is presented in Table III. The four clinical tests identified addressed the dimension of activities. Two tests address the performance of whole body activities and two were specific to locomotor activities.

The table of evidence of the functional tests used in patient with traumatic brain injury is presented in Table IV. Three functional tests were identified. One of them consists of a classification of functional goals used for treatment planning. Two other clinical tests address either whole body activities, though mainly focussed on locomotor and transfer activities.

Discussion

The literature review allowed a selection of functional tests to be identified in four major diagnostic groups treated in community physiotherapy: musculoskeletal disorders (including lower back pain), stroke, elderly, and traumatic brain injuries. The functional tests identified cover essentially the body and activities dimension of the ICF. Along with functional tests allowing one aspect of the patient functioning to be assessed, a few studies described patient classification systems that were mostly used by a physical therapist in order to establish treatment goals and plan patient treatment.

The primary observation is that most of the tests identified cover the body and activity dimensions of the ICF, while very few tests address the participation dimension. Note that tests addressing participation have been developed more recently than tests for the other dimensions of functioning. Although rehabilitation aims at improving the performance of activities and participation in society according to Wade,[1] the clinical practice of physiotherapy puts a higher emphasis on the body dimension of functioning. This traditional approach is probably motivated by the fact that the selection of interventions of physical therapists is related to the treatment goals that are pursued.[50] Especially treatment goals at the level of impairments (body dimension) and at the level of activity limitations (individual dimension) are formulated.[51] Thus, up to now, treatment goals at the level of participation restrictions (social dimension) seem to be less specific

TABLE I.—*Evidence of functional tests used in patients with musculoskeletal disorders.*

Functional test	Clinical utility	Metric properties	General appraisal
Lower back pain			
Diagnostic classification system for non-specific lower back pain Type: classification [9]	Diagnostic subclassification of non-specific lower back pain consisting of pathoanatomically labelled syndromes assumed to refer to a specific pathological condition ICF dimension: body structure, body unction	Reliability: percentages of agreement ranged from 74% to 100% and kappa coefficients ranged from 0.26 to 1.00. Inter-tester reliability of categorisation of the syndromes was acceptable	Not applicable since this study proposes a diagnostic classification system for non specific lower back pain, rather than functional evaluation
Treatment based classification of lower back pain patients (TBC) Type: classification [10]	Treatment-based classification system, based on physical examination and patient self-reports of pain and disability, for use in the evaluation and treatment of patients with acute lower back pain ICF dimension: body structure, body function	Reliability: the interrater reliability was moderate (kappa=0.56) according to criteria of Landis and Koch. The percentage agreement between therapists was 65%	Not applicable since this study proposed a classification process for patients in the acute stage, rather than functional evaluation
Classification of patients with LOWER BACK PAIN Type: clinical test [11]	Impairment-based classification system addressing 28 symptom behavioral items and 22 alignment and movement items in an attempt to define mutually exclusive categories of lower back pain problems ICF dimension: body structure, body function	Validity: content validated by 4 orthopaedic physical therapists Reliability: experienced therapist who had trained together were able to agree on the results of examination and obtain acceptable level of reliability (kappa >=0.75 for all items related to symptoms elicited and >=0.40 in 72% of items related to alignment and movement)	Not applicable since this study proposed a classification of impairments in lower back pain patients, rather than functional evaluation
Modified-Modified Schober Test (MMST) Type: CLINICAL test [12]	Test measuring range of motion of the lumbar spine in the follow-up of lower back pain patients ICF dimension: body function	Validity: Pearson correlation coefficient between the MMST measurements and the gold standard (Xray) is 0.67 (95%CI: 0.44-0.84) Reliability: the intra-rater reliability was excellent (ICC=0.95; 95%CI 0.89-0.97) The inter- rater reliability was excellent (ICC= 0.91; 95%CI 0.83-0.96) Responsiveness: a change over 1 cm on the MMST must be observed to be 95% confident that a true change in ROM occurred	The metric properties of this test support its use in community-based physiotherapy. The test takes 5 minutes to be administered in by the therapist
Physical Impairment Index (PII) Type: clinical test [13]	Evaluate the physical impairment in lower back pain patients, through 7 tests, each scored dichotomously based on published cut-offs ICF dimension: body function, body structure	Validity: convergent validity was supported by significant correlations with disability (r=0.51 with the Roland-Morris questionnaire), work loss in the past year (r=0.43), pain (r=0.27), depression (r=0.26 with the Zung depression inventory), somatisation (r=0.32 with the modified somatic perception questionnaire), nonorganic signs (r = 0.49) and nonorganic symptoms (r = 0.35). The highest correlation were found with pain ratings (r=047), the Oswestry disability questionnaire (r=0.42), nonorganic signs (r=0.42) and nonorganic symptoms (r=0.36). Smaller but significant correlations were found with the physical component score of the SF-6 (r=-0.28) and the physical activity subscale of the fear-avoidance beliefs questionnaire (r=0.24) Reliability: good to excellent reliability for individual items (ICC or kappa coefficients ranging from 0.48 to 0.96). The overall score demonstrated excellent interrater reliability (ICC=0.89); the reliability coefficients for individual components rang	Systematic clinical test using published cut-off values for scoring physical impairment in acute lower back pain

(to be continued)

TABLE I.—*Evidence of functional tests used in patients with musculoskeletal disorders (continued).*

Functional test	Clinical utility	Metric properties	General appraisal
Shuttle walking test Type: clinical test [14]	Measure of walking capacity in patients with chronic airway obstruction, chronic heart failure and lower back pain, requiring the patient to shuttle on a 10-meter course at speeds ranging from 1.8 to 8.53 km/h ICF dimension: activities	Reliability: the test-reliability is excellent (ICC =0.99) with a mean difference of 2.5 m between assessments, and upper and lower limits of agreement of 52 m and -47 m Responsiveness: Patients undertaking fitness training reached an effect size of 1.2 compared to a control group of 0.23 and 0.94 for a group undergoing various orthopaedic treatments	The metric properties of this test show sensitivity to clinically meaningful change, a 10 m walking course is required and the test takes 15 min to be administered by the therapist
Standing flexion test Type: clinical test [15]	Manual palpation of the sacroiliac joint in patients with lower back pain ICF dimension: body function	Reliability: intraexaminer reliability data demonstrated a mean percentage agreement of 68% and a k coefficient of 0.46 indicating moderate reliability. Inter-examiner reliability data, with a mean percentage agreement of 42% and a k coefficient of 0.052, demonstrated statistically insignificant reliability. These results suggest that the reliability of the standing flexion test as an indicator of sacroiliac joint dysfunction still remains questionable	The metric properties do not support the use of this test in community-based physiotherapy. Moreover, the sample size of the study was very small (9 patients)
Aberdeen back pain scale Type: self-administered questionnaire [16]	Clinical assessment of patients with lower back pain made of 19 items of either forced choice or multiple choice, producing a back pain severity score ICF dimension: body function, activities, participation	Validity: developed through a review of the clinical literature and selection of items that reflected areas of importance in the clinical assessment of patients with lower back pain. Small but significant correlation (P<0.01) with two pain scales (r=0.24 with worst pain and 0.25 with best pain), the number of days absent from work over the past 6 months (r=0.26) and the number of visits to a general practitioner for back pain in the last 6 months (r=0.23) Responsiveness: large responsiveness to treatment as assessed by the standardized response mean (-7.64 to -13.36), either at 6 weeks, 6 months and 1 year follow-up	The metric properties do not support the use of this test in community-based physiotherapy since it is less discriminating than the well known Roland Disability Questionnaire
EuroQol (EuroQol) Type: self-administered questionnaire [16]	Generic instrument that incorporates descriptions and valuations of health states, applied in lower back pain patients ICF dimension: participation	Validity: content derived from a review of existing instruments, consisting of five items measuring mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Small but significant correlation (P<0.01) with two pain scales (r=-0.20 with worst pain and -0.29 with best pain), the number of days absent from work over the past 6 months (r=-0.14) and the number of visits to a general practitioner for back pain in the last 6 months (r=-0.28) Responsiveness: Moderate responsiveness to treatment as assessed by the standardized response mean (0.02 to 0.11), either at 6 weeks, 6 months and 1 year follow-up	The metric properties do not support the use of this test in community-based physiotherapy since it is a generic instrument, lacking responsiveness to treatment
Jan Van Breemen Instituut pain and function questionnaire Type: self-administered questionnaire [17]	Self-administered questionnaire consisting of 6 questions addressing pain and 9 questions addressing functional capacity in lower back pain patients ICF dimension: body function, activities	Responsiveness: sensitive to clinically important change (with function scale %change =33 and pain scale % change =37) after treatment of 4 to 6 weeks	The metric properties support the use of this test, nevertheless this scale is multidimensional
Lower back SF-36 physical functioning scale (Lower back SF-36 PF) Type: self-administered questionnaire [18]	Back-specific version of the SF-36 physical functioning scale, combining the advantages of both generic and specific functional questionnaires into a single, parsimonious set of items from	Validity: 18 items pooled from the Australian version of the SF-36 physical functioning scale, the modified version of the Oswestry questionnaire and the Quebec back pain disability scale. The lower back SF-36 PF scale has fewer misfitting items than the original SF-36 PF scale and eliminates its floor and	The low-back SF-36 PF18 proposes a revision of the well established SF-36 PF scale. It shows better psychometric properties than the

(to be continued)

TABLE I.—*Evidence of functional tests used in patients with musculoskeletal disorders (continued).*

Functional test	Clinical utility	Metric properties	General appraisal
	which the original SF-36 PF10-item scale can still be extracted ICF dimension: body function, activities Self-administered questionnaire of disability in lower back pain patients ICF dimension: activities	ceiling effects. Unidimensional and linear scale developed with the Rasch partial credit model Reliability: intrarater reliability of the PCBS test were .94 and .96, respectively. The test-retest of the lower back SF-36 PF (ICC=0.91) was comparable to that of the original SF-36 PF and of the Oswestry Responsiveness: The responsiveness of the lower back SF-36 PF was slightly, though not significantly, improved compared to that of the original SF-36 PF. The minimum detectable change was improved, as compared to the original SF-36 PF, from 16 to 12 points on a 0-100 scale expressing a percentage of the maximum possible score	original scale for the assessment of functioning in patients with lower back pain. The scores are expressed on a linear scale which enables quantitative comparisons of functional states. The test is self-administered and takes 5 min
Oswestry Type: Self-administered questionnaire [17]	Self-administered disability measure in lower back pain reflecting 24 activities of daily living ICF dimension: activities	Responsiveness: sensitive to clinically important change (%change =40) after treatment of 4 to 6 weeks	This self-administered disability measure in lower back pain is a well established test completed in 5 min. Instructions are available
Roland disability questionnaire Type: self-administered questionnaire [16]	Self-administered disability measure in lower back pain reflecting 24 activities of daily living ICF dimension: activities	Validity: 24 items selected form the Sickness Impact Profile represent the areas of greatest relevance to lower back pain. Small but significant correlation (P<0.01) with two pain scales (r=0.26 with worst pain and 0.33 with best pain), the number of days absent from work over the past 6 months (r=0.28) and the number of visits to a general practitioner for back pain in the last 6 months (r=0.29) Responsiveness: Large responsiveness to treatment as assessed by the standardized response mean (-1.04 to -3.62), either at 6 weeks, 6 months and 1 year follow-up	This self-administered disability measure in lower back pain is a well established test completed in 5 min. Instructions are available
Roland-Morris Questionnaire (RMQ) Type: self-administered questionnaire [19]	Self-administered disability measure in lower back pain reflecting 24 activities of daily living ICF dimension: activities	Responsiveness: the magnitude of the measurement error is sufficiently small to detect change after 4 to 6 weeks of physiotherapy treatment in patients with initial scores in the central portion of the scale; however, the magnitude is too large to detect improvement in patients with scores of less than 4 and deterioration in patients who have scores greater than 20	This self-administered disability measure in lower back pain is a well established test completed in 5 min. Instructions are available
Roland-Morris Type: self-administered questionnaire [17]	Self-administered disability measure in lower back pain reflecting 24 activities of daily living. ICF dimension: activities.	Responsiveness: sensitive to clinically important change (%change =40) after treatment of 4 to 6 weeks	This self-administered disability measure in lower back pain is a well established test completed in 5 min. Instructions are available
Sickness Impact Profile (SIP) Type: self-administered questionnaire [20]	Behaviorally based measure of perceived health status applicable across a spectrum of illnesses and among various demographic and cultural subgroups, applied in lower back pain patients ICF dimension: participation	Validity: validity coefficients of moderate magnitude have been reported between the SIP and various biologic and clinical measures, and between subsections of the SIP and sub categories of the Minnesota Multiphasic Personality Inventory Reliability: A high test-retest reliability coefficient (r=0.85) has been demonstrated in a lower back pain patient group Responsiveness: 20 items of the SIP showed change in at least 20% of the patients after a 6.4 weeks treatment and have an item-total score (minus item) cor-	The metric properties support the use of this self-administered test addressing participation in lower back pain patients

(to be continued)

TABLE I.—*Evidence of functional tests used in patients with musculoskeletal disorders (continued).*

Functional test	Clinical utility	Metric properties	General appraisal
		relation greater than 0.30. Seven of the 20 items identified in this study appear in the Roland-Morris disability questionnaire	
Spine disorders			
Fingertip-to-Floor Test Type: clinical test [21]	The finger-to-floor distance might be used to assess spine stiffness and the effects of exercise on spine stiffness in persons with spine disorders ICF dimension: body function	Validity: Spearman's correlation coefficient of trunk flexion assessed by the fingertip-to-floor test and the radiologic measure was excellent (r=-0.96) Reliability: The intra-observer reliability of the fingertip-to-floor test was excellent with an ICC of 0.99. The interobserver reliability of the radiologic analysis was excellent with an ICC of 0.99 Responsiveness: The responsiveness of the fingertip-to-floor test was evaluated by the SRM (0.97) and the effect size (0.87)	The metric properties of this test support its use in community-based physiotherapy. The test takes 1 min to be administered in by the therapist
Passive intervertebral motion of the cervical spine Type: clinical test [22]	Passive motion between adjacent articular segments is assessed when diagnosing and treating dysfunctions of the spine ICF dimension: body function	Reliability: the percentage agreement for both therapists for all tests was 77% (70±87%). The kappa coefficient varied between 0.28 and 0.43, considered to be only fair to moderate. Even if two examiners have equivalent clinical experience and educational backgrounds it is difficult to demonstrate acceptable inter-examiner reliability in the assessment of passive cervical joint motion. The results of this study showed lower concordance than expected in spite of the optimal testing conditions	The metric properties do not support the use of this test in community-based physiotherapy
Chronic pain			
Global Physiotherapy Examination 52 (GPE-52) Type: clinical test [23]	Physical examination for musculoskeletal pain patients in 5 domains: posture, respiration, movement, palpation muscle, palpation skin ICF dimension: body structure, body function	Validity: discriminates between patients with localized versus widespread pain. Discriminates between patients with long-lasting musculoskeletal problems and people who are healthy. Recently examined on data from people who were healthy and patients with long-lasting musculoskeletal pain Reliability: Good to excellent inter rater reliability of the total score (ICC=0.91) and individual components (ICC=0.65 for posture, 0.60 for respiration, 0.89 for movement, 0.83 for muscle palpation and 0.76 for skin palpation) Responsiveness: responsiveness to important change, defined in this study as return to work, was found only for the total GPE-52 score and within the movement and respiration domains. Responsiveness to important change was greater in patients with localized pain than in patients with widespread pain	This test requires a 3-day training which is a disadvantage compared to other tests, and it takes 30 min to be completed
Chronic pain intensity measures Type: telephone interview [24]	Rating through telephone interview of current, worst, least and average pain intensity on Visual Analog Scales over a determined period of time in chronic pain patients ICF dimension: body function	Reliability: the relatively low test-retest stability (0.55±0.65) of some of the individual ratings (average, current and worst) between one pair of assessments (1-month to 2-month follow-up) raises some issues Responsiveness: Each of the individual ratings was able to detect expected changes in pain intensity from pre-treatment to various points after treatment. Although the composite measures appeared more sensitive to treatment effects when compared to the individual ratings, these differences were not statistically significant	The metric properties do not support the use of this test in community-based physiotherapy since it is not reproducible over time

(to be continued)

TABLE I.—*Evidence of functional tests used in patients with musculoskeletal disorders (continued).*

Functional test	Clinical utility	Metric properties	General appraisal
Lower extremity dysfunction			
Lower Extremity Functional Scale (LEFS) Type: self-administered questionnaire [25]	Self-reported, condition-specific, functional measure applicable to a wide range of patients with lower-extremity orthopedic conditions ICF dimension: activities	Validity: correlations between the LEFS and the SF-36 physical function subscale and physical component score were r=0.80 Reliability: Excellent internal consistency (Chronbach's alpha=0.96). Test-retest reliability of the LEFS scores was excellent (r=0.94 [95% lower limit confidence interval = 0.89]) Responsiveness: The sensitivity to change of the LEFS was superior to that of the SF-36 in this population	The metric properties are supported by the methodology of the study. The test can be self-administered in 5 min
Questionnaire Rising and Sitting Down (QR&S) Type: self-administered questionnaire [26]	A self-administered questionnaire that measures perceived and actual functional limitations in rising and sitting down of patients with lower extremity dysfunction ICF dimension: activities	Validity: scale sum scores correlate 0.30 to 0.41 (P<0.001) with doctor's or therapist's global assessment of functional limitations. Sum scores correlate 0.20 to 0.59 (P<0.001) with self assessment of functional limitations in walking outdoors and walking stairs Reliability: The reliability coefficient Rho for object and combination scales ranges from 0.77 to 0.91. The intraclass correlation coefficient ranges from 0.72 to 0.90	The metric properties of this test support its use in community-based physiotherapy. The test is self-administered in 5 min
Lower limb amputation			
Functional Measure for Amputees (FMA) Type: Self-administered questionnaire [27]	A questionnaire collecting long-term functional and prosthetic use information following discharge in lower limb amputees ICF dimension: activities	Reliability: the FMA questionnaire showed moderate test-retest agreement	The metric properties of this test do not support its use in community-based physiotherapy
Ankylosing spondylitis			
World Health Organisation Disability Assessment Schedule II (WHODAS II) Type: Self-, Interviewer- or Proxy- administered questionnaire [28]	Generic instrument measuring the level of disability across various conditions and interventions, applied in ankylosing spondylitis ICF dimension: activities	Validity: the scores on the WHODAS II were significantly correlated with AS oriented questionnaires on physical functioning, disease activity, and quality of life. The WHODAS II and all AS oriented questionnaires also correlated well with all domains of the generic health status questionnaire SF-36. The variables BASDAI and BASFI at baseline were selected as independently associated with WHODAS II scores after five years (adjusted rÇ=0.48) Responsiveness: In all variables an improvement was seen, with moderate to large responsiveness scores. The WHODAS II showed a comparable, moderate responsiveness score. In the short term, a change in WHODAS II was significantly correlated with a change in physical functioning (BASFI and DFI), but not with a change in disease activity or quality of life	The test is useful for measuring disability in Ankylosing spondylitis and can also be self-administered

to physiotherapy, and fall into the realm of other medical and paramedical specialities. This notwithstanding, participation remains one of the key research areas in community rehabilitation according to Wade:[1] "Much research is needed to develop and evaluate the expertise that community-based teams might have. For example there is minimal research into the processes of giving support to patients and their families, monitoring, prevention of complications, and facilitation of social participation. This observation also suggests that outcome evaluation in physiotherapy should be in keeping with this framework in order to monitor functional recovery of the patient and treatment efficacy."

The functional tests reviewed appeared to address the dimensions most related to the clinical manage-

TABLE II.—*Evidence of functional tests used in stroke patients.*

Functional test	Clinical utility	Metric properties	General appraisal
Body dimension - global Fugl-Meyer Type: clinical test [29]	Assess the physical impairment following stroke, in terms of motor performance, balance, sensation, range of movement and pain in 155 items ICF dimension: body structure, body function	Reliability: the standard error of measurement (combining, rater, occasion and error variance) may be significant when only small changes in the patient's level of motor performance are expected. However, the overall inter-rater reliability was high (ICC=0.96) as well as the subscore (ICC=0.85 to 0.97), with exception of pain (ICC=0.61)	This widely used test is hardly applicable, as it is composed of multiple sub-scales that cannot be summed in a total score. The test takes 30 min to be administered by the therapist
Rivermead Motor Assessment (RMA) Type: clinical test [30]	Detailed assessment of gross motor control, upper limb control and lower limb and trunk control after stroke ICF dimension: body function	Reliability: reliability varied between questions (kappa=0.33 to 0.37), but agreements between assessors and between observed and asked performance were always better than expected by chance	The metric properties do not support the use of this test in community-based physiotherapy
Stroke Impairment Assessment Set (SIAS) Type: clinical test [31]	The SIAS assesses various aspects of impairment in hemiplegic patients (e.g.,motor function, sensory function, pain) ICF dimension: body function	Validity: standardized measure of stroke impairment consisting of subcategories of motor function, tone, sensory function, ROM, pain, trunk function, visuospatial function, speech, and sound side function. The scale has been validated with the Rasch model. The item difficulty patterns were similar for the right- and left-sided lesion groups Reliability: Scale reliability is usually analyzed as internal consistency or unidimensionality. The fit statistics were acceptable, except for a few items. Justifications for these discrepancies are presented in the study. The item difficulty patterns were identical at admission and discharge	The metric properties of this test support its use in community-based physiotherapy. The scores are expressed on a linear scale which enables quantitative comparisons of functional states. The test administered by the therapist in approx. 10 min. Published scoring guidelines are available
Stroke Physiotherapy Intervention Recording Tool (SPIRIT) Type: clinical test [32]	Define and describe the content of physiotherapy intervention for postural control (posture, balance, walking) post stroke ICF dimension: body function, body structure, activities	Validity: first study in the development of a recording system for physiotherapy interventions in the rehabilitation of postural control post-stroke. Content based on literature review and validated against expert advice	Not applicable since this study proposes a recording system for physiotherapy interventions in postural control of acute stroke patients, rather than functional evaluation
Stroke Rehabilitation Assessment of Movement (STREAM) Type: clinical test [33]	Objective and quantitative evaluation of the motor functioning of individuals with stroke among 3 subscales: upper-limb movements, lower-limb movements, and basic mobility items ICF dimension: body function	Validity: the total STREAM score was moderately to highly associated with the score of the Barthel Index (rho=0.67) and Fugl-Meyer motor assessment scale (rho=0.95) Reliability: Moderate to excellent agreement inter-rater on scores for individual items (weighted kappa=0.55 to 0.94). Very high inter-rater reliability (ICC=0.96 for the total score, 0.95 for upper extremity, 0.92 for lower extremity and 0.92 for mobility subscales)	This test requires a 2-day training, and was mostly validated on acute patients. Moreover, the total score is composed of three sub-scales, which is not suitable for financing purposes
Body dimension - balance			
Berg Balance Scale (BBS) Type: clinical test [34]	Determine change in functional standing balance over time ICF dimension: body function	Validity: in arm flexion condition, the force platform and accelerometer outcomes were significantly related to BBS performance, with CP-flexion explaining 43% and peak arm acceleration explaining 45% of the variance in BBS scores. It appears that the ability to show activation of postural muscles in advance of focal movement is associated with higher BBS scores. In quiet stance condition, performance on the BBS was found to be significantly related to CP-stance (r^2=58%) Reliability: Paired tests of difference failed to reveal any differences in the variables over the two days	There is evidence in the literature to support the use of the BBS. Established guidelines allow the test to be administered by the therapist in 20 min

(to be continued)

TABLE II.—*Evidence of functional tests used in stroke patients (continued).*

Functional test	Clinical utility	Metric properties	General appraisal
Berg Balance Scale (BBS) Type: clinical test [35]	Determine change in functional standing balance over time ICF dimension: body function	Validity: r=0.80 with gait speed, r>0.80 with Barthel Index, r> 0.70 with Fugl-Meyer, r<0.60 with FIM Reliability: Excellent test-retest reliability (ICC=0.97). Excellent inter rater reliability (ICC=0.98) Responsiveness: A difference of five to seven BBS points is necessary to conclude with 90% certainty that patients receiving rehabilitation following CVA have undergone a real change in BBS performance when assessed in a between-rater situation	There is evidence in the literature to support the use of the BBS. Established guidelines allow the test to be administered by the therapist in 20 min
Functional balance tests post stroke Type: clinical test [36]	Functional test of postural control post stroke ICF dimension: body function	Validity: poor to high: rho=0.33-0.54 with Motor Assessment scale, rho=0.26-0.70 with Berg Balance Scale, rho=0.32-0.61 with Rivermead Mobility Index Reliability: Good to excellent test-retest reliability (ICC: 0.91-1). Good to excellent inter rater reliability (ICC=0.88-1)	The metric properties do not support the use of this test in community-based physiotherapy
Postural Control and Balance for Stroke test (PCBS) Type: clinical test [37]	Assess balance in terms of postural changes, sitting balance, and standing balance in order to help physiotherapists in planning balance interventions and in rehabilitation follow-up ICF dimension: body function	Reliability: the Cronbach alpha for all the items combined was 0.96. The ICC values for the interrater and intrarater reliability of the PCBS test were 0.94 and 0.96, respectively Responsiveness: The PCBS test showed an acceptable level of reliability and the responsiveness results indicated a good level before 120 days, but not between 120 and 360 days after stroke	The metric properties do not support the use of this test in community-based physiotherapy since the criteria for scoring are doubtful and the resulting score is multidimensional. The test takes 30 min to be completed
Body dimension-gait			
Rivermead Visual Gait Assessment (RVGA) Type: clinical test [38]	The current study aims to develop a procedure and assessment form which allows practising therapists to record relevant aspects of the quality of gait observed in patients with neurological disease in a reliable and valid way ICF dimension: body function	Validity: there was a significant correlation between the global RVGA score and the various criterion measures (r=0.53-0.79; P<0.001) and between change in the RVGA score and change in walking time in patients who received treatment (r=0.68; P<0.01) Reliability: inter-rater reliability between multiple raters was reasonable both for the global scores from the gait assessment form (Kendall's coefficient of concordance; P<0.001), and for individual items (complete agreement occurred on 63.8% of all observations)	The metric properties of this test support its use in community-based physiotherapy. The test takes 15 min to be administered by the therapist
Activities - global			
Mobility milestones Type: clinical test [39]	Evaluate mobility after stroke in 4 simple functional tasks providing a quick, simple, and standardized outcome measure ICF dimension: activities	Reliability: three quarters of the physiotherapists showed high intrarater reliability, with only 3 subjects showing poor intrarater reliability. 28 subjects?of the sample showed good or high test-retest reliability, with only 3 subjects showing poor reliability. The percentage agreement on the mobility milestones between the physiotherapists' ratings and our ratings was very high overall. Percentage agreement between therapists between 78.3 and 97.5 for all items	Not applicable since the test addresses early recovery stages after stroke in a hospital setting
Stroke Activity Scale (SAS) Type: clinical test [40]	Measure motor function at the level of disability in stroke patients ICF dimension: body function, activities	Reliability: the internal consistency of the SAS was estimated using Cronbach's alpha as 0.68 at time 1 and 0.69 at time 2. A high level of agreement was observed for the total score (GCC=0.96). Inter-rater reliability for individual items was good to very good. A very high level of agreement was observed for the total score (GCC>0.94)	Not applicable since the test addresses early recovery stages after stroke in hospital setting

(to be continued)

TABLE II.—*Evidence of functional tests used in stroke patients (continued).*

Functional test	Clinical utility	Metric properties	General appraisal
Activities - upper limb			
ABILHAND Type: self-administered questionnaire [41]	Self-reported questionnaire on the perceived difficulty of manual activities of daily living ICF dimension: activities	Validity: content validated against the involvement of the affected limb in each activity. The activities that define the more difficult levels of the scale also tend to require a higher involvement of the affected limb, while the easier activities can be achieved in a movement sequence that does not require the affected limb. ABILHAND measures are significantly correlated to grip strength (r=0.56), motricity (rho=0.73), dexterity (r=0.60), and depression (rho=-0.21). The ABILHAND questionnaire results in a valid, unidimensional, person-centered measure of manual ability in everyday activities. The stability of the item-difficulty hierarchy across different patient classes further supports the clinical application of the scale Reliability: The overall scale precision is summarized by a good between-patient separation reliability of 0.90 in this sample. It appears sufficient to discriminate across patients and, presumably, to capture even subtle functional changes with time	The metric properties of this test support its use in community-based physiotherapy. The scores are expressed on a linear scale which enables quantitative comparisons of functional states. The test can be self-administered in 5 min
Motor Activity Log (for the Assessment of Arm Use in Hemiparetic Patients) (MAL) Type: semi-structured interview [42]	Assess the use of the paretic arm and hand during activities of daily living in terms of the amount of use (AOU) and of the quality of movement (QOM) of the paretic arm ICF dimension: activities	Validity: The cross-sectional construct validity of the MAL is reasonable, but the results raise doubts about its longitudinal construct validity Reliability: Internal consistency was high (Chronbach's alpha=0.88 for the AOU and 0.91 for the QOM). The reproducibility is sufficient to detect an individual change of at least 12% to 15% of the range of the scale Responsiveness: The improvement on the MAL during the intervention was only weakly related to the improvement on the Action Research Arm test (rho=0.16 to 0.22)	The use of the MAL as a primary outcome measure in clinical trials is not recommended because there are reasons to doubt the longitudinal construct validity of the instrument
Activities-gait			
Gait speed Type: clinical test [30]	Timed walk at the patient's preferred speed, using the patient's own selection of walking aid or assistance, over a distance of 5 m, including turn and walk back ICF dimension: activities	Reliability: intercorrelations between 3 repetitions of the test, between 5m and 10m walks, and between test and retest were very high (r=0.95 to 0.99). For individual patients, the test-retest variability ranged from 0 to 40%, with 95% of the patients varying the walking speed between tests by less than 25% of the slowest time	The metric properties of this test support its use in community-based physiotherapy. The test can be administered in 5 min
Activities – mobility			
New mobility scale Type: open-ended questionnaire [43]	Measure mobility inside and outside the home and in the community ICF dimension: activities	Validity: content of mobility scale assessed through open-ended questionnaire sent to 15 experienced physiotherapists. They identified important mobility tasks and important places to access inside the house, outside the house and in the community as well as additional mobility abilities required for independent mobility in the house and community	Preliminary study of test under development; not applicable

ment of patients in physiotherapy. All tests addressed either the body or the activities dimension of the patient's functioning, which are the most related to the clinical management of patients in physiotherapy. Nevertheless, some of the most widely used tests in rehabilitation (*e.g.*,the Functional Independence Measure or the Barthel Index) were not listed, which contrasts with a recent inquiry on outcome measure-

TABLE III.—*Evidence of functional tests used in elderly patients.*

Functional test	Clinical utility	Metric properties	General appraisal
6-minute Walk Test Type: clinical test [44]	Measure of exercise capacity and endurance also used as an outcome measure ICF dimension: activities	Reliability: intraclass correlation coefficients for test-retest reliability were 0.93 Responsiveness: There was no change in 6-minute walk test distance in the intervention group when compared with the control group	The test was unable to measure change in performance expected with a functional training intervention. An indoor 32-meter course is also required. Its applicability is limited
8-item Physical Performance Test (PPT-8) Type: clinical test [44]	Evaluate the ability to use the upper and lower extremities in activities of daily living ICF dimension: activities	Reliability: good internal consistency (Cronbach's alpha=0.78). Intraclass correlation coefficients for test-retest reliability was 0.88. The intraclass correlation coefficient for interrater reliability was 0.96 Responsiveness: The responsiveness index was 0.8, indicating a significant difference of effect between intervention and control group	The metric properties are supported by the methodology of the study. A published test protocol allows the test to be administered by the therapist
Functional Gait Assessment (FGA) Type: clinical test [45]	Assess postural stability during gait tasks in the older adult (greater than 60 years of age) at risk for falling ICF dimension: activities	Validity: Poor to moderate correlation with balance measurements (rho=0.11 to 0.67) Reliability: Good internal consistency (Cronbach's alpha=0.79). Good inter rater reliability (ICC=0.74). Good inter rater reliability (ICC=0.86)	This test has been developed for a very specific group of patients. The results of this study should be interpreted cautiously since the sample size was very small (6 patients)
General Motor Function Assessment Scale (GMF) Type: clinical test [46]	Compound assessment of three components (dependence, pain and insecurity) of daily living activities among older rehabilitation patients ICF dimension: activities	Validity: Principal components factor analysis demonstrated individual FGA item loading across 3 extracted factors that may represent separate domains of performance on the total battery Reliability: Analysis of reliability showed overall high values of percentage agreement (PA=0.70) and of the rank-order agreement coefficient (ra=0.82), and low degrees of systematic disagreement	The metric properties do not support the use of this test in community-based physiotherapy

ment in Physical Medicine and Rehabilitation across Europe.[52] This observation might be explained by the fact that the current literature review was focussed on functional evaluation carried in the community, where patients are able to live at home, and therefore located at the higher range of measurement of such tests. While the tests mostly used in hospital rehabilitation are pretty well documented, functional assessment in the community are less standardised. Moreover, this literature review considered only papers published since 1990. Consequently, some standardised tests were not reported because they were validated before 1990. This is particularly the case for some functional tests in the body dimension like the nine-hole peg test or the Purdue pegboard test.

The body dimension was mostly addressed in terms of body function, except in lower back pain patients or in other spinal disorders that primarily relied on testing body structures through skin palpation. The activities dimension was generally addressed with clinical tests performed by the physical therapist or by a self-reported questionnaire relating to the patient's perceived disability in daily life activities. Clinical tests typically address movement quality or a patient's performance in standardized activities, and are considered more reliable than questionnaires. Nevertheless, questionnaires present several advantages over clinical test. First, they address the perception of disability in activities as they are actually realized by the patient in their own environment. Second, they capture an average perception of disability over a longer period of time (one up to several weeks) and are probably more representative of the patient's disability than the performance of standardized activities in a consultation room. Third, they are very inexpensive and can be completed by the patient in the waiting room in order to reduce the burden of work of the physical therapist. On the other hand, it can be discussed

TABLE IV.—*Evidence of functional tests used in patients with brain injury.*

Functional test	Clinical utility	Metric properties	General appraisal
Clinical Outcomes Variable Scale (COVS) Type: clinical test [47]	Evaluation of functional ability following traumatic brain injury in a large range of motor tasks retrained by physiotherapists including a measure for transfer ability to and from bed and floor surfaces as well as wheelchair skill ICF dimension: body function, activities	Reliability: the intra-class correlation coefficients (ICC) were very high for both inter-tester reliability (ICC>0.97 for total COVS scores, ICC>0.93 for individual COVS items) and intra-tester reliability (ICC>0.97)	The metric properties do not support the use of this test in community-based physiotherapy
Step length and step width measurement Type: clinical test [48]	Walking performance after traumatic brain injury assessed by step length and step width measures ICF dimension: body function	Validity: concurrent validity was excellent, with correlations between the procedures ranging from 0.93 to 1.00 Reliability: the inter-rater reliability of step length and width measurements was very high, with intraclass correlation coefficients between 0.94 and 1.00, for both procedures	Potentially interesting test for neurological patients, but its metric properties should be interpreted cautiously since the sample size was small (20 patients). A 14-meter walk course is required
Taxonomy of patient goals in acquired brain injury comprising 21 categories within five domains, utilizing 125 descriptors Type: classification [49]	Classification of client goals in community-based acquired brain injury rehabilitation allowing treatment focus and changes in service delivery to be investigated ICF dimension: activities, participation	Reliability: the taxonomy demonstrated good inter-rater consistency and was able to discriminate between similar but related data sets comprising goal statements. Out of the 140 goal statements 128 (92%) were placed in the same category by at least three of the four, and there was full agreement between all four on 91 of the goal statements (65%). This indicated that the taxonomy and descriptor statements had been refined to a stage where there was significant inter-rater consistency	Not applicable since this study proposes a classification of patient goals, rather than functional evaluatio

TABLE V.—*Selected tests according to the dimensions of the International Classification of Functioning, Disability and Health (ICF).*

Body	Activities	Participation
Musculoskeletal disorders		
Fingertip-to-floor test Modified-Modified Schöber Test (MMST)	Low back SF-36 physical functioning Lower Extremity Functional Scale (LEFS) Oswestry Questionnaire Rising and Sitting Down (QR&S) Roland-Morris Shuttle walking test World Health Organisation Disability Assessment Schedule II (WHODAS II)	Sickness Impact Profile (SIP)
Stroke		
Berg Balance Scale (BBS) Rivermead Visual Gait Assessment (RVGA) Stroke Impairment Assessment Set (SIAS)	ABILHAND Gait speed	
Elderly		
	8-item Physical Performance Test (PPT-8)	
Brain injury		
Step length and step width measurement		

whether it is feasible to base reimbursement on self-reported questionnaires.

The psychometric qualities of the test were reported as presented in each study. These statistics can be interpreted by the reader in order to make an optimal choice of functional tests for their specific purpose.. Some precautions, however, must be kept in mind in the comparison of statistical indices. For instance, although validity was generally assessed by correlation analyses, it is well known that correlation coefficients are sensitive to the range and distribution of the values correlated.[53] The intensity of the reported correlations is therefore influenced by the difference in sampling strategies used between studies. Moreover, different inclusion criteria were used between studies (*e.g.,*elderly patients with or without a previous stroke). The selection bias should be considered from a clinical perspective in comparing statistical indices since it influences the relation between the functional test and the gold standard.

Responsiveness was also assessed with different types of indicators. Responsiveness is generally defined as the ability of a test to elicit clinically meaningful change. Nevertheless, all evaluations with functional tests are accompanied by a certain amount of measurement error, and some of the statistics used to assess responsiveness do not take the error of measurement into account (*e.g.,*percentage of change score). Thus, when assessing change with such statistics, scores obtained at both occasions and regarded as significantly different might very well be within the error of measurement and therefore the possibility that the change is due to chance cannot be excluded. Moreover, the reviewed functional tests used different levels of measurement.[54] Though the large majority of results were reported at the ordinal level (*i.e.,*allowing for ranking comparisons), some tests reported their results on nominal scales (*i.e.,*only allowing for equivalence comparison) or on interval or ratio scale (*i.e.,*allowing difference, or change score, to be computed on a linear scale). Misinferences due to linear interpretation of data at the nominal or ordinal level of measurement has been largely reported in the literature.[55]

Conclusions and perspectives

Functional evaluation is part of the process of rehabilitation.[1] This process involves identifying the problems and needs of individuals, defining therapy goals, planning and implementing interventions and assessing the effect of interventions using measurements of relevant variables.[56] The literature review presented in this report provides a selection of functional tests in the major diagnostic groups treated in community physiotherapy. These functional tests essentially cover the body dimension and the activities dimension of the ICF, which are the dimensions that are most pertinent to the current practice of physiotherapy.

A selection of tests in each dimension of the ICF is presented in Table V according to their metric properties and quality appraisal for the four diagnostic groups emerging from this review. The metric properties include published evidence of reliability, validity and responsiveness, but also the unidimensionality and linearity of each scale. The latter properties are requisite for making quantitative comparisons of functional status over time. Quality appraisal includes a general evaluation of the methodology of each study, and the practical applicability of each test in community physiotherapy (*i.e.,* evaluation time, required equipment). It is important to emphasize that time and quality of evaluation are not strictly related. Indeed, very short, well calibrated scales can be much more efficient than time consuming tests. The body and activities dimensions of the ICF are the most well-represented. Musculoskeletal disorders are the diagnostic groups for which the largest number of high quality tests have been identified. The body dimension in musculoskeletal disorders is mostly evaluated with active mobility tests. Although these tests are widely used, one of their major drawbacks is that they measure a maximal performance, *i.e.,* maximal mobility, which can be hindered by pain or other factors.[57] The activities dimension in musculoskeletal disorders is the most represented among the four diagnostic groups. Most tests used in this domain are widely used, validated questionnaires that can be easily and inexpensively applied in community physiotherapy. The body dimension in stroke patients is addressed either with global tests and/or tests of specific functions (*e.g.,* balance). The activities dimension is evaluated either with upper limb or with lower limb specific tests; no global activity instrument was identified in stroke community rehabilitation. Although few tests are available for elderly and brain-injured patients, one test in each diagnostic group presented enough quality to be retained in this selection. The summary

table presented here indicates that 1) high quality functional tests were found for the main diagnostic groups treated in community physiotherapy (*e.g.,*musculoskeletal disorders, stroke); 2) the functional tests identified do not cover all dimensions of the ICF for all diagnostic groups, 3) in some dimensions, various valid scales are proposed, hence emphasising the need for a standard in each diagnostic group.

This literature review has identified a selection of valid tests that could be used as a basis for the standardisation of the functional evaluation of the major diagnostic groups treated in community physiotherapy. This means that standards are available for reporting and following the evolution of patients. Nevertheless, no attempt in using functional outcomes as a rationale for financing physiotherapy was found to date in the current literature review.

References

1. Wade DT. Community rehabilitation, or rehabilitation in the community? Disabil Rehabil 2003;25:875-81.
2. Streiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. 2nd edition. Oxford: Oxford University Press; 1995.
3. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bulletin 1979;86:420-8.
4. Cohen J. A coefficient of agreement for nominal scales. Psycho Meas 1960;20:37-46.
5. Soeken KL, Prescott PA. Issues in the use of kapp to estimate reliability. Medical Care 1986;24:733-41.
6. Maclure M, Willet WC. Misinterpretation and misuse of the kappa statistic. Am J Epidemiol 1987;126:161-9.
7. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27S:178-89.
8. Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171-8.
9. Petersen T, Olsen S, Laslett M, Thorsen H, Manniche C, Ekdahl C *et al*. Inter-tester reliability of a new diagnostic classification system for patients with non-specific low back pain. Aust J Physiother 2004;50:85-94.
10. Fritz JM, George S. The use of a classification approach to identify subgroups of patients with acute low back pain. Interrater reliability and short-term treatment outcomes. Spine 2000;25:106-14.
11. Van Dillen LR, Sahrmann SA, Norton BJ, Caldwell CA, Fleming DA, McDonnell MK *et al*. Reliability of physical examination items used for classification of patients with low back pain. Phys Ther 1998;78:979-88.
12. Tousignant M, Poulin L, Marchand S, Viau A, Place C. The Modified-Modified Schober Test for range of motion assessment of lumbar flexion in patients with low back pain: a study of criterion validity, intra- and inter-rater reliability and minimum metrically detectable change. Disabil Rehabil 2005;27:553-9.
13. Fritz JM, Piva SR. Physical impairment index: reliability, validity, and responsiveness in patients with acute low back pain. Spine 2003;28:1189-94.
14. Taylor S, Frost H, Taylor A, Barker K. Reliability and responsiveness of the shuttle walking test in patients with chronic low back pain. Physiother Res Int 2001;6:170-8.
15. Vincent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. Man Ther 1999;4:87-93.
16. Garratt AM, Klaber Moffett J, Farrin AJ. Responsiveness of generic and specific measures of health outcome in low back pain. Spine 2001;26:71-7.
17. Stratford PW, Binkley J, Solomon P, Gill C, Finch E. Assessing change over time in patients with low back pain. Phys Ther 1994;74:528-33.
18. Davidson M, Keating JL, Eyres S. A low back-specific version of the SF-36 Physical Functioning scale. Spine 2004;29:586-94.
19. Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. Phys Ther 1996;76:359-65.
20. Stratford P, Solomon P, Binkley J, Finch E, Gill C. Sensitivity of Sickness Impact Profile items to measure change over time in a low-back pain patient group. Spine 1993;18:1723-7.
21. Perret C, Poiraudeau S, Fermanian J, Colau MM, Benhamou MA, Revel M. Validity, reliability, and responsiveness of the fingertip-to-floor test. Arch Phys Med Rehabil 2001;82:1566-70.
22. Smedmark V, Wallin M, Arvidsson I. Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine. Man Ther 2000;5:97-101.
23. Kvale A, Skouen JS, Ljunggren AE. Sensitivity to change and responsiveness of the global physiotherapy examination (GPE-52) in patients with long-lasting musculoskeletal pain. Phys Ther 2005;85:712-26.
24. Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. Pain 1999;83:157-62.
25. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. Phys Ther 1999;79:371-83.
26. Roorda LD, Roebroeck ME, Lankhorst GJ, van Tilburg T, Bouter LM. Measuring functional limitations in rising and sitting down: development of a questionnaire. Arch Phys Med Rehabil 1996;77:663-9.
27. Callaghan BG, Sockalingam S, Treweek SP, Condie ME. A post-discharge functional outcome measure for lower limb amputees: test-retest reliability with trans-tibial amputees. Prosthet Orthot Int 2002;26:113-9.
28. van Tubergen A, Landewe R, Heuft-Dorenbosch L, Spoorenberg A, van der Heijde D, van der Tempel H *et al*. Assessment of disability with the World Health Organisation Disability Assessment Schedule II in patients with ankylosing spondylitis. Ann Rheum Dis 2003;62:140-5.
29. Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. Phys Ther 1993;73:447-54.
30. Collen FM, Wade DT, Bradshaw CM. Mobility after stroke: reliability of measures of impairment and disability. Int Disabil Stud 1990;12:6-9.
31. Tsuji T, Liu M, Sonoda S, Domen K, Chino N. The stroke impairment assessment set: its internal consistency and predictive validity. Arch Phys Med Rehabil 2000;81:863-8.
32. Tyson SF, Selley A. The development of the Stroke Physiotherapy Intervention Recording Tool (SPIRIT). Disabil Rehabil 2004;26:1184-8.
33. Wang CH, Hsieh CL, Dai MH, Chen CH, Lai YF. Inter-rater reliability and validity of the stroke rehabilitation assessment of movement (stream) instrument. J Rehabil Med 2002;34:20-4.
34. Stevenson TJ, Garland SJ. Standing balance during internally produced perturbations in subjects with hemiplegia: validation of the balance scale. Arch Phys Med Rehabil 1996;77:656-62.
35. Stevenson TJ. Detecting change in patients with stroke using the Berg Balance Scale. Aust J Physiother 2001;47:29-38.
36. Tyson SF, DeSouza LH. Reliability and validity of functional balance tests post stroke. Clin Rehabil 2004;18:916-23.
37. Pyoria O, Talvitie U, Villberg J. The reliability, distribution, and

responsiveness of the Postural Control and Balance for Stroke Test. Arch Phys Med Rehabil 2005;86:296-302.
38. Lord SE, Halligan PW, Wade DT. Visual gait analysis: the development of a clinical assessment and scale. Clin Rehabil 1998;12:107-19.
39. Baer GD, Smith MT, Rowe PJ, Masterton L. Establishing the reliability of mobility milestones as an outcome measure for stroke. Arch Phys Med Rehabil 2003;84:977-81.
40. Horgan NF, Finn AM, O'Regan M, Cunningham CJ. A new stroke activity scale-results of a reliability study. Disabil Rehabil 2003;25:277-85.
41. Penta M, Tesio L, Arnould C, Zancan A, Thonnard JL. 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.
42. van der Lee JH, Beckerman H, Knol DL, de Vet HC, Bouter LM. Clinimetric properties of the motor activity log for the assessment of arm use in hemiparetic patients. Stroke 2004;35:1410-4.
43. Stanko E, Goldie P, Nayler M. Development of a new mobility scale for people living in the community after stroke: content validity. Aust J Physiother 2001;47:201-8.
44. King MB, Judge JO, Whipple R, Wolfson L. Reliability and responsiveness of two physical performance measures examined in the context of a functional training intervention. Phys Ther 2000;80:8-16.
45. Wrisley DM, Marchetti GF, Kuharsky DK, Whitney SL. Reliability, internal consistency, and validity of data obtained with the functional gait assessment. Phys Ther 2004;84:906-18.
46. Aberg AC, Lindmark B, Lithell H. Development and reliability of the General Motor Function Assessment Scale (GMF)a performance-based measure of function-related dependence, pain and insecurity. Disabil Rehabil 2003;25:462-72.
47. Low Choy N, Kuys S, Richards M, Isles R. Measurement of functional ability following traumatic brain injury using the Clinical Outcomes Variable Scale: a reliability study. Aust J Physiother 2002;48:35-9.
48. van Loo MA, Moseley AM, Bosman JM, de Bie RA, Hassett L. Inter-rater reliability and concurrent validity of step length and step width measurement after traumatic brain injury. Disabil Rehabil 2003;25:1195-200.
49. Kuipers P, Foster M, Carlson G, Moy J. Classifying client goals in community-based ABI rehabilitation: a taxonomy for profiling service delivery and conceptualizing outcomes. Disabil Rehabil 2003;25:154-62.
50. Dekker J. Application of the ICIDH in survey research on rehabilitation: the emergence of the functional diagnosis. Disabil Rehabil 1995;17:195-201.
51. Dekker J, van Baar ME, Curfs EC, Kerssens JJ. Diagnosis and treatment in physical therapy: an investigation of their relationship. Phys Ther 1993;73:568-77.
52. Haigh R, Tennant A, Biering-Sorensen F, Grimby G, Marincek C, Phillips S *et al*. The use of outcome measures in physical medicine and rehabilitation within Europe. J Rehabil Med 2001;33:273-8.
53. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
54. Stevens SS. On the theory of scales of measurement. Science 1946;103:677-80.
55. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Arch Phys Med Rehabil 1989;70:308-12.
56. Steiner WA, Ryser L, Huber E, Uebelhart D, Aeschlimann A, Stucki G. Use of the ICF model as a clinical problem-solving tool in physical therapy and rehabilitation medicine. Phys Ther 2002;82:1098-107.
57. Watson P. Non-physiological determinants of physical performance in musculoskeletal pain. In: Mitchell M, editor. Pain 1999, an updated review refresher course syllabus; 9th World Congress of Pain; Seattle: IASP press; 1999. p. 153-7.