Outcome measurement in cognitive neurorehabilitation

Nadina Lincoln and Roshan das Nair

Introduction

The aim of this chapter is to consider the criteria for selecting outcome measures for evaluating the effects of cognitive neurorehabilitation. The International Classification of Function, Disability and Health (ICF) (World Health Organisation, 2001) is used as a framework for deciding what to measure. The properties of the ideal outcome measure are discussed. Examples of outcome measures commonly used in clinical studies are provided and their strengths and limitations considered. The focus is on self-report measures rather than neuro-psychological tests as these reflect the effect of cognitive rehabilitation on daily life.

Outcome

Outcomes may be assessed at the levels of impairment, activity or participation.

Activity measures are the most important outcomes for cognitive rehabilitation.

Quality of life is best assessed as component domains rather than a single measure.

Rehabilitation may be considered in terms of process, structure and outcome (Donabedian, 1966). Process consists of the activities which are designed to improve the functioning of the individual, such as the treatment techniques used by members of the multidisciplinary team to foster recovery or adaptation. Structure refers to the facilities provided to enable the treatments to be administered, such as the environment, staff and equipment. Outcome refers to the result of the rehabilitation endeavor. It is the endpoint against which the effectiveness of rehabilitation is judged. In cognitive rehabilitation, the aim is to help the patient to function to the maximum level of ability possible within the constraints of deficits resulting from brain damage. In addition, that individual should be as contented and satisfied with his or her condition as is possible, and so should the relatives. The assessment of outcome is the means by which we determine whether rehabilitation has achieved these aims.

Measurement of outcome

The ICF (World Health Organisation, 2001) is recognised as providing a useful framework for the selection of appropriate outcome measures (Heinemann, 2005; Jette & Haley, 2005; Mermis, 2005; Wade, 2003). The concepts are as follows.

(a) Body functions and deficits: these are impairments or the loss or abnormality of psychological, physiological or anatomical structure or function. They include cognitive deficits such as disorders of memory, attention and language.

(b) Activity limitation: the difficulties an individual may have in executing activities, including learning and applying knowledge; self-care; domestic life; interpersonal interactions; and community, social and civic life. It includes the effects of cognitive impairments on daily life, such as difficulties in telling the time and losing
items around the home, and the disruption to interpersonal relationships that may occur following a head injury.

(c) Participation is the involvement in life situations at a societal level. It includes the social, cultural, economic and environmental effects of activity limitation.

An impairment, such as visual inattention, usually will give rise to an activity limitation, such as the inability to dress independently, which in turn may affect participation, through loss of personal independence. An effective rehabilitation programme would reduce the impairments and the activity limitations which are consequent upon those impairments. In many instances, it is not possible to ameliorate the impairment, but nevertheless significant gains may be made by attempting directly to increase activity. Although it would be desirable to improve participation, this can rarely be achieved directly and may not entirely be within the remit of the rehabilitation service. Provided the activity limitation has been reduced, the assumption is that there will be a beneficial effect on participation.

Quality of life is an elusive outcome which relates to participation. Most assessments of quality of life incorporate several domains, which include both activities and participation. It is beyond the scope of assessment procedures to assess adequately all domains which contribute to quality of life and produce a satisfactory single quality-of-life measure. For this reason it does not seem practical even to attempt to assess general quality of life, but rather only the specific domains, even though improving the quality of life must be the ultimate goal of rehabilitation.

**Selection of measures**

- The psychometric properties of measures need to be considered.
- Measures of motor, sensory and cognitive impairment provide standardised descriptions of patients.
- Cognitive tests measure cognitive impairment and not the effect in daily life.

In order to choose an outcome measure it is important to know that the measure meets the requirements of a measurement tool. These requirements are:

(a) Validity: any measure must measure what it purports to measure. For example, measures of activities of daily living should include those activities people would consider to be essential for independence in daily life. Measures should relate to other measures of the same underlying ability and include all the relevant aspects of the attributes they measure. Hypotheses generated based on the measure should be upheld. For example, one would predict that head-injured people will do less well on a measure of memory than normal participants, and if this is supported it indicates that the measure has construct validity.

(b) Reliability: any outcome measure should provide the same information if used by different assessors (inter-rater reliability) or by the same assessor on different occasions (intra-rater reliability). If the assessment is to be used to monitor change it needs to have minimal practice effects and to show no variation simply as a result of repeating the assessment (test-retest reliability).

(c) Sensitivity: outcome measures in rehabilitation need to be sensitive to change, i.e., able to detect change in ability when change has occurred, and responsive to differences between rehabilitation programs.

(d) Practicality: the selection of outcome measures is dominated by practical constraints. An outcome measure must be short, easy to administer and acceptable to patients. Those measures which are tiring, detailed, intrusive or repetitive will not be tolerated. It must also be easy to communicate the findings to others.

Many studies evaluating the effects of rehabilitation have used single-case experimental designs. Measures are repeated frequently and therefore need to be very short, in order not to induce fatigue and not to interfere with the treatment program. They also need to have minimal practice effects so
that stable baselines can be achieved. Alternative versions of tests may be used but few standardised tests have sufficient alternative versions available to be suitable for monitoring progress in single-case experimental designs. When using a randomised controlled trial to evaluate an intervention, it is necessary to use measures which have been used in other studies, so that comparisons between trials are possible. It is also important to include sufficient patients to be sure that a small difference in outcome has not been missed, yet from a patient’s perspective even a small gain in function may be worthwhile. A common strategy to resolve this is to conduct a meta-analysis of several trials, which is facilitated by the use of a common outcome measure.

**Measures of impairment**

Cognitive rehabilitation is designed to improve cognitive abilities. Most cognitive impairments can be assessed by a range of measures but few have been designed as measures of outcome. Some cognitive assessments are intended as screening devices, to identify cognitive impairments which require further evaluation. Others are diagnostic tools, to detect cognitive impairment and differentiate particular impairments from each other. Assessments for screening or diagnostic purposes may not be suitable measures of outcome. Their validity will be based on their ability to classify and therefore those which give clear cut-off points will be most robust. In contrast outcome measures need to have continuous scales such that they are sensitive to changes in ability rather than whether a patient fits a particular category. Measures of specific cognitive impairments can be found in later chapters and therefore are not reviewed in detail here. However, impairments other than cognitive deficits may affect a patient’s response to a cognitive rehabilitation program, and it is therefore appropriate to assess these. They will not be used as outcome assessments, as few cognitive rehabilitation techniques would be expected to decrease, for example, motor impairments. However, the treatment of visual inattention could be predicted to reduce sensory inattention or visual field deficits. For this reason, such assessments are considered.

Motor function may be assessed using summed indices, such as the Rivermead Motor Assessment (Lincoln & Leadbitter, 1979) and the Motor Assessment Scale (Carr et al., 1985), to indicate the level of motor impairment. The Rivermead Mobility Index (Collet al., 1991) is a comprehensive assessment which includes sitting, transfers, walking and getting up stairs. It has been used as an outcome measure in rehabilitation studies on stroke (Wade et al., 1992) and is sensitive to change in people with multiple sclerosis (Vaney et al., 1996). The Expanded Disability Status Scale (Kurtzke, 1983) and Guys Neurological Disability Scale (Sharrack & Hughes, 1999), despite their names, are predominantly measures of impairments in multiple sclerosis and include aspects of motor function.

Sensory impairment is assessed as part of a clinical examination but there are only a few standardised scales available (Lincoln et al., 1998, Winward et al., 2002). Tactile inattention, proprioception and stereognosis may affect the outcome of cognitive rehabilitation, so it is important that they are assessed. Visual impairment will affect patients’ ability to participate in cognitive rehabilitation but conventional acuity measurement techniques require language skills, and visual field assessment may be confounded by the presence of inattention (Walker et al., 1991).

Cognitive impairments are assessed by a wide range of psychological tests. Assessments are available to determine the severity of deficits in language, perception, memory, reasoning, attention, movement disorders and other cognitive functions. Standardised cognitive tests may assess “pure” levels of impairment specific to one cognitive domain. For example, memory can be assessed in terms of encoding ability or working memory capacity. However, such tests may not be of much value as outcome measures. In order for a cognitive assessment to be used as an outcome measure, it must be
sensitive to changes over time but have minimal practice effects. For example, although recognition memory tests are sensitive to differences between individuals, they are not appropriate for evaluating change unless there are parallel versions available. Many cognitive tests, while reliable over time, will show sufficient improvement simply as a result of practice to make them insensitive to small differences between interventions.

Activities

The main aim of cognitive rehabilitation is an increase in activity, which includes functional, cognitive, emotional and social activities.

- The most important outcomes for cognitive rehabilitation are cognitive activities, yet the measurement of these is poorly developed.
- Independence in activities of daily living and mood measures provide proxy measures for the outcome of cognitive rehabilitation.
- The Extended Activities of Daily Living (EADL) and Functional Independence Measure (FIM) are the most suitable measures of independence in activities of daily living.
- Mood should be assessed on questionnaires designed to detect change.

Limitations in functional activities

The ability to perform functional activities in everyday life may be assessed by scales of activities of daily living (ADL), including basic personal self-care skills and instrumental activities of daily living. The choice of scale is governed by practical considerations and there is little consensus on the ‘best’ measures (Jette & Haley, 2005). Most ADL scales consist of summed indices, but concerns have been expressed about treating such scales as ordinal, though some have been demonstrated to be acceptable using Guttman scaling, e.g., EADL and Barthel, or a Rasch model, e.g., Barthel and FIM.

The Barthel Index (Mahoney & Barthel, 1965) is a widely used measure of personal activities of daily living which has almost become the gold standard for stroke rehabilitation studies. It is sensitive to differences between rehabilitation interventions (Indredavik et al., 1991) and reliable when administered verbally or by post. It is sensitive in the inpatient phase of rehabilitation (Houlden et al., 2006) but less sensitive to change in community-based patients. The original index was scored on a 0–100 scale, but this implies a spurious degree of accuracy, and the revised 20-point version (Collin et al., 1988) has become the standard. The main limitation is its ceiling effect; therefore, rehabilitation studies may include a measure of instrumental activities of daily living in addition to the Barthel Index.

Two widely used measures of instrumental activities of daily living are the Nottingham Extended Activities of Daily Living (EADL) scale (Nouri & Lincoln, 1987) and the Frenchay Activities Index (Holbrook & Skilbeck, 1983). Each includes domestic activities such as preparing meals and washing up, mobility outside the home, leisure and social activities. Both scales have been found to be sensitive to differences between rehabilitation interventions (Forster & Young, 1996; Fuller et al., 1996). The EADL is suitable for multi-center studies as it has been validated for postal administration. Cognitive rehabilitation programs are more likely to have an effect on instrumental activities of daily living than on personal self-care skills. Therefore, these scales may be appropriate for assessing the generalisation of cognitive retraining to daily life skills.

The main measure of personal ADL used with a wide range of patients is the Functional Independence Measure (FIM) (Granger et al., 1986). This covers personal self-care and motor activities and has cognitive items on comprehension, expression, social interaction, problem solving and memory. Inter-rater reliability has been found to be higher for physical disability than communication and social cognition sections (Brosseau & Wolfson, 1994; Kidd et al., 1995; Pollak et al., 1996). Because the effect of a cognitive rehabilitation program is most relevant to these items, further improvement to the reliability is needed. Rasch analysis has indicated that the motor and cognitive items form two
distinct scales, though this has been questioned by Dickson & Kohler (1996), who identified six factors from a factor analysis of the FIM. Differential item functioning has also been identified in different patient groups (Dallmeijer et al., 2005), particularly in the motor domain and disordered thresholds, leading to the suggestion that the number of response categories should be reduced (Dallmeijer et al., 2005; Nilsson et al., 2005). The cognitive scale has shown ceiling effects in people with multiple sclerosis (van der Putten et al., 1999) and administration requires prolonged observation of the patient, which means the scale would be difficult to administer in the context of a randomised controlled trial by an independent assessor. Houlden et al. (2006) demonstrated comparable responsiveness between the Barthel and the FIM.

The Functional Assessment Measure (FAM), which was developed to assess the specific problems of brain-injured patients (Hall et al., 1993), contains additional items emphasising cognitive, communicative and psychosocial function. McPherson et al. (1996) found high inter-rater reliability, but greatest discrepancies occurred on cognitive, communication and behavioral items. Hobart et al. (2001b) compared the psychometric properties of the Barthel, FIM and FIM+FAM and found them similar, and highlighted that the FIM and FIM+FAM have significant redundancy of items and confer few advantages over the shorter simpler Barthel for patients receiving inpatient rehabilitation. For the FIM+FAM to be used to evaluate the outcome of cognitive rehabilitation, further work is needed to check the reliability and sensitivity.

The Rivermead Head Injury Follow-Up Questionnaire (Crawford et al., 1996) was developed as an outcome measure for patients with mild to moderate head injury. It is short, simple and can be administered by post or at interview. It was found to have good inter-rater reliability, to be sensitive to changes over time and to detect differences in outcome in a randomised controlled trial (Wade et al., 1997). The Brain Injury Community Rehabilitation Outcome 39 (BICRO39) (Powell et al., 1998) covers aspects of personal and social functioning for brain injury patients living in the community. It has good test-retest and inter-rater reliability and has also been found to be sensitive to the effects of intervention (Powell et al., 2002).

Measures of activity for people with multiple sclerosis seem to be few. The FIM (Brosseau & Wolfson, 1994) and Assessment of Motor and Process Skills (AMPS) (Doble et al., 1994) and Functional Assessment of MS (Cella et al., 1996) have been used, but there are few data to indicate the most appropriate measure for this group. The AMPS requires patients to perform tasks and so takes longer, but it covers instrumental activities of daily living, which may be more important to assess than personal activities of daily living, when the effects of cognitive rehabilitation are being evaluated and has adequate reliability (Doble et al., 1999).

In addition there are general rehabilitation outcome measures, which are predominantly measures of activity limitation. The Sickness Impact Profile (Bergner et al., 1981), the British version of which is the Functional Limitations Profile (Charlton et al., 1983), assesses the impact of sickness on daily activities and behavior. It provides subscales in 12 areas, including ambulation, body care and household management. It is lengthy and complex to administer.

**Mood**

Although mood disorders might be considered either as an impairment, i.e., a direct consequence of some underlying pathology, or as a consequence of impairments, in the context of cognitive rehabilitation they are probably best considered as emotional disabilities. Many mood scales are available but few have been validated for patients with neurological disorders. Those which have items affected by physical disability are likely to be insensitive to mood changes. Many mood questionnaires were developed as screening devices to detect significant levels of depression or anxiety. To evaluate the outcome of cognitive neurorehabilitation, measures need to be sensitive to change and therefore not all
These scales have good validity and reliability, but their sensitivity to change in response to cognitive rehabilitation has not been established.

Limitations in cognitive activities

Limitations in cognitive activities are a major concern for patients and carers and are primary targets for cognitive rehabilitation programs. (For further discussion see Cicerone, Chapter 7, this volume.) Outcome is assessed in several ways, including semi-structured interviews, questionnaires or patient observation. The most common strategy has been to employ a questionnaire that includes items on the behavioral manifestations of cognitive impairments. These may be completed by patients or carers, to determine the subjective effects of cognitive impairment on daily life. These are important indicators of the outcome of cognitive neurorehabilitation, but there are few scales available and not all possible cognitive deficits are included. Another problem is the reliability of eliciting this kind of information from individuals who may not be able to judge their functioning accurately due to their cognitive deficits. For example, frontal lobe damage after brain injury results in impaired metacognitive processing (Hanten et al., 2000), patients with MS underestimate their memory problems on questionnaires (Beatty & Monson, 1991) and patients with memory problems and epilepsy, tended to overstate memory problems on questionnaires, compared with “objective” tests of memory (Piazzini et al., 2001). Semi-structured interviews may provide valuable information regarding the patients’ experience of cognitive rehabilitation, but may not be accurate measures of outcome.

Ecological validity

Cognitive assessments can use ecologically valid tests that assess cognitive functions in the context of everyday tasks. For example, the Rivermead Behavioral Memory Test (RBMT) (Wilson et al., 1985) has tasks, such as remembering where an
object was placed in a room or remembering to do things at appointed times. This task-based performance is related to a number of cognitive functions, and not a specific one, but provides a clinical assessment which approximates the patient’s functioning in everyday life. Two approaches to ecological validity (Franzen & Wilhelm, 1996) have been adopted. One is developing tests with high face validity which simulate daily tasks (requiring the underlying cognitive functions to complete these tasks to be intact). The other relates performance on pre-existing (traditional) tests to daily functioning (Chaytor & Schnitter-Edgecombe, 2003). Assessments of cognitive function which are designed to reflect the cognitive skills needed in everyday life, such as the RBMT (Wilson et al., 1985), Behavioral Inattention Test (BIT) (Wilson et al., 1987), Behavioral Assessment of the Dysexecutive Syndrome (Wilson et al., 1996) and Test of Everyday Attention (TEA) (Robertson et al., 1994), maybe considered to measure limitations to cognitive activities rather than impairment. They include items which are likely to be predictive of everyday performance rather than assessing everyday performance itself. Although they have the advantage of ecological validity, unlike many assessments of cognitive impairment, they are probably not true measures of activity limitation, because they comprise artificial activities and not those which people necessarily perform on a daily basis. Ecologically valid tests of memory and attention (e.g., RBMT and TEA) have been found to be better at predicting functional disability than memory questionnaires (Higginson et al., 2000). However, there is a trade-off when developing ecologically valid tests: the more they approximate real-world scenarios, the less structured they are; and consequently have poor psychometric properties (Van Zomeren & Spikman, 2003). Assessments in the main cognitive domains will be considered.

Memory

Subjective memory impairment has been investigated using the Everyday Memory Questionnaire (EMQ) (Sunderland et al., 1983) in studies of stroke (Tinson & Lincoln, 1987), head injury (Sunderland et al., 1984) and MS (Taylor, 1990). The questionnaire has been used both for patients to assess their own problems and for relatives. Five factors, reflecting the underlying memory processes, have been identified in healthy individuals: retrieval, task monitoring, conversational monitoring, spatial memory and memory for activities (Cornish, 2000). The EMQ has validity in that it correlates moderately with tests of memory, and its reliability is acceptable, though not good. The Subjective Memory Assessment Questionnaire (Davis et al., 1995) is short and has been validated for stroke patients. The Memory Failures Questionnaire (Gilewski et al., 1990) contains more items in four subscales, including one on the use of mnemonics. There is conflicting evidence on the extent to which it correlates with prospective memory (Kinsella et al., 1996; Zelinski et al., 1990).

In general, memory questionnaires appear to have adequate reliability, but low validity, particularly when completed by patients (Ruisel, 1991). To complement memory questionnaires, and to compensate for some of their limitations, memory “performance” tests are frequently used. Such tests, which approximate real-life scenarios, may be more valid measures of rehabilitation outcome than traditional memory tests. The Rivermead Behavioral Memory Test (RBMT) (Wilson et al., 1985) was developed to assess everyday memory problems in individuals with acquired brain damage. The extended version (RBMT-E) (Wilson et al., 1998), with two parallel forms, is sensitive to milder memory deficits. It has age norms (years 11–95), and has been translated into many languages.

Visual neglect

The behavioral manifestations of visual neglect have also been assessed by questionnaires. Towle & Lincoln (1991) developed the Problems in Everyday Living Questionnaire as a subjective measure of visual neglect. The patient has to report how often problems, such as bumping into door frames and making errors when dialling the
telephone, have occurred. The Catherine Bergego Scale (Azouvi et al., 1996) contains ten items which the patient has to rate according to their severity. It has been found to have good inter-rater reliability and validity. Test-retest reliability and sensitivity to change need to be checked. Neither scale has been demonstrated to be sensitive to differences between interventions. An alternative approach has been to ask patients to carry out practical tasks and to observe their performance. Zoccolotti et al. (1992) described a scale which differentiated tasks involving the exploration of external space, dealing cards and serving tea, from those which related to one’s own body, using a comb or razor. The scales were found to have high inter-rater reliability and internal consistency, and concurrent validity in relation to conventional impairment measures.

An ecologically valid test of unilateral visual neglect that reflects patients’ daily life is the Behavioral Inattention Test (BIT) (Wilson et al., 1987). This short, easy to administer test has six conventional tests (such as line cancellation) and nine behavioral tests (such as telephone dialling). Hartman-Maeir & Katz (1995) found construct and predictive validity of most of the behavioral subtests as functional measures of neglect. To assess the effects of domain-specific interventions, tests of neglect for body space, peripersonal space and loco-motor space may be used. Robertson et al. (1998) used a variant of the Hair Combing Task (Zoccolotti & Judica, 1992) to measure neglect of body space, the Baking Tray Task (Tham & Tegnér, 1996) to assess peripersonal space and a tailor-made navigation task to assess locomotor neglect. The Shapes Task (Maddicks et al., 2003), in which the patient has to name 20 shapes on a wall three meters away, has also been used for this purpose. Although some neglect tests have been designed for diagnosis, imaginative adaptations can render them suitable outcome measures.

Attention

discovery, lack of control (focused and divided attention) and poor sustained attention. Some traditional tests, e.g., the Stroop Color Word and Trail Making, have been used to assess these, but as tests of impairment, they can only be used as proxy measures of rehabilitation outcome. An alternative has been to use questionnaires.

Changes associated with attention training have been assessed with the Attention Questionnaire (Sohlberg et al., 2000), the Dysexecutive Questionnaire (Wilson et al., 1996) and the Attention Rating and Monitoring Scale (ARMS) (Cicerone, 2002). The latter consists of 15 items measuring concentration, mental effort, and cognitive symptoms associated with attentional difficulties. An observational rating scale, the Moss Attention Rating Scale (MARS) (Whyte et al., 2003) is completed by occupational and physical therapists. The inter-rater reliability is good but further evaluation of reliability and validity is needed.

The Rating Scale of Attentional Behavior (Ponsford & Kinsella, 1991) showed moderate correlations with neuropsychological measures of attention, good internal consistency and intra-rater reliability but agreement between raters working in different contexts was less satisfactory. Scores showed change over time with treatment but the correspondence to neuropsychological measures of attention was low. Discrepancies seemed to occur as a result of emotional factors and expectations of the therapists. This highlights the difficulty of validating such scales, and Ponsford & Kinsella (1991) suggested that more concrete descriptions of scale items might reduce this subjectivity.

The Test of Everyday Attention (TEA) (Robertson et al., 1994) is an ecologically valid test of selective attention, sustained attention and attentional switching. It consists of eight subtests that approximate attentional tasks carried out by people in daily life, such as listening to winning numbers in a lottery, searching telephone directories and scanning maps. It takes about 60 minutes to complete and normative data exists for ages 18–80. It has three parallel forms and high test-retest reliability. Discriminative validity has been established in brain injury (Chan, 2000).
Executive functions

People with impairment of executive function display problems in initiating and stopping behaviors, shifting set, paying attention and being aware of themselves and others. Cognitive rehabilitation of executive dysfunctions has aimed to reduce the barriers to participation and activity limitations. Worthington (2005) recommended that outcomes be measurable as “socially meaningful (as opposed to statistically significant) change” (p. 259). A rating scale for problem-solving behaviors was developed by Von Cramon et al. (1991) to evaluate the behavioral effects of treatment. Aspects of problem-solving behavior were rated according to the frequency of their occurrence. The scale was found to be reliable and sensitive to improvements.

The Behavioral Assessment of the Dysexecutive Syndrome (BADS) (Wilson et al., 1996) assesses problems related to planning, organising, initiating, monitoring and adapting behavior. It comprises six tests which simulate real-life scenarios. The Dysexecutive Questionnaire (DEX) has 20 items on a Likert scale that describe behaviors related to the dysexecutive syndrome. Patient and carer versions exist. Reliability has been evaluated by the authors, and validity by other small studies (Norris & Tate, 2000). Significant correlations have been reported between executive tests, such as Wisconsin Card Sorting (0.37) and phonemic fluency (0.35), and ratings on the DEX (Burgess et al., 1998), and with the Disability Rating Scale (0.52) (Hanks et al., 1999).

Language

The effects of language problems on everyday life have been investigated in detail. The Communicative Activities for Daily Living (Holland et al., 1999) presents language tasks through role play. This is more sensitive to communication strengths than traditional testing but is not a naturalistic observation. The Functional Communication Profile (Sarno, 1969) and Edinburgh Functional Communication Profile (Skinner et al., 1984) provide ratings of everyday language behavior but are very subjective. The Profile of Functional Impairment in Communication (Linscott et al., 1996) contains a detailed analysis of communication skills but requires an experienced assessor. Aphasia batteries have been employed as measures of change over time, but these are not sensitive (Weniger, 1990), with some subtests being sensitive but having few stimuli, and overall scores being too generic (Nickels, 2005).

Assessment of general cognitive functions

In addition there are scales which do not attempt to assess specific cognitive domains but consider cognitive ability in general.

The Level of Cognitive Functioning Scale (LCFS) (Hagen et al., 1972) (also referred to as Rancho Los Amigo Scale) has been used to assess cognitive functioning in post-coma patients. This scale has an 8-level classification of functioning, ranging from “no response” to “purposeful-appropriate.” It has good test-retest and inter-rater reliability and concurrent and predictive validity (Gouvier et al., 1987). However, the categories are broad, making detection of small changes difficult.

The Cognitive Failures Questionnaire (CFQ) (Broadbent et al., 1982) is a 5-point self-rating scale that determines the frequency with which cognitive slips (arising from failures in perception, memory, and motor functions) have occurred, with versions for both patients and significant others. It is less specific to memory and includes the behavioral consequences of other cognitive deficits. It has been used as a measure of outcome of treatment, including light therapy for neuropsychological functions in seasonal affective disorder (Michalon et al., 1997) and cognitive assessment in stroke rehabilitation (McKinney et al., 2002). A similar scale, the Cognitive Log (Cog-Log) (Alderson & Novack, 2003), is suitable for assessing recovery of higher neuropsychological processes with inpatients. The Log assesses verbal recall, attention, working memory, motor sequencing and response inhibition, and is reported to have good inter-rater
reliability and high internal consistency (Alderson & Novack, 2003).

**Limitation in social and occupational activities**

Behavioral and psychosocial problems are common consequences of traumatic brain injury and need to be assessed, particularly in the later stages of rehabilitation (see Dawson & Winocur, Chapter 14, this volume). Assessment procedures have been criticised for their lack of rigorous evaluation (Hall, 1992). The Neurobehavioral Rating Scale (Levin et al., 1987) is based on behavior, symptoms and skills measured in a structured clinical setting. The Neurobehavioral Functioning Inventory can be used with informants to record their perceptions of everyday problems (Kreutzer et al., 1996). Neither scale has well-established reliability or validity nor shown to be a sensitive indicator of rehabilitation outcome. However, the scales would seem to be worth developing further as they tap the activity limitation associated with cognitive impairment.

**Participation**

- There are few measures of participation.
- The best developed measures are the London Handicap Scale and Short Form 36.
- Global measures which include impairment, activity and participation are unlikely to be sensitive to the effects of cognitive rehabilitation.

Although it is unlikely that cognitive rehabilitation will have a direct effect on participation, because multiple factors will determine participation and cognitive problems are only one of many, it may be useful to assess participation as an effect of an overall rehabilitation package. Several measures have recently been developed (Heinemann, 2005) and some overlap with indicators of quality of life. Participation is subjective and inherently more difficult to measure than activity limitation, particularly in people with cognitive impairments.

The London Handicap Scale (Harwood et al., 1994) generates a profile of handicaps in the six survival roles: orientation, physical independence, mobility, occupation of time, social integration and economic self-sufficiency and an overall severity score. Other potential measures include the Community Integration measure (McColl et al., 2001), Craig Handicap Assessment and Reporting Technique (Whiteneck et al., 1992) and MS Impact Scale (Hobart et al., 2001a).

Generic measures of quality of life may also be used as indicators of participation. The Short Form 36 (Stewart & Ware, 1992) is a short questionnaire with good construct validity (Ware et al., 1993). However, there are doubts about its applicability with the elderly (Hill & Harries, 1994) and it has floor effects in physically disabled groups (Freeman et al., 1996). Wade (2003) questioned whether measurement of quality of life was appropriate in the context of rehabilitation.

**Future developments**

- The psychometric properties of most measures need further evaluation.
- Activity measures are needed for the evaluation of outcome in single case experimental design studies.
- Researchers conducting randomised control trials should attempt to reach consensus on a few standard activity measures in order to facilitate meta-analyses.

For each of the measures mentioned, further work is needed to establish the validity and reliability of the scale. In particular, the validity is often only established with one diagnostic group. Validation studies should be carried out in several groups of patients to confirm the underlying construct. The reliability needs to be checked in a variety of situations (inpatient, outpatient, hospital, community), conditions and over a variety of time intervals. For most scales this task has hardly yet begun. Sensitivity to the effects of intervention will not be established until there are far more efficacy studies. Most single case
experimental design studies use measures of impairment to assess the effect of intervention. However, it is the effect on daily life that is of most concern to patients and their families. Therefore, activity measures need to be developed for use in this context. Few of the measures described above are sensitive to the small changes in ability that need to be detected and many are too long to be administered with the frequency that is necessary in single case experimental design studies. The alternative approach to treatment evaluation is the randomised controlled trial. Several measures described above have been found to be sensitive to differences between rehabilitation procedures in randomised controlled trials. However, there have been few well-designed, methodologically sound trials of cognitive rehabilitation (see Cicerone, Chapter 7, this volume; Lincoln & Bowen 2006). It is hard to recruit sufficient patients within a single center; therefore multi-center trials of cognitive rehabilitation are needed. This requires consensus about which outcome measure to use. In addition, consistency of outcome measures is important for meta-analysis. The main way forward, therefore, seems to be to agree on a group of outcome measures suitable for trials of cognitive neurorehabilitation. These measures need to be refined in terms of their psychometric properties. If this is achieved in the context of research, it will also then be possible to use the measures for the audit of clinical services and the evaluation of the progress of an individual patient in a rehabilitation setting.

Conclusions

Standardised measures are available for evaluating the outcome of cognitive neurorehabilitation. These include measures at the levels of impairment, activity and participation. At each level, the measures chosen should be reliable, valid, sensitive to the effects of intervention and consistent with those used by other researchers. There is a particular need for the development of measures of limitations in cognitive activities.

REFERENCES


patients with unilateral visual neglect. *Clinical Rehabilitation*, 5, 135–140.


