APPENDIX F
LITERATURE REVIEW INSTRUMENTS

October 2003

Development of Quality Indicators for Inpatient Rehabilitation Facilities

Literature Review

Prepared for

Jean C. Scott, Dr.P.H., R.N.
Rita Shapiro, M.A., P.T.
Centers for Medicare & Medicaid Services
CBC/QMHAG/DAPAC, S3-02-01
7500 Security Boulevard
Baltimore, MD 21244

Prepared by

RTI International
Health, Social, and Economics Research
Research Triangle Park, NC 27709

RTI Project Number 07964.004.004

RTI Project Number
07964.004.004

Development of Quality Indicators for Inpatient Rehabilitation Facilities

Literature Review

October 2003

Prepared for

Jean C. Scott, Dr.P.H., R.N.
Rita Shapiro, M.A., P.T.
Centers for Medicare & Medicaid Services
CBC/QMHAG/DAPAC, S3-02-01
7500 Security Boulevard
Baltimore, MD 21244

Prepared by

RTI International
Health, Social, and Economics Research
Research Triangle Park, NC 27709

*RTI International is a trade name of Research Triangle Institute.

Acknowledgements

This report was prepared under contract 500-00-0024 by Leslie Greenwald, Ph.D.; Lee Mobley, Ph.D.; Claudia Squire, M.S.; Diamante Tornatore, M.S.P.H.; and Shula Bernard, Ph.D.

CONTENTS

Section **Page**

Cognitive Functioning

Cognitive Functioning

Although some studies found no significant relationship between cognition and recovery of function, numerous other researchers report that cognitive status does affect function (Resnick and Daly, 1997).

Of the 96 articles selected for review that covered approximately 27 distinct instruments, 24 were chosen for more in-depth study. Selection from the final Medline search of instruments for further examination was based on the amount of literature available (i.e., those that have been reviewed extensively were generally felt to have face and content validity) and/or the persons with whom the instruments were tested (i.e., elderly or rehabilitation-specific populations). Articles for inclusion were not limited solely by population-type because this would have excluded valuable information. In addition, upon the advice of Dr. Margaret Stineman, appropriate members of our technical expert panel (TEP) and national organizations were contacted and questioned regarding the cognitive assessment instruments currently used in the field and/or clinical guidelines for the assessment of cognitive readiness for rehabilitation, problem solving functional skills, daily cognitive tasks or executive functioning. These national organizations included the American Occupational Therapy Association (AOTA), American Physical Therapy Association (APTA), American Psychological Association (APA) Division 22**,** American Speech-Language-Hearing Association (ASHA), American Therapeutic Recreation Association (ATRA), and the National Therapeutic Recreation Society (NTRS).

The general consensus among those contacted was that there is no general consensus regarding the use of cognitive assessment instruments. As one person wrote, "there are no universally accepted instruments or procedures/guidelines at present." Another stated that the test chosen depends on the individual patient and so a wide variety of instruments are used on a regular basis, all of them modified as deemed necessary. It was reported that the Mini-Mental State Examination, the Cognitive Capacity Screening Examination, the Cognistat (also known as the Neurobehavioral Cognitive Screening Examination), the Assessment of Language-Related Functional Activities, Ross Information Processing Assessment-Geriatric, the Mattis Dementia Rating Scale, the newest versions of the Wechsler Scales (WAIS-III and the Wechsler Memory Scale), and the Ravens Coloured Progressive Matrices are currently utilized in the rehabilitation field. Another suggested that cognitive assessment items might be selected from the measures developed by Dr. David Loewenstein.

The articles reviewed are seminal studies and included all of the information needed to populate the accompanying matrix and overview manuscripts that provide a context for the examination of the selected instruments. The instruments ultimately selected for closer examination include:

- Mini-Mental State Examination (MMSE)
- Functional Independence Measure (FIM)—cognitive items N-R
- Cognitive Capacity Screening Examination (CCSE)
- Lowenstein Occupational Therapy Cognitive Assessment (LOTCA)
- Direct Assessment of Functional Status Scale (DAFS)
- Executive Interview (EXIT)

- Clock Drawing Task 1 and 2 (CLOX)
- Cognitive Impairment Diagnosing Instrument (CIDI)
- Cognistat/Neurobehavioral Cognitive Status Examination (NCSE)
- Assessment of Language-Related Functional Activities (ALFA)
- Ross Information Processing Assessment-Geriatric (RIPA-G)

Mini-Mental State Examination (MMSE) [14, 16] (E)	
Domain	Cognitive Functioning
Purpose (constructs measured, target population)	This instrument measures cognitive function among a wide range of subjects [14] and cognitive status among post-stroke patients in geriatric rehabilitation [16].
Description	
Number and types of questions	11 open-ended and performance-based items with a maximum score of 30. These items are grouped into 7 categories of cognitive domain/function: Orientation to time (5 points); Orientation to place (5 points); Registration of three words (3 points); Attention and Calculation (5 points); Recall of three words (3 points); Language (8 points); Visual Construction (1 point).
Method of administration	In-person interviews.
Time to administer	5 to 10 minutes [14].
Instrument has been tested with the following populations	Tested with a wide variety of subjects, ranging from individuals with various types of dementing illnesses causing severe cognitive impairment to community residents who are cognitively intact [14, 16]
Sensitivity to measure changes in status	Several longitudinal studies with test-retest periods ranging from 1 month to 3 years revealed a significant decline in MMSE scores over time among dementia patients. Alzheimer's patients had a substantial degree of variability due to uneven progression of AD and psychometric properties of the MMSE. MMSE becomes less sensitive to the progressive decline of functioning associated with AD for severely demented patients [14]. All patients performed better at 6 month post-stroke than 2-8 weeks post-stroke in absolute MMSE scores [16].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Used in screening for cognitive impairment severity [14, 16] and/or serial documentation of cognitive change [14].
Used for population estimates, facility-level estimates or individual level	Individual level [14, 16].
Used for quality improvement, quality monitoring, quality measurement	No [14, 16].

Note = Reference information for cognitive functioning domain begins on page 107.

Validity/Reliability	
Internal consistency, list relevant statistics	**(MMSE)** Among a mixed group of medical patients, the alpha level was 0.92 (all correlationswere significant at $p<0.05$). However, the alpha levels ranged from 0.51 to 0.73 among community samples. Two month test-retest reliability coefficients fell between 0.76 and 0.91 for both cognitively unimpaired and impaired individuals. Delirium patients' coefficient of 0.53 is attributed to the changing course of the illness [14].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	MMSE sensitivity results among the studies examined for this meta-analysis ranged from 20% (Pfeffer et al.) to 96% (Folstein et al.). • The MMSE lacks sensitivity to mild cognitive impairment and fails to adequately discriminate between normal patients and those with mild AD. • Individual items of the MMSE are differentially sensitive to disease severity. [14]. • For general neurology and psychiatry patients, the sensitivity of the MMSE is usually low (21% to 76%). The MMSE is fairly insensitive to right hemisphere damage, most likely due to its bias toward verbal items, thus causing an increase in false negatives. The positive predictive values are also highly variable, ranging from 31% to 100%. Lowered specificity was found when the comparison group included psychiatric patients [14]. The correlations between MMSE and neuropsychological test scores were modest-to-high. Decreased independence is shown to be related to lower MMSE scores by correlations of 0.39 and 0.71 between MMSE scores and ADL scaled scores. Higher correlations with IADLs, have been shown in several studies. MMSE scores are more sensitive to declines in functional behavior that is more cognitively demanding than they are to physical health and mobility [14]. Several longitudinal studies revealed a significant decline in MMSE scores over time among dementia patients and those with severe AD [14]. MMSE scores were significantly lower for the stroke patients than the controls ($p<0.001$). MMSE, revealed sensitivity of 53%, specificity or 76%, a false positive ratio of 24 and a false negative ratio of 33. [16].
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	None of the MMSE items are currently in the IRF-PAI. IRF-PAI items #25, 26, and N, O, P, Q, and R of the FIM/IRF-PAI items, which also measure cognitive status/function, might be supplanted.

How well does this instrument distinguish people compared with existing questions on IRF-PAI	Among moderately demented patients, the verbal items make the MMSE sensitive to a profound decline in memory, but they lose their discriminative ability among more severely demented patients. As such, as the lower end of the scale is approached, the MMSE becomes less sensitive to the progressive decline of functioning associated with AD [14].
Limitations	Both sensitivity and specificity are affected by number of years of education. MMSE scores decrease with increasing age. MMSE scores have been shown to be affected by social class, socioeconomic status, years of education, and age [14]. The content of the MMSE lacks sufficient visuospatial and/or constructional praxis measures and is instead highly verbal and perhaps overly simplistic [14].
Other Comments	In at least one of the studies reviewed, the Modified MMSE was found to have greater reliability and validity than the MMSE [14]. Addition of more spatial items on the MMSE might improve its validity when used with stroke patients [16].

Functional Independence Measure and/or Functional Assessment Measure (FIM/FAM) [34, 37, 38, 86] (H)	
Domain	Cognitive Functioning
Purpose (constructs measured, target population)	Functional independence across motor and cognitive domains among rehabilitation patients [37, 86], or mature adults in multilevel care residences [38].
Description	
Number and types of questions	18 items, each scored on a scale of 1-7, the highest score reflecting complete independence. Possible scores range from 18 to 126. Motor domain items include eating, grooming, bathing, dressing upper body, dressing lower body, toileting (all self care); bladder and bowel management (sphincter control); bed, chair, wheelchair, toilet, tub, shower (transfer); walk/wheelchair, stairs (locomotion). Cognitive domain items include comprehension and expression (communication), social interaction, problem-solving, and memory (social cognition) [37, 86, 38 (excludes toileting)].
Method of administration	Clinical observation and in-person interviews/self-report conducted at admission and in some cases also at discharge. Different portions of the FIM are scored by different members of the rehab team [37, 38, 86]
Time to administer	10 to 20 minutes [34]
Instrument has been tested with the following populations	Tested in the following populations: elderly, stroke, residing in SNF and assisted living, and NWART facilities.
Sensitivity to measure changes in status	Sensitivity was not assessed, but cross-sectional discriminative ability was demonstrated by the differences in FIM score patterns across impairment categories [37]. Sensitivity to change was not assessed, but there were statistically significant differences between the 3 groups on both the motor and cognitive measures thus showing concurrent validity [38]. Across all impairment groups, significant improvements ($p < 0.0005$) were seen between admission and discharge scores[86]. Results of the FIM were used to document improvements in various impairment groups over time [86]. FIM also found to detect cross-sectional discriminative ability across impairment categories [37, 38].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment, case-mix [37]. Assessment (of level of assistance required/burden of care) [38].
Used for population estimates, facility-level estimates or individual level	Individual level [37, 38, 86].
Used for quality improvement, quality monitoring, quality measurement	No [37, 38, 86].

Validity/Reliability

Internal consistency, list relevant statistics	**(FIM/FAM)** Cronbach's alphacorrelations ranged from 0.84 to .0.93 for the total FIM across the 20 impairment categories. For the motor-FIM, the range was 0.82 to 0.93. For the cognitive FIM , it was 0.82 to 0.91. There was no dependence by the total FIM or the 2 subscales on any single item for cohesiveness as a test [37]. For the motor subscale there was high test-retest reliability with an ICC of 0.9. The test-retest reliability of the cognitive subscale was good (ICC = 0.8). Among the subjects in skilled nursing facilities, the test-retest reliability was higher for the motor subscale (r = 0.9) when compared with the cognitive subscale (r = 0.6). The item difficulty calibrations revealed that those for upper body dressing (-.27), toilet transfers (-.28), and bed/chair/wheelchair transfers (-.33) were nearly the same. [38]. Nearly all Cronbach's α on the FIM and its subscales were high, thus indicating that the FIM's individual items are highly correlated and that the FIM is a reliable/internally consistent instrument. The α for overall admission FIM scores was 0.89, and 0.91 for the overall FIM discharge score. It was also found, however, that there was low internal consistency (overall Locomotion discharge scores α = 0.64) for the Locomotion category, which consists of the ambulation and stair walking items, particularly for subjects with spinal cord injury (α = 0.40) or amputations (α = 0.34). In contrast, the Communication category demonstrated high internal consistency (overall α = 0.89) [86].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	More than 91% of the motor-FIM items were at least moderately related (r ≥ .39) to the total score calculated from other items in the motor group within each impairment category. All cognitive-FIM items were related (r ≥ .39) to the total score computed from the other items in the cognitive group. In addition, factor analysis revealed that in 16 of the 20 impairment categories the motor and cognitive domains were perfectly distinguished. Operational definitions for problem solving and comprehension influenced FIM ratings because problem solving includes the ability to solve financial problems in addition to being able to solve complex problems. As such, those subjects who had given others authority to manage financial matters received lower scores. With comprehension, its operational definition is "understanding of auditory or visual communication" and so a "modified independence" score (a 6 on the FIM scale) was obtained if subjects needed a hearing or visual aid (excluding normal spectacles). More than half of the subjects in this study used a hearing aid thus possibly explaining why the second most difficult item on the cognitive subscale was auditory comprehension. The FIM seems to have broad discriminative ability, but may not be as effective in detecting more subtle important differences. First of all, as hypothesized, FIM scores were lower for older subjects or those with comorbid conditions, thus demonstrating its capability to discriminate on the basis of age and comorbidity. Those without comorbid conditions, on average, achieved higher FIM discharge scores (93 vs. 91, p < 0.005) than those with comorbid illness. Likewise the overall mean FIM discharge score for those under age 43 was 101 while it was 88 for those over age 71 (33% of the sample population); such decreases with increasing age were consistently revealed across all but one of the impairment categories.

Construct validity & description of relationships with existing IRF-PAI (FIM) questions (continued)	Furthermore, among spinal cord injured patients with different severity impairment levels there were statistically significant differences ($p < 0.005$) in discharge FIM scores (i.e., mean FIM of 105 for incomplete paraplegic subjects vs. 93 for complete paraplegic subjects vs. 91 among incomplete quadriplegic subjects vs. 68 for complete quadriplegic subjects). As hypothesized, patient discharge destinations caused a variation in FIM scores. In fact, the greatest differences in discharge FIM scores were associated with discharge destination (e.g., among all impairment categories, greater independence was demonstrated by FIM scores of subjects discharged to home than those discharged to other destinations). Unpredictably, subjects with double above the knee amputations had higher average FIM discharge scores than those with single below the knee amputations although the difference was not statistically significant. In evaluating stroke patients, the FIM also has discriminative ability. The mean FIM admission and discharge scores significantly differed ($p < 0.005$ and 0.05, respectively) between right-body involved stroke patients (FIM admission score = 1) and left-sided stroke patients (FIM admission score = 64). Such differences were revealed only among the FIM's Communication items [86].
Ceiling or floor effects	For most items and patients, floor and ceiling effects were relatively small.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	These are the cognitive measures used on the IRF-PAI (i.e., question 33, items N, O, P, Q, and R).
How well does this instrument distinguish people compared with existing questions on IRF-PAI	The same.
Limitations	Small sample, selected from a specific retirement community, FIM training methods were specific to this study, the toileting item was excluded, and the investigator knew the residential situations/level of care provided for all 3 subject groups [38]. Inter-rater reliability was not assessed and may be a factor because the FIM is usually not scored by a single evaluator but rather by different team members. Furthermore, a bias may be introduced when a team member who treats the patient then also scores the FIM rather than a more objective evaluator. The comorbidity measure used in this study (the Charlson Comorbity Index) is not the best one for the rehab population and there were no recorded comorbid conditions for many of the subjects. The FIM assesses disability or performance impairment by measuring burden of care, but does not measure handicap (disability which causes a socially defined disadvantage). FIM does not capture the social, psychological, or vocational impact and only reveals minimal functional limitation [86]. Assessment of individuals who have suffered cerebrovascular accidents, traumatic brain injuries, or other cognition-affecting impairments may be more difficult with the FIM because it has few cognitive and community-related functional items relevant to such persons [91].

Other Comments	The motor and cognitive domains are nested within the global concept of burden of care that is measured by the total FIM and are more appropriate than the total FIM score for answering more clinically focused questions about general types of disabilities. The ability of these subscales to generalize across heterogeneous impairments allows direct comparisons of patterns of disability across groups of patients who are clinically different [37]. Because FIM scores were different on both subscales for the 3 different groups of subjects in accordance to the 3 different levels of care provided in their respective residential communities, it would seem that the FIM measures levels of assistance (which may be thought to be synonymous with burden of care, which the FIM purports to measure) [38]. The FIM has broad discriminative abilities: FIM scores decrease with increasing age or comorbidity, they vary with discharge destination, they decrease in accordance with increased injury severity, they decrease with ascending amputation severity, they were greater among left body-involved stroke patients than among right body-involved patients. However, it is unclear if the FIM is adequately sensitive to more subtle yet clinically important differences. The FIM is a good generic indicator of disability [86].

Cognitive Capacity Screening Examination (CCSE) [71]	
Domain	Cognitive Functioning
Purpose (constructs measured, target population)	Mental status (as the most important predictor of cognitive impairment) in elderly hospitalized patients [72]. Cognitive impairment in a meta-analysis of the five most frequently cited bedside cognitive screens [73].
Description	
Number and types of questions	30-items designed to "diagnose diffuse organic mental syndromes in nonpsychiatric patients" [72]. The content areas measured include orientation, memory, attention, calculations, and concept formation, with scores ranging from 0-30 [71]. Scores of less than 20 are indicative of cognitive impairment [72, 73].
Method of administration	In-person interviews [71, 72, 73]. Subjects required to speak and understand English [72].
Time to administer	5-15 minutes [72, 73].
Instrument has been tested with the following populations	Tested with elderly medical-surgical patients [72], hospital staff, and non-patient community residents [73] all over the age of 65 years, with the ability to speak and understand English, and with the ability to answer a 30-40 minute questionnaire (other instruments were also used in this study for validation purposes.
Sensitivity to measure changes in status	Sensitivity to change was not assessed [72, 73].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment of mental health status [72] and the presence and severity of cognitive impairment [73].
Used for population estimates, facility-level estimates or individual level	Individual [72, 73].
Used for quality improvement, quality monitoring, quality measurement	No [72, 73].
Validity/Reliability	
Internal consistency, list relevant statistics	**(CCSE)** The CCSE was revealed to have high internal consistency ($\alpha = 0.965$) [72]. The only reliability information reported was 96% inter-rater reliability across six subjects, using three examiners [73].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	In terms of content validity, the CCSE was found to be the most comprehensive because it measures 9 of the 11 components of mental status the CCSE scores were correlated with the APA clinical diagnoses of global cognitive impairment. The Spearman correlation coefficient was 0.83 (again, higher than that found for the other 2 instruments), and was statistically significant at $p < 0.001$. Sensitivity, predictive value of a positive test, specificity, and the predictive value of a negative test were all 1.00 thus demonstrating excellent sensitivity

	and specificity. Convergent validity was determined by calculating
Construct validity & description of relationships with existing IRF-PAI (FIM) questions (continued)	intercorrelations among the scores obtained with the CCSE and the other instruments, resulting in high Pearson product moment correlations (ICC = 0.60 with the SPMSQ, and ICC = 0.84 with the MMSE) that were statistically significant at $p < 0.001$ [72]. In one study with 24 subjects (18 diagnosed with organic mental disorders), there was significant correlation between scores of <20 and clinical diagnoses of organic mental disorder, with one false-positive and one-false negative (for a sensitivity of 91% and specificity of 82%). It was found in another study of 25 psychiatric inpatients that 84% of those with functional diagnoses obtained scores of ≥ 20. In a study involving 62 hospitalized elderly and 57 community dwelling elderly, those with diagnosed mental disorders all scored <20 (96% sensitivity), as did more than 71% of those who had presented with stroke or hip fractures. False negative rates of 33% and 46% were found in two studies, one of which revealed that only 2 of 7 patients with unilateral right hemisphere damage scored below 20. In another study, six of seven patients with abnormal EEGs scored <20 (one false-negative) and two of three with normal EEGs scored ≥ 20 (one false-positive), for a sensitivity of 84% and a specificity of 71%. Yet another study demonstrated that scores decline with increasing age [73].
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	IRF-PAI items #25, 26, and N, O, P, Q, and R of the FIM/IRF-PAI items, which also measure cognitive status/function, might be supplanted.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	Unlike the FIM, the CCSE was not used in any of the reviewed studies to assess physical and cognitive function across impairment groups. As such comparison of discriminative capabilities is not possible [72, 73].
Limitations	Because the instrument contains 30 items and takes longer to complete than some alternatives, it may present too great a burden for individuals with moderate to severe cognitive impairment [72]. Too many false negatives because (1) a unidimensional scoring system is used and so cognitive dysfunction that is limited to one area most likely will not produce an abnormal result, (2) the breadth and depth of item content are limited, and (3) the tests probably are not hard enough for patients with high levels of pre-morbid intelligence and education [73].
Other Comments	TEP member Christine Baron (co-director of Stroke & Recovery Program, and manager in Speech & Pathology Service at Nat'l Rehab Hospital) indicated that this instrument is regularly used in the field.

Loewenstein Occupational Therapy Cognitive Assessment (LOTCA) and Direct Assessment of Functional Status Scale (DAFS)—both developed by Dr. David Loewenstein [41	
Domain	Cognitive Functioning
Purpose (constructs measured, target population)	LOTCA: Cognitive abilities of brain-injured (cerebrovascular accidents—CVAs—and TBIs) patients [42]; relation to functional outcomes of stroke patients [41]. DAFS: Functional capacity of older patients with suspected cognitive impairments [87] and older patients with schizophrenia [92].
Description	
Number and types of questions	LOTCA: 20 subtests across four areas of orientation, perception, visual motor (?) organization, and thinking operations. 22 items with score range from 22 to 87 [41]. DAFS: Assesses 7 functional areas with specific tasks associated with each area. The areas are time orientation, communication, transportation, financial, shopping, grooming, and eating. A total of 106 items, with a total maximum score of 106 points for correct performance [87].
Method of administration	LOTCA: In-person interviews with direct observation of performance (usually conducted by an occupational therapist). No information was provided about literacy level [41, 42, 92]. DAFS: In person interview with direct observation of performance. No information was found regarding literacy level.
Time to administer	LOTCA: 30 to 43 minutes [42]; DAFS: 25 minutes per subject [92].
Instrument has been tested with the following populations	LOTCA: 94 TBI, CVA, and non-brain injured (control group) adults [42]; 1 stroke patients in a geriatric neurologic rehabilitation department in Israel [41]. DAFS: 52 outpatients with schizophrenia and 68 normal control subjects, ages 43-82.
Sensitivity to measure changes in status	LOTCA: Small (6.5 +/- 0.5 SD) but statistically significant ($p < 0.001$) changes in LOTCA scores were observed from admission to discharge [41]. Average improvement was observed when a second assessment was made of both patient groups 2 months later. More improvement was observed for the TBI group than the CVA group [42]. DAFS: sensitivity to measure change in status was not assessed [92].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment [41, 42, 87, 92]
Used for population estimates, facility-level estimates or individual level	Individual level [41, 42, 87, 92]

Used for quality improvement, quality monitoring, quality measurement	No [41, 42, 87, 92].
Validity/Reliability	
Internal consistency, list relevant statistics	**(LOTCA & DAFS)** LOTCA: Interrater reliability was calculated and Spearman's rank correlation coefficients between the raters ranged from 0.78 to 0.93 on the various subtests. Internal consistency by alpha coefficientsfor perception was 0.83, for visuomotor organization 0.91, and for thinking operations 0.81. Among the subtests, the correlation coefficients ranged from 0.39 to 0.76 [42]. DAFS: The interrater reliability was good, with an intraclass correlation coefficient of 0.87 [92].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	LOTCA: Concurrent validity was determined through analyses of the Pearson correlation coefficients for the LOTCA, the MMSE, and the FIM cognitive subscale. They ranged from 0.45 (between the FIM and LOTCA) to 0.63 (between the FIM and the MMSE) and were all statistically significant. The Pearson correlation coefficients of the LOTCA at admission were lower with the FIM cognitive domain and higher with the MMSE, but both were statistically significant [41]. Using the Wilcoxon two-sample test to determine concurrent validity, each patient group was compared to the control group. With the exception of the identification of objects subtest, all of the subtests successfully differentiated (at the 0.0001 level of significance) between the control and each of the patient groups upon the first assessment, with *z* scores ranging from 4.0 to 6.2. At the second assessment, it was demonstrated that the same subtests differentiated the groups above the 0.02 level of significance, with *z* scores ranging from 2.5 to 4.5. In terms of convergent validity, the visuomotor organization area was compared with the Block Design of the Wechsler Adult Intelligence Scale (WAIS) among the TBI group. Between the score on the Block design subtests and the mean score Visuomotor Organization subtests, a correlation coefficient of r = 0.64 was found. Factor analysis revealed that all of the 7 visuomotor organization subtests loaded on factor 1/visuomotor organization, as did the sequencing subtests of the thinking operations domain. Coupled with the sequence subtests, visuomotor subtests accounted for 28% of the variance. Praxis loaded on factor 2/perception along with spatial perception, object identification, and overlapping figures, but shape identification, categorization and classification all loaded on factor 3/thinking operations. Different results/loadings were found for the control group. For example, three of the perception subtests, along with praxis and pictorial sequence, loaded highly on factor 1 and explained 33% of the variance. Among the patient group, the most variance in performance was explained by visuomotor organization whereas among the control group perception explained the most variance, followed by thinking operations. The total amount explained in both groups, was more than 57% [42].

Construct validity & description of relationships with existing IRF-PAI (FIM) questions (continued)	DAFS: The patients with schizophrenia had lower mean scores on both the MMSE and the DAFS than the control group subjects. Upon conducting univariate analyses of variance with Bonerroni corrections, it was found that the patients were significantly more impaired than the normal subjects in the areas of communication, transportation, finance, and shopping ($p < 0.0001$). Differences were not significant in the areas of grooming, eating, or time orientation [92].
Ceiling or floor effects	LOTCA: All of the control group subjects performed well on almost all of the subtests (a ceiling effect) which may explain why the profile of that group and the importance of the factors explaining the variance were so different in the factor analysis [42].
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	None of the LOTCA or the DAFS items are currently in the IRF-PAI. IRF-PAI items #25, 26, and N, O, P, Q, and R of the FIM/IRF-PAI items, which also measure cognitive status/function, might be supplanted.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	LOTCA: The ability of the LOTCA to discriminate between individuals with brain injury and healthy adults was demonstrated by this study [42], and lengthy administration [41]. DAFS: Differences in cognitive function between patients with schizophrenia and those without the disease were demonstrated [92].
Limitations	LOTCA: Small samples, no established relationship between LOTCA scores and functional evaluation and ADL scales [42], lengthy administration [41]. DAFS: This sample of schizophrenic patients was much less impaired than those found in institutionalized settings, the schizophrenic group differed from the control group by age and gender (but controlling for these variables did not affect the analysis of differences between the two groups) [92].
Other Comments	LOTCA: The LOTCA is slightly superior to the FIM cognitive scale and the MMSE as it relates to functional outcome parameters. However, it is lengthy and burdensome to administer, and neither it nor the FIM are significantly better than the MMSE, which requires less administration time and expertise [41]. DAFS: In an effort to increase its sensitivity to earlier dementia, the DAFS is currently being revised. Medication management, food preparation, and taking a telephone message subtests are now included. *TEP member Peter Lichtenberg (director of the Institute of Gerontology and associate professor of psychology at Wayne State University) recommended the DAFS as a "good instrument" from which to choose some possible items for the IRF-PAI.*

Assessment of Language Related Functional Activities (ALFA) [93]	
Domain	Cognitive Functioning
Purpose (constructs measured, target population)	A measure of functional skills on a set of language-related tasks. The ALFA is designed for persons between the ages of 16 and 91 who can understand the directions of the subtests, who are able to formulate the necessary responses, and who have some familiarity with the functional areas assessed.
Description	
Number and types of questions	The ALFA contains 10 subtests: Telling Time, Counting Money, Addressing an Envelope, Solving Daily Math Problems, Writing a Check and Balancing a Checkbook, Understanding Medicine Labels, Using a Calendar, Reading Instructions, Using a Telephone, and Writing a Phone Message. Subtests 1, 2, 3 and 4 are timed. The first 9 subtests contain 10 questions each and the final subtest contains 5 questions.
Method of administration	The ALFA is administered from a durable easel-bound picture book. Answers are recorded in a dichotomous 1-0 fashion.
Time to administer	30 min to 2 hours.
Instrument has been tested with the following populations	471 patients with neurologic traumas and 148 normally functioning adults served as the standardization sample.
Sensitivity to measure changes in status	No studies of stability in absence of treatment, also known as test-retest reliability, appear in the manual. Without test-retest reliability data it is not possible to know if change over time is due to real change or the instability of the test.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment.
Used for population estimates, facility-level estimates or individual level	Individual, but can also be used as evidence of positive outcomes in program evaluation.
Used for quality improvement, quality monitoring, quality measurement	No comments on application for quality improvement noted at this time.
Validity/Reliability	
Internal consistency, list relevant statistics	**(ALFA)** Coefficient alpha provides evidence of internal consistency. For the normative sample, coefficients for the 10 subtests ranged from .76 to .84. Coefficient alphas for six clinical subgroups ranged from .69 to .89. That the six clinical subgroups had similar reliability coefficients was taken as evidence that the ALFA contains little or no bias relative to these groups. Interscorer reliability was assessed with two staff people rescoring 30 protocols. Correlations between the examiners' scoring for the 10 subtests was r = .95.

Construct validity & description of relationships with existing IRF-PAI (FIM) questions	The ALFA manual does not present adequate validity studies to allow judgment. A study of 103 patients involved administration of the Reading Comprehension Battery for Aphasia and the ALFA. Correlations between these two measures were moderate (range = .42 to .67). However, the variance in cognitive ability in the sample of patients taking the test is unknown. Comparing mean scores of the normal sample and the clinical sample showed statistically significant differences on all 10 subtests of the ALFA.
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	No comments noted at this time.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	No comments noted at this time.
Limitations	There are some technical weaknesses including normative sample selection, lack of reported validity studies, lack of support for the claim that the ALFA is an unbiased test, and lack of reported reliability data.
Other Comments	*TEP member Christine Baron (co-director of Stroke & Recovery Program, and manager in Speech & Pathology Service at Nat'l Rehab Hospital) indicated that this instrument is regularly used in the field and is, in fact, a favorite because the norms are more applicable and it is more functionally-based.* ALFA appears to be appropriate as an assessment tool for assessing pretreatment functioning, planning therapy goals, and perhaps for evaluating treatment outcomes.

Ross Information Processing Assessment—Geriatric (RIPA)	
Domain	Cognitive Functioning
Purpose (constructs measured, target population)	Designed to identify, describe, and quantify cognitive-linguistic deficits in the geriatric population following traumatic brain injury.
Description	
Number and types of questions	The battery consists of 10 subtests, 2 supplemental subtests, and a record form with 3 subsections to include background, medical, RIPA-G test, and retest information. Measures various cognitive or linguistic processes including memory, orientation, organization, problem solving, auditory processing, knowledge of general information, reading, and word finding
Method of administration	The test protocol directions expect the examiner to record patient's responses to RIPA-G test items, and to observe and record confabulation, denial, delayed, error, irrelevant, perseveration, partially correct, repetition, self-corrected, and tangential behaviors.
Time to administer	Takes 45 to 60 minutes to administer, with an additional 10 minutes for the two optional supplemental subtests
Instrument has been tested with the following populations	The test was standardized on 84 adults living in skilled nursing facilities. Validation data is very weak.
Sensitivity to measure changes in status	This article does not provide information on re-test data and instruments' ability to measure change in status.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment
Used for population estimates, facility-level estimates or individual level	Individual level
Used for quality improvement, quality monitoring, quality measurement	No
Validity/Reliability	
Internal consistency, list relevant statistics	**(RIPA-G)** Reliability estimate measures, coefficients alpha (composite scores ranging from .84 to .93), and SEMs (composite scores between 3-5) are within acceptable ranges.
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Attempts to establish the test validity were made by seeking professional reviews, by conducting item analysis, and by using construct validity analyses. The item discrimination, subtest interrelationships, and item validity coefficients fall within moderate range. Strong validation data is not presented.
Ceiling or floor effects	No ceiling or floor effects noted at this time.

IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	No comments noted at this time.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	No comments noted at this time.
Limitations	The test fails to meet many of the generally used criteria for adequate reliability and validity in order to warrant a strong recommendation for its use. Much work remains to be done regarding norming, reliability, and validity of this battery before the RIPA-G can be solidly recommended as a measure of choice for assessing a variety of cognitive and linguistic abilities in older adults.
Other Comments	*TEP member Christine Baron (co-director of Stroke & Recovery Program, and manager in Speech & Pathology Service at Nat'l Rehab Hospital) indicated that this instrument is regularly used in the field. She also noted, however, that it is not well-liked because the norms are poor/don't hold true and it is not very functionally-based. She provided a copy of the instrument* ***[94]****.*

Executive Interview (EXIT) and CLOX—both instruments developed by Dr. Don Royall [67]	
Domain	Cognitive Functioning
Purpose (constructs measured, target population)	EXIT: Executive cognitive function (ECF) among the elderly [67]. CLOX: Executive impairment in the elderly [95]
Description	
Number and types of questions	EXIT: 25 items, scored 0-48, with higher scores indicating greater executive dyscontrol [67]. The cutoff point of 15/48 best discriminates between non-demented individuals and individuals with cortical or non-cortical dementing illness [95]. CLOX: The first part of the exam (CLOX1) involves drawing a clock that says 1:43. CLOX2 involves the examiner drawing the 1:43 clock and having the subject copy it. Scores range from 0-15 on both CLOX, with lower scores reflecting greater impairment. Scores of <10/15 on the CLOX1 and <12/15 on the CLOX2 are considered impaired [95].
Method of administration	EXIT: In-person, performance-based interview. Literacy level information was not provided in the article, but the subjects were required to speak/understand English [67]. CLOX: In-person, performance-based interview [95].
Time to administer	EXIT: 10 minutes [67]. CLOX: Information not provided.
Instrument has been tested with the following populations	EXIT: 39 randomly selected elderly subjects across four levels of care (10 per level) in an extended care community, ages 67-92 [67]. CLOX: 86 elderly subjects from independent living apartments in a retirement community and outpatients diagnosed with probable Alzheimer's disease—19 with and 26 without gross constructional impairment on the MMSE) and 59 young adult undergraduates as controls [95].
Sensitivity to measure changes in status	EXIT: Sensitivity to measure change in status was not assessed [67]. CLOX: Sensitivity to measure change in status was not assessed [95].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	EXIT: Assessment [67]. CLOX: Assessment [95].
Used for population estimates, facility-level estimates or individual level	EXIT: Individual level [67]. CLOX: Individual level [95].
Used for quality improvement, quality monitoring, quality measurement	EXIT: No [67]. CLOX: no [95].

Validity/Reliability	
Internal consistency, list relevant statistics	**(EXIT/CLOX)** EXIT: Of the final 25 items selected for the EXIT, the internal consistency was high (Cronbach's alpha = 0.83). Interrater reliability was also high (Pearson's r = 0.86) [67]. CLOX: In this sample, the internal consistency of the CLOX was high (Cronbach's alpha = 0.78). A range of r = 0.32 to 0.73, with a mean r = 0.40, was found for item total correlations. Interrater reliability in a subset of 27 elderly subjects. For the CLOX1, it was r = 0.90 and for CLOX2 it was r = 0.89, both $p < 0.001$ [95].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	EXIT: The concurrent validity of the EXIT is shown by its ability to distinguish between groups that it should theoretically be able to distinguish between. That is, there were statistically significant differences ($p< 0.001$) between non-institutionalized and institutionalized subjects on both the EXIT and the MMSE. Strong correlations were found between the EXIT and traditional measures of ECF CLOX: CLOX scores correlated strongly with cognitive impairment as measured by the EXIT and the MMSE. The EXIT accounted for 64% of the variance in CLOX1 scores (partial R-squared = 0.64) in a forward stepwise least squares regression model. The MMSE explained 68% of the variance in CLOX2 scores after adjusting for age and education. Combined, the EXIT and CLOX2 scores accounted for 70% of the variance in CLOX1 scores. As indicative of concurrent validity, the CLOX subscales were capable of discriminating subjects with Alzheimer's disease from the elderly control subjects after adjusting for age, education and MMSE test performance (MANCOVA: R(2,77)+3.6, $p < 0.03$). 79.1% of cases were correctly identified by the two CLOX subscales (Wilkes' lambda = 0.46; $F(2,82)=44.27$, $p < 0.0001$). The combination of the EXIT and the MMSE correctly identified 85.9% of the cases (Wilkes' lambda=0.29; $F(2,82)=103.8$, $p < 0.0001$). After adjusting for the EXIT and MMSE scores, the CLOX2 scores were capable of distinguishing between the Alzheimer's disease patients with grossly disorganized MMSE pentagons and those without such gross disorganization (ANCOVA: $F(1,33)=39.13$, $p < 0.0001$); the CLOX1 scores did not (ANCOVA: $F(1,33)=0.58$, NS). The pattern of performance on CLOX1 x CLOX2 subscales in a discriminant model correctly classified 87.9% of these Alzheimer's disease subgroups (Wilkes' lambda = 0.31; $F(2,34)=37.8$; $p < 0.001$); a less satisfactory performance was shown by the combination of the EXIT and MMSE scores (71.7% of correct identifications; Wilkes' lambda = 0.69; $F(2,34)=6.4$; $p < 0.005$). Furthermore, 14% (6) of the elderly control subjects failed the CLOX1 subscale and 27% (12) failed the EXIT at 10/48, while only 2.2% (1) failed the MMSE at 24/30 [95].
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	None of the EXIT items or the CLOX are currently in the IRF-PAI. IRF-PAI items #25, 26, and N, O, P, Q, and R of the FIM/IRF-PAI items, which also measure cognitive status/function, might be supplanted.

How well does this instrument distinguish people compared with existing questions on IRF-PAI	EXIT: Comparisons have not been made with the FIM, but both instruments have discriminative ability. In comparisons with the MMSE, the EXIT was found to have more subtle discriminative capability among residents at different levels of functional impairment [67]. That is, the EXIT is more sensitive than the MMSE to early cognitive impairment and non-cortical dementia in elderly patients [95]. CLOX: Again, comparisons with the FIM have not been made.
Limitations	No limitations noted at this time.
Other Comments	The EXIT is simple and has clinical face validity because many of the items are derived from routine clinical procedures. Sensitivity to a broad spectrum of cognitive function is provided by the wide range of scores elicited by the EXIT. As such, it might be useful for assessment of very early cognitive impairment that might not be detected by other instruments, and could also be used to make level of care determinations for patients being discharged from a hospital or admitted to a nursing home [67]. What's remarkable about this is that verbal, memory and ECF measures (e.g. WCST) DON'T Summary: 1. ECF is multifactorial. The EXIT and WCST (Wisconsin Card Sort Test) don't load on the same dimension of ECF (but neither loads on memory either). 2. The EXIT and not the WCST explains variance in disability. 3. The EXIT's correlation with disability is stronger than most other putative ECF measures (trails, verbal fluency, CLOX1, WCST etc.) 4. Nonetheless, the EXIT correlates only about .48 w/ IADL's (e.g. 25% of variance)." *A copy of the 15-page EXIT instrument and the 1-page CLOX instrument are included with a letter from Dr. Royall [96].*

Cognitive Impairment Diagnosing Instrument [10]	
Domain	Cognitive Functioning
Purpose (constructs measured, target population)	Cognitive function in the elderly [10].
Description	
Number and types of questions	73 items across 10 subscales (i.e., short-term memory, long-term memory, orientation in time, orientation in place, memory registration, concentration/calculation, judgment, object naming, abstract thinking, and higher cortical functions). 64 items are pass-fail (1=pass, 0=fail); 2 long-term memory and 7 abstract thinking items are scored as 0 (incorrect), 0.5 (partially correct), and 1.0 (completely correct), with the maximum possible score of 73 [10].
Method of administration	In-person, semi-structured interview. The mean years of education for the first sample was 3.4 and 4.1 for the second [10].
Time to administer	20 to 52 min with the nursing home subjects; 10 to 47 min with the hospital subjects [10].
Instrument has been tested with the following populations	63 non-educated elderly (aged 57 years or more) in a Korean nursing home and 246 in- patients and out-patients in the psychiatry department of a Korean hospital. Of the latter group, 71 were clinically diagnosed with dementia [10].
Sensitivity to measure changes in status	Sensitivity to measure change in status has not yet been assessed, but cross-sectional discriminative capability was demonstrated by the statistically significant differences in all subscale and total CIDI scores between the demented and non-demented groups [10].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Comprehensive assessment [10].
Used for population estimates, facility-level estimates or individual level	Individual level [10].
Used for quality improvement, quality monitoring, quality measurement	No [10].
Validity/Reliability	
Internal consistency, list relevant statistics	**(CIDI)** The range for interrater kappas was 0.2 to 1.0, with a mean of 0.658. 8 of the 67 (11.3%) items had kappas of less than 0.4. Kappas could not be calculated for six of the items and so concordance rates were used instead. The range was 48-96%, with a mean of 83.25%. Test-retest reliability calculated with Pearson's correlations ranged from 0.783 to 0.986 across the subscale scores and was 0.980 for the total CIDI score; all were statistically significant at $p < 0.001$. Pearson's correlation coefficients were calculated between each subscale score, and ranged from 0.479 to 0.780, with a mean of 0.644. The Cronbach's alpha for the subscale total was 0.890.

Internal consistency, list relevant statistics (continued)	Across individual subscale items, Cronbach's alphas were acceptably high, ranging from 0.875 for the concentration/calculation subscale to 0.662 for orientation to place [10].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Concurrent validity was demonstrated by statistically significant differences in all subscale and total CIDI scores between the demented and non-demented groups. Using a score of 54.0/54.5 as the optimal cut-off for presence of dementia, the sensitivity of the instrument was assessed at 89.3% and the specificity was 89.8%. That is, 5 of the 71 subjects classified as demented by the DSM obtained scores of more than 54.5 on the CIDI. 11 of the 172 non-demented patients received scores of less than the cut-off. Of those incorrectly assessed by the CIDI as non-demented, their total scores ranged from 55 to 56 (i.e., borderline) except for one with a score of 64 (that DSM diagnosis was based on memory impairment and personality change). Of those incorrectly assessed by the CIDI as demented, 6 of the scores were over 48 and 5 were over 28. Ten of the subjects had never been to school and 1 had only received 1 year of education. Furthermore, 5 of them had been diagnosed with major depression, 5 had limited premorbid intelligence, 2 had had cerebral stroke, one had a variety of mental disorders and some others had comorbid mental disorders. As evidence of convergent validity, the total CIDI scores of the 63 nursing home patients correlated with the Blessed Dementia Rating Scale (-0.567; $p < 0.001$), the raw MMSE-Korean score (0.878; $p < 0.001$), and the corrected (for the non-educated elderly) MMSE-K score (0.868; $p < 0.001$). The 10 CIDI subscale scores correlated with the Blessed with statistically significant correlations of between -0.312 and -0.602. Furthermore, subjects who scored a 0 on the MMSE-K also scored a 0 on the CIDI while those who scored 30 on the MMSE-K had CIDI scores of 70.5 to 73.0. Those with CIDI scores of between 48 and 64.5 were likewise found to have questionable cognitive impairment as assessed by the MMSE-K with scores of 21-24 [10].
Ceiling or floor effects	The ceiling effect is lessened because of the coverage of a broader sphere of cognitive function. However, due to the profound demented subjects, a floor effect was unavoidable in this study [10].
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	The CIDI instrument is not included with the article (but is available upon request from the authors) and so it is unknown if any of the questions might supplant any IRF-PAI items. If they could, IRF-PAI items #25, 26, and N, O, P, Q, and R of the FIM/IRF-PAI items, which also measure cognitive status/function, might be supplanted [10].
How well does this instrument distinguish people compared with existing questions on IRF-PAI	Both the FIM and the CIDI have broad discriminative capabilities and can assess global cognitive impairment, but the authors of this article claim that the CIDI can also detect circumscribed impairments and determine cognitive status on a specific sphere of cognitive function [10].

Limitations	Almost all psychotropic drugs and a variety of mental disorders could impede cognitive function, especially in the elderly. The instrument may only be applicable to a Korean-speaking population [10].
Other Comments	No comments noted at this time.

Cognistat (previously known as the Neuro)	
Domain	Cognitive Functioning
Purpose (constructs measured, target population)	Cognitive status among individuals with traumatic brain injury (TBI) to determine readiness for rehabilitation [50], cognitive functioning [51], and cognitive impairment among psychiatric patients [53].
Description	
Number and types of questions	11 subtests assessing consciousness, attention and orientation, language construction, memory, calculation, and reasoning. Results of performance are presented using a multidimensional cognitive profile. [50]. Subjects who pass the screening items are considered to be cognitively intact in that domain [51].
Method of administration	In-person interview [50, 51, 53].
Time to administer	10 to 20 minutes [54]. 10 to 30 minutes [55]. 5 minutes for cognitively intact persons and 30 minutes for cognitively impaired individuals [53]. 25 to 39 minutes for impaired individuals [51].
Instrument has been tested with the following populations	48 TBI patients with a mean age of 29.8 years who were admitted to rehabilitation within 6 months of their NCSE screening [50]. Subjects undergoing cognitive assessments at a VA Medical Center, with a mean age of 59.1 years [51]. 188 psychiatric patients from a county general hospital with a mean age of 33.5 years [53].
Sensitivity to measure changes in status	Sensitivity to measure change in status was not assessed [50, 51]. Upon assessing test-retest reliability, it was revealed that there was a minimal overall tendency for improvement in subjects' performance on each scale [53].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment [50, 51, 53].
Used for population estimates, facility-level estimates or individual level	Individual level [50, 51, 53].
Used for quality improvement, quality monitoring, quality measurement	No [50, 51, 53].
Validity/Reliability	
Internal consistency, list relevant statistics	**(Cognistat/NCSE)** The frequency of misidentification between the screens and the metric scores ranged from 4% (on the Comprehension scale) to 36% (on the Repetition scale). That is, these subjects failed the respective screens, but achieved equal or higher scores on the subsequent metric items than was required to pass the screening items. High rates of such misidentification (as with Repetition) indicate that the screening item is not a good predictor of impairment as it is assessed by the metric items. The screen items and the metric items reflecting the same construct/scale do not yield similar results. Analysis of the test-retest (conducted over 5-10 days

Internal consistency, list relevant statistics (continued)	intervals) results revealed that the Construction, Memory, and Calculation scales have the lowest reliability (with coefficients of 0.75, 0.49, and 0.77, respectively), but remarkable consistency was revealed by the performance ratings on the 7 other scales [53].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Upon examining the association between the NCSE and subsequent neuropsychological (NP) evaluation, the resulting Pearson product-moment correlation coefficient was $r = 0.53$ ($p < 0.001$). This association was not affected by age, education, Glasgow Coma Scale score, or design factors (e.g., time lapse between injury and NCSE administration or time lapse between injury and the NP evaluation). Evaluation of the extent of agreement in classification between the NCSE and the NP evaluation revealed the accuracy to be 0.75 (Kappa statistic = 0.43), which is indicative of fair agreement between the two testings. Sensitivity was calculated as 0.88 and specificity was 0.22. Examination of the association between the NCSE subtests and the NP tests paired with them revealed poor classification agreement. The Kappa statistics for the 11 test pairs ranged from 0.03 (Judgment) and 0.38 (Memory), accuracy ranged from 0.49 (Judgment) to 0.71 (Comprehension), sensitivity ranged from 0.08 (Comprehension) to 0.86 (Memory), and sensitivity ranged from 0.48 (Memory) to 1.00 (Comprehension). Additional analysis of all of the pairs revealed overall accuracy of 0.63, sensitivity of 0.49, specificity of 0.75, and the kappa statistic was 0.32 [50]. Again, significant correlations were found between the subtests and paired NP tests, with correlations ranging from 0.39 to 0.79. Poor agreement was found between the NCSE subtests and the NP tests, with a majority demonstrating no more than a 'fair' rate of classification, as revealed by Kappa analysis (among subjects classified as impaired, Kappas ranged from 0.31 for Naming to 0.65 for Construction/Visual Reproduction). A range of 0% (Naming) to 44% (Memory) was found for false positives [51].
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	The NCSE instrument is not included with any of the articles and so it is unknown if any of its questions might supplant any IRF-PAI items. If they could, IRF-PAI items #25, 26, and N, O, P, Q, and R of the FIM/IRF-PAI items, which also measure cognitive status/function, might be supplanted [50, 51, 53].
How well does this instrument distinguish people compared with existing questions on IRF-PAI	Comparisons have not been made with the FIM, which has demonstrated broad discriminative capability. In contrast, the ability of the NCSE to correctly distinguish between impaired and unimpaired subjects and to flag individuals most likely to be impaired in a given cognitive domain is questionable. In addition, the use of the NCSE to delineate functioning by domain may be inaccurate [51].
Limitations	Weak classification agreements between the NCSE subtests and the paired NP evaluations indicate that the NCSE cognitive profile would generally not be consistent with the NP evaluation profile. As such, while the NCSE has demonstrated its utility in identifying the presence of cognitive impairment, its identification of cognitive strengths and weaknesses is unlikely to be similar to those delineated by the NCSE [50].

Limitations (continued)	Small sample, use of published normative studies to establish impairment on the NP evaluations [51].
Other Comments	*A copy of the Cognistat/NCSE instrument is not included with any of the reviewed articles.* *Two TEP members (Dr. Eliot Roth of the Rehabilitation Institute of Chicago and Dr. Bruce Gans of the Kessler Rehabilitation Corporation) stated that Cognistat/NCSE is commonly used in the field to assess cognitive function.*

Depression

Depression

Several studies found depression to be a common functional psychiatric disorder of both institutionalized and non-institutionalized elderly. Post stoke depression is the focus of several major literature reviews. Although there is considerable disagreement in estimates of its prevalence in the population, ranging from 25% to 79%, there is general agreement that depression is associated with various negative outcomes (Gillen, Eberhardt, and Tennen, 1999). Similar conclusions have been reached from studies of institutionalized adults suffering from brain injury, dementia, and frailty.

Depression has been studied as an outcome measure, as well as a predictor of length of stay in a hospital or rehabilitation unit, rehabilitation efficiency, mortality, and as a case mix adjuster. Most of the existing depression rating scales have been developed and validated in younger, general population studies, and may not be appropriate for use with older adults. These instruments tend to be heavily loaded toward measuring somatic items such as loss of appetite and sleeplessness. While these symptoms can be indictors of depression in the elderly, particularly the frail elderly, they can also be side effects from medication, the result of physical deficits, or the aging process itself. Thus these instruments may produce false positive results.

Another problem with using a general population depression measure with the elderly is the difficulty of the instrument. For example, a self-rating 4-point scale is likely to be more confusing than an instrument with a yes/no format because it involves a greater number of choices and subtle discrimination that must be made by the respondent. This is particularly problematic for cognitively impaired respondents. Most researchers agree that a simple, easily understood instrument that is sensitive enough to distinguish between depression and other conditions with similar symptoms is essential for use with the geriatric population.

Yale Depression Screen	
Domain	Depression
Purpose (constructs measured, target population)	Geriatric Depression. This instrument was recommended by the Yale Task Force on Geriatric Assessment to be used as an initial screener for depression [12].
Description	
Number and types of questions	1 yes/no question. "Do you often feel sad or depressed?"
Method of administration	Measure was initially developed as a question that would be asked in person by a health care professional; specific information on literacy was not available [12, 35].
Time to administer	Less than 5 minutes.
Instrument has been tested with the following populations	Tested with elderly patients. No information on whether it has been tested with cognitively impaired patients or proxies [12].
Sensitivity to measure changes in status	No; measure was designed as a screener for depression.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Mainly used for assessment [12, 35].
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	No.
Validity/Reliability	
Internal consistency, list relevant statistics	N/A.
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	A study comparing the 1-item Yale Depression Screen to the 30-item GDS found that the Yale Depression Screen had a sensitivity of 69% and a specificity of 90% (compared to 54% and 93% for the GDS). The Yale Depression Screen correctly identified 9 depressed and 38 non-depressed subjects for a total of 47 of 55 patients correctly identified (85.4%). The 30-item GDS identified 44 of 55 patients correctly (80%). Although the Yale Depression Screen correctly identified 5.4% more patients than the GDS, the difference between the two was not significant (P = .50). With the 24% prevalence of depression in the study sample, the Yale Depression Screen had a positive predictive value of 85.4% and a negative predictive value of 90% [12].

Note = Reference information for depression domain begins on page 117.

Ceiling or floor effects	The Yale Depression Screen is able to identify patients with both major and minor depression as candidates for further evaluation, but it cannot distinguish between levels [12, 35].
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	No.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	Depression is not part of current instrument.
Limitations	Cannot distinguish between levels of depression or change in status [12, 35].
Other Comments	A study comparing the Yale Depression Screen to the 30-item GDS found that the Yale Depression Screen appeared to provide a quick reasonable alternative to more lengthy questionnaires such as the GDS. Results also suggest that the GDS provides no additional information beyond the 1 question [12].

4-Item Geriatric Depression Scale (GDS)	
Domain	Depression
Purpose (constructs measured, target population)	Geriatric depression.
Description	
Number and types of questions	4 yes/no questions.
Method of administration	Self-report paper-and-pencil instrument (PAPI); Specific information on literacy was not available.
Time to administer	10 minutes or less.
Instrument has been tested with the following populations	Tested with the elderly—both cognitively impaired and intact. Specific subpopulations studied include, stroke patients, the physically ill, arthritics, and elderly subjects undergoing cognitive treatment for senile dementia. Note: Brink (1984) has shown that in severe cases of dementia the subjects may fail to comprehend some questions. This suggests that the usefulness of the GDS in this population might be limited to subjects with mild to moderate degree of dementia [1, 22].
Sensitivity to measure changes in status	Most studies report fairly low sensitivity levels (compared to 30-item GDS). The 4-item GDS is a good measure to rule out nondepressed respondents and to identify individuals that should be evaluated by a clinician [1, 2, 9, 24, 26].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Mainly used for assessment at admission, but could also be used as a case-mix measure [1, 20, 21].
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	No.
Validity/Reliability	
Internal consistency, list relevant statistics	The 30-item GDS has been shown to be both a valid and reliable measure of depression. The 4-item was derived from the 30-item using statistical methods to determine consistency and correlation with 30-item measure [2, 16, 22, 26, 29].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Studies comparing the 4-item GDS to the 30-item GDS found that the 4-item measure to have acceptable levels of construct validity [2, 26].
Ceiling or floor effects	Studies indicate the 4-item GDS is able to identify patients with mild to severe depression as candidates for a clinical evaluation (and rule out the nondepressed), but the measure is not able to distinguish between levels of depression [2, 26].

IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	No.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	Depression is not part of current instrument.
Limitations	Cannot distinguish between levels of depression or change in status.
Other Comments	The GDS stands out as one of the preferred measures of depression in the elderly. The GDS was designed specifically for the elderly and does not include somatic items that have been shown to create false positives in some of the other measures designed for the general population. The 4-item version has been shown to be reliable and valid, and may be favorable to a longer version or the GDS for our purpose. [Note: There's also a 15-item GDS which has been extremely well tested with good results. However, we thought 15 items would be too long for our purpose. Additional research on the 4- and 5-item GDS is expected] [1, 2, 9, 20, 21, 22, 26, 33].

1-Item GDS	
Domain	Depression
Purpose (constructs measured, target population)	Geriatric depression.
Description	
Number and types of questions	1 yes/no question (Do you feel that your life is empty?) [2]
Method of administration	The GDS was designed to be self-administered, but has been interviewer-administered in at least one study. Specific information on literacy was not available.
Time to administer	Less than 5 minutes.
Instrument has been tested with the following populations	The GDS has been well tested both with cognitively impaired and intact patients. However, the 1-item version has not been well tested on its own.
Sensitivity to measure changes in status	Data suggests the 1-item GDS is useful to help identify elderly patients with depressive symptoms. A study of elderly primary care patients found that the 1-item GDS identified 84% of depressive cases and the agreement with a 15-item version was 79% [2].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Mainly used to assess depressive symptoms.
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	No.
Validity/Reliability	
Internal consistency, list relevant statistics	Not a lot of data on the reliability of the 1-item measure. There have been numerous studies that support the reliability of longer versions of the GDS (30-item, 15-item) [1, 2, 9, 20, 21, 22, 24, 26, 27, 28, 31, 33].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Sensitivity and specificity for the 1-item GDS were 59 and 75% in a study of elderly primary care patients. More data on elderly patients in a hospital setting is needed to establish validity for our purpose [2].
Ceiling or floor effects	The 1-item GDS is thought to be able to identify patients with mild to severe depressive symptoms, though more inclined to pick up moderate to severe cases. The item is not able to distinguish between levels of depression [2].

IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	No.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	Depression is not part of current instrument.
Limitations	Cannot distinguish between levels of depression or change in status [2].
Other Comments	The 1-item version of the GDS has not been well tested with our target population but has shown promising signs as a quick and simple method of identifying elderly patients with depressive symptoms that may require a follow-up evaluation.

Center for Epidemiological Study of Depression Scale (CES-D)	
Domain	Depression
Purpose (constructs measured, target population)	The CES-D was originally developed for use in the general population but has since been studied extensively for use with the elderly [8].
Description	
Number and types of questions	20-item questionnaire that incorporates items from previously developed surveys. Respondents are asked to rate depressive symptoms during the past week on a 4-point scale (0 = rarely—less than 1 day per week to 3 = most or all of the time—5-7 days per week) [11].
Method of administration	Used as both a self-report and interviewer-administered instrument [8].
Time to administer	The CES-D has been described as "brief and quick to administer," but no time estimates were given. However, one study conducted with nursing home patients found that the CES-D took longer than other instruments to administer (GDS), and required the interviewer to frequently repeat several items [13].
Instrument has been tested with the following populations	Although originally designed for use in the general population, the CES-D has been validated with a number of elderly populations including stroke patients, the physically disabled, and the frail elderly. Although most of the studies found the CES-D to be appropriate and valid, one study of nursing home patients found that the CES-D fared poorly on all methods of diagnostic test evaluation, particularly in comparison to other instruments, including the GDS [8, 10, 11, 13].
Sensitivity to measure changes in status	The CES-D was designed to measure the current level of depressive symptomatology. It is primarily used as a screening instrument.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Designed to be a screening instrument, but can also be used for other purposes that require the identification of depressive symptoms.
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	No.
Validity/Reliability	
Internal consistency, list relevant statistics	Good inter-rater reliability has been shown when the instrument is interviewer-administered. Inter-rater reliability was assessed by Kappa statistics, which measure the level of agreement, adjusted for agreement due to chance for items measured on a nominal scale [8]. To measure Internal consistency, the Spearman rho correlations between each item and the total test score were calculated on the CES-D data obtained from the nurse examiner. All but one of the correlations were significant ($p < .05$), ranging from .39 to .75 [11].

Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Many studies have been conducted to assess the validity of the CES-D with the elderly population. The literature indicates that the CES-D correlates highly with other well-tested depression instruments, including Psychiatric Diagnosis based on the DSM-III manual. This provides strong support for the construct validity of the CES-D. A study of elderly stroke patients established the validity of the CES-D by evaluating how well it could distinguish between no depression and minor or major depression, (i.e., sensitivity to and specificity for depression). The CES-D appeared to yield no false positives. The estimated sensitivity at the 16 cutpoint was 73%. When the CES-D was less than 16, a very high proportion of the patients were not depressed (84%), but a number of depressed patients were missed (27% false negative rate) [11]. Other studies found sensitivity ranges of 86-92% and specificity ranges of 84-90% [8, 10, 15].
Ceiling or floor effects	The CES-D has been shown to have excellent properties for use as a screening instrument for major depression in the elderly, but some researchers think that the GDS is a better instrument for identifying patients with mild depressive symptoms.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	No.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	Depression is not part of current instrument.
Limitations	Cannot distinguish between levels of depression. One study of nursing home patients found the CES-D to be difficult and time consuming to administer (compared to other instruments) and that it fared poorly on methods of diagnostic evaluation.
Other Comments	There have been conflicting reports concerning the issue of whether high scores on the somatic subscale, influenced by physical illness, bias findings of depressive symptoms among the elderly. Some studies have concluded that such items do distort reports of depressive symptoms in this population, whereas others found they do not. In a qualitative review, Radloff and Teri (1986), found that somatic scores were not overrepresented to the detriment of the total score. A follow-up study by Davidson, Feldman, and Crawford (1994), found no evidence that the somatic subscale score contributes disproportionately to the total symptom score in the elderly population.

Hamilton Depression Rating Scale (HAM-D)	
Domain	Depression
Purpose (constructs measured, target population)	Designed in the late 1950s as a standardized scale for the measurement of the severity of depressive symptoms [5].
Description	
Number and types of questions	A 21-item scale administered by a trained clinician. The clinician considers both the intensity and frequency of a symptom and assigns it a rating value [13].
Method of administration	Designed to be administered by a trained examiner (after a 30-minute clinical interview to rule out "endogenous" depression). It has since been adapted for use as a self-rating scale but is most commonly used as an examiner-administered interview [5, 13].
Time to administer	Estimates range from 30 minutes to an hour or more (assuming clinical interview to rule out endogenous depression an adequate time for clinician to observe patient). The instrument was designed for use if a hospital setting [D5, D7, D13, D18].
Instrument has been tested with the following populations	The HAM-D has been widely tested and validated with both younger and geriatric populations, including various subpopulations such as stroke patients and the disabled [5, 9].
Sensitivity to measure changes in status	Yes, intended to be a measure of treatment outcomes. Is widely used to study the differential effect of various treatments on specific depressive symptoms [5, 9, 13].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Used mainly as a measure of treatment outcomes. It is not widely used as a screening device because it requires adequate time to observe patients [7, 13].
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	Can be used to measure quality and outcomes.
Validity/Reliability	
Internal consistency, list relevant statistics	The HAM-D has been proven reliable in a number of studies dating back to the 1950s. One often-cited study by Yesavage and Brink (1983) found the HAM-D had a high level of internal consistency (correlation with total score was .56; inter-item correlation was .34). An Alpha coefficient was utilized in order to provide an overall measure of internal consistency (.90) [16]. Also, a study by Williams (1988) found that the use of a structured interview guide increased the test-retest reliability of the HAM-D [5].

Construct validity & description of relationships with existing IRF-PAI (FIM) questions	The HAM-D has been shown to have high concurrent and differential validity in numerous studies. It is regarded in the literature as the "foremost' examiner-rating scale [5]. A study of convergent validity found a correlation of .83 between the HAM-D and the GDS and .80 between the HAM-D and the SDS [16].
Ceiling or floor effects	Capable of distinguishing between degrees of depression as well as changes over time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	No.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	Depression is not part of current instrument.
Limitations	Requires an experienced examiner to administer. Takes a longer amount of time than other depression instruments [5].
Other Comments	While the HAM-D is a well-tested, reliable, and valid instrument, it is probably not appropriate for our purposes.

Zung Self-Rating Depression Scale (SDS)	
Domain	Depression
Purpose (constructs measured, target population)	Designed as a measure of depression in the general population [1].
Description	
Number and types of questions	A 20-item self-rating symptom list. Answer categories = never, sometimes, usually, always [1].
Method of administration	Self-administered.
Time to administer	Time estimates not available.
Instrument has been tested with the following populations	Tested with elderly including stroke patients and other physically ill/disabled populations [1, 14, 18].
Sensitivity to measure changes in status	The SDS is a screening instrument that provides information on current depressive symptoms.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Mainly used for assessment.
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	No.
Validity/Reliability	
Internal consistency, list relevant statistics	Studies show internal consistency with split half reliability coefficients in the range of .73-.79 [16].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Validity coefficients have shown great variability across studies. Correlations with the HAM-D have ranged from .22–.95. Studies have found that the mean score for elders is significantly higher than that for younger subjects, and many normal aged wind up as false positives, while the scale misses those depressives whose disorder is in the guise of somatic illness. As a result, some researchers have questioned the appropriateness of the SDS for research or clinical assessment of geriatric depression [1, 4, 16, 18].
Ceiling or floor effects	Measures depressive frequency from never to always, but does not distinguish between major and minor depression.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	No.

How well does this instrument distinguish people compared with existing questions on IRF-PAI	Depression is not part of current instrument.
Limitations	May not be appropriate for use with the elderly due to false positives and varying validity coefficients.
Other Comments	No comments noted at this time.

Pain

Pain

The personal nature of pain makes it difficult to assess in a standardized fashion. There are no biological markers of pain and so it cannot be directly observed by clinicians. As such, assessment is primarily dependent upon patient self-report. This presents difficulties in measurement because individuals may perceive pain differently (Jensen and Karoly, 1992). In addition, assessment of this complex perceptual experience may be more challenging among certain populations (e.g., children, the elderly, and cognitively impaired persons). To complicate matters further, it has been demonstrated that pain is associated with psychosocial factors and depression in the elderly (AGS Panel on Chronic Pain in Older Persons, 1998).

Nonetheless, effective pain management is not possible without a valid, reliable, and clinically useful way to measure the experience. However, the literature lacks an integrated overview of pain assessment techniques and critical evaluation of the methods commonly used, and there is an increasing need for broader, more operational definitions of pain. The continued use of scales such as the numerical rating scale (NRS), the visual analogue scale (VAS), and others is accompanied by the built-in assumption that pain is a unidimensional experience which varies only in intensity, and other dimensions such as pain location and pain affect are not measured (Chapman, Casey, Dubner et al., 1984). While the use of such scales with relatively few questions in the same format maximizes reliability coefficients, more revealing and broad content is sacrificed (Caraceni, Cherny, Fainsinger et al., 2002).

The pain measure that is currently part of the Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) is classified as a NRS, a single line instrument that ranges from 0 on one end to 10 on the other end, with a "0" indicating "no pain," a "5" indicating "moderate pain," and a "10" indicating "worst possible pain." Some may describe the IRF-PAI pain measure as a VAS (a line instrument similarly anchored by "no pain" and "pain as bad as it could be"), but the original conception of the VAS included no corresponding numerical values on the scale, it was personally marked by the patient (or the patient indicated where s/he wanted it marked by pointing), and the 10-cm line was measured by the interviewer to determine the patient's pain rating (Jensen, Karoly, and Braver, 1986). The NRS does not require this extra step for scoring, which may add additional time to the administration of the instrument and an additional source of error, and allows the patient to make a verbal rating. In addition, older people tend to have conceptual difficulty with the VAS (Jensen and Karoly, 1992).

Of the 137 reviewed articles, which covered approximately 50 distinct instruments, 17 were examined more closely. Studies involving palliative pain measures were excluded in favor of instruments that assess chronic, neuropathic, musculoskeletal, or arthritic pain (i.e., types of pain that are likely to be encountered in a rehabilitation setting). Measures of pain specific to certain body parts (e.g., back, head, neck, knee, etc.) were excluded because their focus was too narrow and not necessarily generalizable to other types of pain. The instruments then selected for further examination were chosen by the amount of literature available (i.e., those that have been reviewed extensively were generally felt to have face and content validity) and/or the persons with whom the instruments were tested (i.e., elderly, cognitively impaired, illiterate, or rehabilitation-specific populations). Articles for inclusion were not limited solely by population-type because this would have resulted in a paucity of information. The articles retained are

seminal studies which included all of the information needed to populate the accompanying matrix and overview manuscripts that provide a context for the examination of the selected instruments. The instruments ultimately reviewed include:

- Numerical Rating Scale (NRS)
- Geriatric Pain Measure (GPM)
- Minimum Data Set (MDS)
- Faces Pain Scale (FPS)
- Musculoskeletal Form of the Medical Rehabilitation Follow-Along (MRFA)
- McGill Pain Questionnaire (MPQ)
- Chronic Pain Experience Instrument (CPEI)

Numerical Rating Scale (NRS) [7, 9, 14, 16, 100, 115] (C)	
Domain	Pain
Purpose (constructs measured, target population)	Chronic non-malignant pain [9]. How the number of pain assessments affects reliability and validity of average pain intensity measures in 200 chronic pain patients [7]. Number of levels needed in pain intensity measurement as determined in a study examining pre- to post-txt changes in pain intensity in 122 chronic pain patients [14]. The validity and practicality of 6 pain intensity measures (a 100-point NRS, 11-point Box Scale rating, 6-point Behavioral Rating Scale (BRS), a Visual Analogue Scale (VAS), a 4-point Verbal Rating Scale (VRS), and a 5-point VRS completed by 75 chronic pain patients [16]. Pain evaluation methods for illiterate rheumatic disease patients [100]. Pain experience in geriatric patients [115].
Description	
Number and types of questions	11-point, 21-point, and 100-point scales seem to be the ones used most often [General]. 11-point (0-10) numeric scale, with a “0” indicating ‘”no pain” and a “10” indicating pain ”as intense as you could imagine”; patients asked to rate worst, least, average, and current pain (4 items) over the previous 2 weeks using the scales and then 3 different composites were created from various combinations of the means of the 4 individual measures (it was hypothesized that the composites would be more sensitive to changes in pain than the individual measures) [9]. Patients were required to complete 2 weeks of pain diaries before their evaluation and were then asked to provide multiple (hourly) pain ratings on the diaries using a 0-10 NRS [7]. Patients assessed their pre- and post-txt current pain, as well as their worst, least, and average pain over the previous week, on a 100-point scale (8 items total) [14]. 4 kinds of pain (present, least, most, and average) were rated, using the 6 scales (24-items totals) [16]. Three different items (VAS, NRS, and VRS) to be scaled on a continuum of either a 10cm line, 11 numbers from 0-10, or a list of descriptive words (no pain, mild, moderate, severe, or unbearable pain), administered before and immediately after regular medical consultation [100]. Three 1-item scales (NRS, VAS, and graphic rating scale [GRS]) administered twice (with a 5-minute gap in between) and then a question about whether they experienced pain, ache, hurt (PAH) or other symptoms. All 3 scales consisted of a 10 cm line with extreme endpoints marked "no pain" and "worst pain imaginable"; the NRS was also graded 1-10, and the GRS has the words "mild," "moderate," "moderately severe," "severe," and "unbearable" between the endpoints [115].
Method of administration	Could be administered through in-person interview or using a paper-and-pencil interview (PAPI) scale, or over the telephone [General]. Telephone interviews conducted just before or w/in 2 days after beginning the program, and then 2 weeks, 1 month, and 2 months after completion of the program; subjects required to read, write, or speak English [9]. 2 weeks of hourly, self-reported ratings (method of administration not detailed) [7] (presumably in-person although method of administration not detailed: "subjects were asked to complete four 100-point NRSs before and immediately after txt") [14].

Note = Reference information for pain domain begins on page 121.

Method of administration (continued)	In-person interview (NRS, BRS, VRS) and PAPI (the Box and the VAS must be given in written form). No information is provided about the literacy level, but all 6 instruments are provided in the article [16]. In-person, PAPI administration (patients were asked to score their pain severity for the previous week, marking each scale with a pen); each scale was presented twice (before and after medical consultation) [100]. PAPI for the 3 scales (many marked by pointing b/c had difficulty marking with a pen) and short, structured in-person interview for the PAH question. Subjects were required to have the ability to read the written information [115].
Time to administer	< 5 minutes for VAS (although measuring/ scoring may take less than an add'l 2 minutes), NRS, or VDS [47]. Information not provided, but it would seem to be minimal [9, 7, 14, 16, 100]. 20-40 minutes (for all 4 measures, 3 of which were completed twice) [115].
Instrument has been tested with the following populations	General population of chronic pain sufferers, with an age range or 21-78 years and a variety of primary pain sites [9]. Chronic pain patients screened for possible txt at the UW Multidisciplinary Pain Center (but who did not undergo txt), with an average age of 43.83 years and a variety of primary pain sites [7]. Chronic pain patients who participated in a process analysis of the UW inpatient, multidisciplinary pain program [14]. 75 chronic pain patients consecutively admitted to St. Joseph's pain unit, a 10-bed inpatient program designed to teach patients to control and cope with their pain [16]. 66 literate and 25 illiterate (could not read/write Portuguese and were also innumerate) Brazilians with rheumatoid arthritis [100]. 167 Swedish patients who had acceptable visual function and sufficiently good hearing and no terminal illness or diagnosed dementia. Age range = 60-96 years (mean = 80.5 years) [115].
Sensitivity to measure changes in status	Each individual and composite measure changed significantly from pre-treatment to the 2-week, 1-month, and 2-month assessments (as predicted, the 3- and 4-item composite measures tended to reflect greater change, based on the size of the F-value, than did the 2-item composite or individual measure, but no F-value differed significantly from any other), thus supporting the instrument's ability to detect expected changes. The relatively low test-retest stability for some individual ratings indicates that ratings of individual pain may be rather variable w/in each individual even when systematic group changes are not seen. In regard to the effects of variables such as activity and medication use, individual pain ratings may be more sensitive and so individual pain ratings may not be as useful as composite measures for assessing average pain intensity experienced over a period of time [9]. Averages of daily pain reports were calculated and regression equations were performed for each subject to examine the reports' consistency. To test the accuracy of the equations, standard deviations were compared using a t-test. The results indicated that the regression equations accurately distinguished patients who report consistent pain levels from those whose daily pain fluctuates (the majority of the study patients had average daily ratings that were not similar from one day to another) [7]. After recoding the 100-point NRS to 2-, 3-, 4-, 6-, 11-, and 21-point scales, the pre- to post-txt changes in terms of average, current, worst, and least pain intensity were examined through paired t-tests for the original 100-point NRS and the 6 recoded scales. The 11-, 21-, and 100-point measures had essentially the same means, standard deviations , t-values, and P-values, indicating that these are

Sensitivity to measure changes in status (continued)	generally equally sensitive measures for detecting change. The scales with less than 11 points had the greatest changes in sensitivity/ability to detect txt effects [14]. All 6 scales are relatively sensitive, but VRS are usually assumed to be less sensitive (because they provide fewer response categories) than other scales, such as the VAS (with potentially infinite responses). In this study, incorrect responses (i.e., left blank, included 2 answers, etc.) to the pain intensity measures were examined. All scales had some degree of incorrect responses (the rate ranged from 2.7% for the VRS to 8% for the BRS). The difference in the rates was not significant as revealed by a chi-square test for correlated dichotomous data. It was found from the correlation between incorrect responding and subject age that it was only significant for the VAS ($p < 0.01$) [16]. "The evaluation of the validity and sensitivity of these pain scales in each pain group is now in progress" [100]. On the NRS, there was no significant difference ($z = -1.251$, $p < 0.211$) in the level of pain between the first test (mean = 3.15) and the second one (mean = 3.03) administered 5 minutes later. This may indicate that the NRS is more reliable and valid than the VAS or GRS, or it might mean that the 10-point NRS is less sensitive to changes in pain experience. Ratings of pain experience were significantly lower at the second test than the first for the VAS and the GRS [115].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Used for assessment before and after pain txt program [9]. Used for assessment of average pain intensity over a certain period of time [7]. Assessment pre- and post-txt [14]. Correct pain assessment [16]. To evaluate the reliability of 3 pain instruments with an illiterate population; the scales themselves are used for assessment [100]. Assessment of pain experience [115].
Used for population estimates, facility-level estimates or individual level	Individual level [9, 7, 14, 16, 100, 115].
Used for quality improvement, quality monitoring, quality measurement	No [9, 7, 14, 16, 100, 115].
Validity/Reliability	
Internal consistency, list relevant statistics	The hypothesis that little average change occurs in pain intensity between the 2-week through the 2-month assessment was tested in order to examine the reliability of the measures after txt. No significant change in pain during that period, as determined by the individual and composite measures, were shown by the repeated measures ANOVAs (F-values ranged from 0.01 to 1.01, all $P > 0.05$). Test-retest stability correlation coefficients were calculated for each of the pain intensity measures between adjacent assessments (2-week and 1-month follow-up scores and 1- and 2-month follow-up scores); as expected, the stability coefficients were generally higher for the composite measures (range = 0.7-0.88, median = 0.81) than for the individual ratings (range = 0.55-0.84, median = 0.72) although the rating of least pain was comparable in stability to the composite measures [9]. 24 composites/scales were created based on the number of pain ratings per day and the number of days of rating,

Internal consistency, list relevant statistics (continued)	with a range of 2-28 pain intensity ratings per composite/scale, and internal consistency coefficients were then calculated and compared across them. These coefficients increased as the number of assessments increased. Internal consistency coefficients of 0.94 or greater were reached in 3 days w/at least 3 assessments per day, but great internal consistency was also found for a single daily rating composite across 7 days. In a comparison between scales computed by averaging across multiple days versus those calculated by averaging the ratings from a single day, little difference was found in the internal consistency coefficients (median of 0.93, as opposed to 0.95). The test-retest stability for single measures of pain intensity were estimated by correlating the pain rating reported at one point in time during week 1 to a similar point in time for week 2; for 2 ratings, it was calculated by correlating the average of 2 pain intensity ratings of week 1 with an average of 2 ratings of week 2; and for aggregate pain intensity measures, it was computed from the average of multiple ratings during week 1 correlated with similar ratings from week 2. Increases in the number of assessments (i.e., the aggregate pain intensity measures) likewise increase the stability of the measurement. The stability coefficient of a single measure was 0.63 while that for the average of 28 measures (i.e., 4 assessments per day across all 7 days) was 0.94. In a comparison between scales computed by averaging across multiple days versus those calculated by averaging the ratings from a single day, the stability coefficients of the 7 scales estimated from a single day's ratings were inadequate (median = 0.79) while those calculated from ratings over multiple days were adequate (median = 0.91) [7]. After recoding the 100-point NRS to 2-, 3-, 4-, 6, 11-, and 21-point scales, correlation coefficients were estimated between each recoded score and the original 100-point scale. The correlation coefficients for the 11-, 21-, and 100-point measures were all greater than 0.98, and the lowest/weakest coefficients were found for the 2-point measure (range = 0.59-0.80), suggesting that pain intensity measures with 11 or more levels are probably adequate for detecting pain intensity changes among chronic pain patients [14]. Again, incorrect responses were only significantly correlated with age for the VAS. By intercorrelating responses to the 6 scales (and separately for each of the 4 descriptions of pain) and conducting a series of principal axis factor analyses on the correlations, the strength of the relationship between the shared variance of all the measures and each individual scale was determined (subjects who responded incorrectly to any of the measures were excluded from these analyses). All coefficients were significant ($p < 0.001$, two-tailed tests), indicating a large amount of shared variance/internal consistency among the pain intensity measures [16]. The Pearson product moment correlations between assessment in Times 1 and 2 for each pain scale show higher correlation coefficients for the literate patient group although it was only statistically significant in the VAS ($p < 0.001$). All 3 scales showed high correlation coefficients between Times 1 and 2 in the literate patient group. Conversely, only the NRS and VRS showed high correlation coefficients in the illiterate patient group. The pain scores were systematically higher in the illiterate patient group ($p < 0.05$), regardless of the type of scale used [100]. Between the first and second tests in the total sample, there was a significant ($p < 0.001$) and consistently high ($r = 0.75–0.83$) correlation for all three rating scales. Furthermore, when alternative-forms reliability was calculated, the correlation between the 3 rating scales at both the first and second test was significant ($p < 0.001$) [115].

Construct validity & description of relationships with existing IRF-PAI (FIM) questions	(NOTE: the IRF-PAI also includes a 0-10 NRS to assess pain at admission and discharge). Previous studies showed that a 2-item composite comprised of an average of least and average pain ratings has greater validity than other composites when measuring actual average pain, but the same composite in this study appeared no more valid than composites that included worst and current pain ratings [9]. As the number of ratings reported increased, so did the validity of estimates as measures of average pain. A validity coefficient of 0.74 was found for a single rating, but excellent estimates (i.e., a coefficient of 0.94 or greater) required a minimum of 2 measures each day over 4 days of assessment. In a comparison between scales computed by averaging across multiple days versus those calculated by averaging the ratings from a single day, little difference was found in the validity coefficients (median of 0.94 versus 0.89) [7]. Almost all of the patients treated the 100-point scale as a 21-point scale for each of the pain ratings, at both pre- and post-txt (as many as 97% and consistently more than 90% rated their pain in multiples of 5 or 10 points); a substantial number (more than one-half) provided responses in multiples of 10 only, treating the 100-point scale as an 11-point scale [14]. Although there was some range between the 6 scales for incorrect response rates, none of the differences were significant, suggesting that the sample chronic pain subjects generally did not give more incorrect responses to one scale rather than another. As such, each scale may be considered a useful measure of subjective chronic pain intensity [16]. The evaluation of the validity and sensitivity of these pain scales in each patient group is now in progress" [100]. Patients rated significantly ($p < 0.001$) lower on all 3 rating scales if they verbally denied experiencing PAH. However, it was found that the probability of consensus between the patients' pain ratings and the verbal PAH report decreased with greater age, particularly for the VAS. In addition, despite the high test-retest reliability found, patients rated their pain experience significantly lower on the second test as compared to the first for the VAS and GRS scales [115].
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	Again, the IRF-PAI already includes a 0-10 NRS, but does not use composite measures or assess pain at more than two points in time. However, the authors note that although some small increase in sensitivity can be obtained by using composites, the improvement may not be large enough to make multiple measures of pain intensity (e.g., rating of worst, current, and least pain) necessary [9]. The IRF-PAI does not measure average pain over time as was studied here [7]. The IRF-PAI already has a NRS and this study supports the use of an NRS-100 over the other 5 scales examined [16]. Results indicate that the NRS (such as the scale currently used on the IRF-PAI) has the higher reliability among both literate and illiterate patients [100]. The IRF-PAI presently includes a 0-10 NRS, but it resembles a VAS in that it is 10 cm long with the same extreme endpoints (in fact, some studies refer to such numerical scales as VASs or VRSs) as well as the GRS because it includes a middle point of "moderate" [115].

How well does this instrument distinguish people compared with existing questions on IRF-PAI	The composite measures used here, as opposed to the individual measures also used in the study and on the IRF-PAI presently, did not show a statistically significant superiority to the individual ratings in terms of their ability to detect changes in pain intensity from pre-txt to various points after txt. But the composite scores did demonstrate greater stability than did the individual ratings after txt. The authors concluded that 0-10 individual pain ratings have adequate psychometric strengths for use in chronic pain research (especially in studies that involve comparing groups with relatively large samples or when assessing txt effects on a specific pain dimension), but that 0-10 composite ratings might be used when maximal reliability is required (e.g., studies with small sample sizes or clinical settings where monitoring of pain intensity changes is needed) [9]. The conclusion that measures made up of multiple pain ratings are more valid indicators of average pain (which is what most clinicians attempt to control) than a single measure is supported by this study [7]. It doesn't. Rather this study indicates that the 11-point scale used on the IRF-PAI is as equally valid as a 100-point scale [14].
Limitations	Reliability was assessed only after multidisciplinary txt which may have changed how attentive patients are to their pain and how they rate it, which, in turn, might affect the stability of the ratings. In addition, only 0-10 scales were compared, and other types of scales may have better psychometric properties for individual pain intensity ratings. The generalizability of these findings to other populations (i.e., other samples of pain center patients, chronic pain sufferers not seeking txt in pain centers, and cancer-related pain) is not known [9]. Examination of pain report patterns over time is not possible when many measures of pain are averaged, and the exact number of assessment required to maximize the psychometric strengths of average pain intensity measurement may differ among different patient populations [7].
Other Comments	Problems in pain measurement: (1) the literature lacks an integrated overview of pain assessment techniques and critical evaluation of methods commonly employed, (2) there is a growing need for broader, more operational definitions of pain, (3) there is a built-in assumption with the use of scales such as the NRS, VAS, and others, that pain is a unidimensional experience that varies only in intensity [1], and while the use of such scales/relatively few questions in the same format maximizes reliability coefficients, more telling and broad content is sacrificed [3]. The NRS-100 is extremely simple to administer and score, and can be administered either verbally or in written form. *Copies of all 6 scales (i.e., 100-point NRS, 11-point Box Scale, 6-point BRS, VAS, 4-point VRS, 5-point VRS) are included in this article* [16]. **VRS** (*Strengths* = easy to administer and score, good evidence for construct validity, high compliance with measurement task; *Weaknesses* = may be difficult for persons with limited vocab, relatively few response categories compared with other scales, ranked scores represent ordinal data and should be statistically treated as such, subjects forced to choose 1 word even if none of the words adequately express their pain); **VAS** (*S* = easy to administer, many/infinite response categories, scores can be treated as ratio data, good evidence for construct validity; *W* = scoring the measures requires an extra step which may add time and add'l source of error, older people tend to have difficulty with the VAS); **NRS** (*S* = Easy to administer, many response categories if use NRS-100, easy to score, high compliance with measurement task, good evidence for construct validity, sources can be

Other Comments (continued)	treated as ratio data; *W* = no NRS-100 txt sensitivity evidence as compared with other measures, limited response categories if NRS-11 used, NRS-11 may be less sensitive to txt effects than VASs); **BRS** (*S* = easy to administer and score, limited evidence for construct validity; *W* = no evidence about relative compliance rates, limited number of response categories, no evidence regarding relative txt sensitivity, scores need to be statistically treated as ordinal data, pain interference may be confounded with pain intensity); **Box** (*S* = easy to administer and score, good evidence for construct validity, measurement task compliance rate is high, scores may be treated as ratio data; *W* = no relative txt sensitivity evidence as compared with other measures, limited response categories) [135; this article also contains many instruments]. Among literate and illiterate patients with rheumatic arthritis, significant differences across all pain scales used (i.e., NRS, VAS, and VAS) was systematically observed although results do indicate that the NRS has greater reliability with this population. The 3 different scales are included in this article [100]. A study by Williams et al. (1978) found that elderly patients with hip fractures tended to rate randomly on the VAS [115].

Geriatric Pain Measure (GPM) [89] (Q)	
Domain	Pain
Purpose (constructs measured, target population)	**Multidimensional pain** among 186 subjects in ambulatory geriatric clinics [89].
Description	
Number and types of questions	24-items (22 dichotomous, yes/no questions and 2 items scored categorically on a 0 to 10 scale) [89].
Method of administration	Can be interviewer-administered or self-administered (in this study, there was a standardized, assisted interview conducted by research assistants). No information provided about the literacy level [89].
Time to administer	Less than 5 minutes [89].
Instrument has been tested with the following populations	176 geriatric subjects (with a mean age of 84.7 years, multiple medical problems and took multiple medications) in ambulatory geriatric clinics, 73% of whom had a history of pain complaints in their medical records [89].
Sensitivity to measure changes in status	"The instrument has not been analyzed for sensitivity to change or to detect change over time" [89].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment of pain intensity, pain-related functional status, mood, and quality of life among cognitively intact ambulatory older adults [89].
Used for population estimates, facility-level estimates or individual level	Individual estimates [89].
Used for quality improvement, quality monitoring, quality measurement	No [89].
Validity/Reliability	
Internal consistency, list relevant statistics	Standardized Cronbach's alpha was 0.9345; homogeneity ratio was 0.457, and average inter-item correlation was 0.415 (range 0.166-0.830; $p < 0.05$). Based on factor analysis, 5 unique subscales were identified (related to disengagement because of pain; pain intensity; pain with ambulation; pain with strenuous activities; pain with other activities); none of the items failed to load onto one of the subscales and only 2 loaded on 2. The instrument was also administered to subsample of 50 patients on 2 separate occasions, 48 to 72 hours apart. Using Pearsons r correlations for continuous data and the Kappa statistic for categorical data, test-retest reliability/stability was analyzed. A Pearsons r of 0.9018 ($p < 0.0000$) was revealed for the total GPM score with an average item agreement of 78% and an average Kappa of 0.596 [89].
Construct validity & description of relationships with existing	In regard to face validity, an expert panel of 3 geriatricians and 2 pain specialists reviewed the initial 33-item draft of the GPM and were asked 3 questions for each item (is it valid/could it be used to measure pain in ambulatory older people?; assignment to 1 of 4 content areas: functional

IRF-PAI (FIM) questions	status, pain intensity, mood, and utilization; how should the item be changed to enhance validity?). After minor changes, there was 99% agreement on the face validity and 86% agreement among the experts on the content area of each item. Upon pilot testing with 25 older volunteers,
Construct validity & description of relationships with existing IRF-PAI (FIM) questions (continued)	nine items were eliminated because they were highly correlated with other items and thus considered redundant, and also because all 33 items required substantial time and interviewer assistance. To evaluate concurrent validity, another subsample of 50 subjects also completed the MPQ. Pearsons r correlations were used between total and subscale scores, and the correlations between the 2 scales were highly significant. In addition, it seems to be valid when compared with other constructs associated with pain and instruments used to measure them, such as disease burden, depression, and impaired ambulation [89].
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	There is an 11-point NRS (item 19) on the GPM which measures current pain that is very similar to that found on the IRF-PAI. In addition, there are 23 other questions [89].
How well does this instrument distinguish people compared with existing questions on IRF-PAI	This instrument provides more information which may be pertinent to rehabilitation than the unidimensional scales that measure only pain intensity undergoing comprehensive geriatric assessment. This additional information includes the effect of pain on function, mood, engagement in activities, and quality of life. For example, functional status and how it is impacted by pain is particularly important among older people. Rehabilitation can be delayed, there can be an increased need for care, and physical function as well as quality of life can be reduced because of pain [89].
Limitations	The analysis was limited to 2 geriatric clinic convenience sample populations and so validity and reliability need to be evaluated further for other general populations and the cognitively impaired. Also, sensitivity to change or to detect change over time was not analyzed, and because this study involved an assisted interview, unassisted methods for subject response, such as a mail or telephone survey, were not explored [89].
Other Comments	According to the authors, the McGill Pain Questionnaire has not been used extensively in older populations but may be a suitable reference for the development of new instruments, as the correlations between the subscale constructs of the GPM and the MPQ were consistent and moderate to strong. That is, both instruments capture highly statistically significant similar constructs. *The full GPM instrument is available in the article* [89].

Minimum Data Set (MDS) [98, 129] (AA)	
Domain	Pain
Purpose (constructs measured, target population)	Associations between certified nursing assistant (CNA) reports of **pain**, MDS reports of pain, and analgesic medication use in **cognitively impaired** nursing home residents. Investigation of correlations between cognitive status, pain reports, and analgesic medication use was a secondary aim [98]. Validity of a pain scale for the MDS **pain** assessment instrument (by comparing it with a VAS/NRS, the article authors refer to the instrument as a VAS), and prevalence of pain in major nursing home subpopulations (including those based on type of admission and cognitive status) [129].
Description	
Number and types of questions	3-item proxy pain questionnaire (PPQ) that assesses presence, frequency, and intensity of pain, the first of which is dichotomous (yes/no) while the other two are rated on a 13-point Likert scale anchored by verbal descriptors (a fourth item assessed CAN beliefs about the relationship between disruptive behavior and pain but was not included in the analyses), and the 2-item MDS pain measure (assessed on a 3-point Likert scale with verbal descriptors). Data collected in two separate phases [98]. 3-items (pain frequency has 3 levels: no pain—if so, skip other pain items, pain less than daily, pain daily; pain intensity has 3 levels: mild pain, moderate pain, times when pain is horrible or excruciating; the third item is a checklist of pain presence/location w/in the last 7 days at the listed sites). The VAS/NRS used for comparison was a 1-item vertical scale, anchored on the bottom by the label of "no pain (0)" and "worst possible pain (10)" at the top with equally spaced intervals of intermediate values representing 1, 2, 3...9 (a copy of the instrument used is not included in the article, and it is unclear if the numerical values are printed on the scale or if they are assigned after the tick marks are made). A simple pain scale was then developed to profile the prevalence of pain [129].
Method of administration	In-person standard interview. No information provided about the literacy level [98]. The MDS reports are based on information provided by the resident, facility staff, resident's physician, and the medical chart. VAS/NRS scores were obtained through in-person interview by the research nurse during which the residents were required to make a tick mark on the scale which was then coded to the corresponding 0-10 number, with 0.5 used for tick marks falling between the whole numbers. No information provided about literacy level. The prevalence of pain analyses conducted among a larger sample of patients, using the validation sample results, was based on MDS reports [129].
Time to administer	Information not provided [98]. No information provided about either instrument [129].

Instrument has been tested with the following populations	The PPQs were completed by 40 CNAs/proxies for 57 residents from 3 nursing homes who were cognitively impaired (mean Mini-Mental State Examination [MMSE] scores of 11.1) and were part of a clinical trial investigating the efficacy of a psychosocial intervention for disruptive vocalization. MDS data collection varies with nursing homes, but the 3 in this study reported that if a resident was able to self-report and the administrator believed in the validity of the self-report then that was what was reflected in the MDS. However, despite the fact that there were no data indicating which residents were directly questioned, the article authors assumed that none of the MDS results were obtained via self-report because of the low average MMSE scores (rather they obtained by direct care staff observation or medical chart review). The pain management medications prescribed during the pain assessments were recorded on a detailed tracking form [98]. Data were collected from 25 MA Medicare-certified SNFs (MDS data were based on the each patient's first assessment). The 94 participants (average age = 81) were presently receiving rehabilitation services or daily skilled nursing, had been admitted from an acute care hospital and stayed in the nursing home for more than 24 hours, and were expected to return—with or w/o home care services—to a community-based setting. Patients were excluded if it was expected that they'd be discharged w/in 24 hours, were dying, or were unable to communicate. A pain scale was derived after the comparison of the MDS and the VAS, which was then used to describe the prevalence of pain for more than 34,000 patients (i.e., all MI nursing home patients from 10/97-10/98) for whom there were individual MDS assessments and also to examine differences in pain levels by admission type and cognitive status [129].
Sensitivity to measure changes in status	Results indicate that the PPQ was a more sensitive measure of pain because while 46% of the sample population were on analgesic medication (aspirin and baby aspirin were excluded from such classification because of the common practice of prescribing them for anticoagulation effects rather than pain relief), pain was reported for only 20% of the sampled residents using the MDS; the PPQ identified 48% of the sample as experiencing pain. This finding of greater PPQ sensitivity is supported by the greater degree of correlation found between the PPQ and analgesic medication use (for frequency, r = .55; for intensity, r = .41); the correlation between the MDS and analgesic use for both construct was 0.33 [98]. "The VAS was chosen as the external standard for its demonstrated sensitivity to pick up a continuum of pain severities and efficiency of administration" [129].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment [98, 129].
Used for population estimates, facility-level estimates or individual level	Individual level [98]. Individual and population levels [129].
Used for quality improvement, quality monitoring, quality measurement	No [98]. The purpose of the Resident Assessment Instrument (RAI)/MDS is to improve care through better care planning based on improved, individualized assessment of a resident's condition which would improve care and lead to better outcomes [129].

Validity/Reliability	
Internal consistency, list relevant statistics	For all of the analyses involving use of analgesics and reports of pain, including estimates of test-retest reliability, Spearman rank-order correlation tests were used. At the two time points for pain report, the 3 items of the PPQ were highly and significantly correlated with each other (for presence, r = 0.84 and p = 0.007; for frequency, r = 0.87 and p = 0.003; for intensity, r = 0.84 and p = 0.006). The PPQ and MDS pain items were significantly correlated with the amount analgesic medications received (PPQ: for presence of pain, r = 0.37, p = 0.0075; for frequency, r = 0.55, p = 0.0001; for intensity, r = 0.41, p = 0.0022. MDS: for both frequency and intensity, r = 0.33, p = 0.02) [98]. 32% of patients indicated no pain, 16% had mild pain, 27% reported moderate pain, and 25% had horrible pain using the MDS assessment items. On the VAS, 41% indicated no pain (the difference between the 32% who reported no pain on the MDS was not significant—p < 0.17) and there was generally good agreement regarding presence of pain between the MDS and the VAS as 87% of those who reported no pain on the MDS scored a 0 on the VAS (kappa = 0.707). In contrast, when the MDS pain scale used in the validation sample was applied to the larger, more representative sample, 47% had no pain and only 4% had horrible pain. When average VAS scores were imputed for each type of patient, post-acute admissions had the highest at 1.5, then other admission (1.3), and then post-admissions (1.0). The percentage of residents reporting no pain increased, as expected and as supported by previous studies) with increasing cognitive impairment in each of the three MI subgroups (e.g., 31% of cognitively intact patients reported no pain as compared to 69% of the most severely impaired among the post-acute admissions group) as shown through application of the Pain Scale developed from the MDS and used to predict VAS scores/prevalence of pain [129].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	For both the frequency and intensity items, significant correlation was not found between the PPQ and the MDS (for frequency, r = 0.18 and p = 0.18; for intensity, r = 0.22 and p = 0.11). In fact, the CNAs reported some level of pain for ~49% of residents while the MDS only did so for ~20% [98]. An index predicting the VAS score was developed from multiple MDS variables (i.e., direct measures of pain, presence of diseases likely to cause substantial pain, measures of physical function, and the CPS scale to identify dementia that could change perception of pain) by using the tree-generation approach of Automatic Interaction Detection. This analysis indicated that the direct measures of pain (frequency, with a variance explanation of 53%; and intensity with a VE of 42%) were much better at predicting VAS-scored pain when compared to all other measures considered (the variable of walking in room had the next largest VE at 11%). The best MDS Pain Scale for predicting the VAS score was a four level one that included no pain, pain less than daily, and daily divided into two groups (moderate and excruciating) which, when used with the 2 MDS variables of frequency and intensity, resulted in a VE of 56% [129].
Ceiling or floor effects	No ceiling or floor effects noted at this time.

IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	The PPQ has a pain intensity item that is similar to that currently on the IRF-PAI although no numeric categorical values are assigned. The MDS does not [98]. The VAS used to validate the MDS/MDS Pain Scale is similar to the NRS/VAS currently on the IRF-PAI if the numbers are indeed marked on the scale although in this study the patients had to manually mark the scale, whereas it is probably done by the interviewer with the IRF-PAI (i.e., the method of administration differs slightly). The MDS questions/those used to develop the Pain Scale are not part of the current IRF-PAI [129].
How well does this instrument distinguish people compared with existing questions on IRF-PAI	Rather than just pain intensity, both of the measures examined here also measure frequency of pain [98]. This is unclear, especially given that the instrument used to validate the one put forth here is very similar to the one that is currently on the IRF-PAI. In fact, the authors state that "the VAS is recognized as a "gold standard" measurement of pain [129].
Limitations	Limited psychometric data are available on the PPQ because it was only one measure from a larger, ongoing clinical trial. Also, there is no patient self-report or direct observational data on patient pain [98]. The validation sample was relatively small and predominantly cognitively alert. Because the VAS cannot be used for those who are unable to communicate or the severely cognitively impaired, the Pain Scale probably would not work with such individuals, either. Generalization to particular subgroups in nursing home populations may not be possible [129].
Other Comments	*A copy of the PPQ instrument is included in the article [98]. The 3 MDS direct pain questions are included in this article* [129].

Faces Pain Scale (FPS) [29, 120, 121] (D)	
Domain	Pain
Purpose (constructs measured, target population)	Accuracy, reliability, construct validity, posdictive validity, and bias susceptibility of 4 **pain intensity** measures: the VRS of the MPQ; the FPS; the 21-point Box Scale; and the Gracely Box Scale (adapted from the Descriptor Differential Scale) among cognitively impaired and unimpaired older adults [29]. **Pain intensity (remembered**, not current) in the elderly [120]. **Acute pain in mature (> 55 years old) hospitalized adults** [121].
Description	
Number and types of questions	Three, 1-item pain scales and one, 2-item scale (VRS = 5 descriptors: mild, discomforting, distressing, horrible, or excruciating, scored 1-5 with higher numbers representing more intense pain; FPS = visual depictions of faces representing increasing levels of pain intensity across a 7-face continuum associated with a range of 1-7; BS = row of 21 boxes labeled from 0/no pain to 99/worst pain in increments of 5; GBS = 2 columns, one measuring intensity and the other unpleasantness, of 21 boxes numbered 0-20 with irregularly spaced, based on ratio-scaled values, verbal descriptors attached to 13 of the boxes) were administered 3 times per day for 7 days and patients were asked to assess their current pain each time using each instrument. In addition, during the evening administration patients made retrospective pain ratings (after the current ratings were completed) with each scale according to “usual,“ “worst,“ and “least“ pain levels. On the 7th day, retrospective ratings were also obtained for the previous week using each of the 4 scales in reference to “usual,“ “worst,“ and “least“ pain levels. Throughout week 2, patients were visited only once per day and made the same retrospective ratings they'd made during the evening visits of week 1. On day 14, the patients made retrospective ratings for all of week 2 and for weeks 1 and 2 combined [29]. 1-item scale with 7 faces depicted (Bieri, 1980) [120]. 1-item scale with 7 faces (1980). It was compared against a 0-10 NRS, a vertical 10-cm VAS, and a verbal descriptor scale (VDS) with 6 options. Subjects were asked to rate, using each of the 4 scales, current pain, worst pain since injury/surgery, and the pain of "getting a shot in the arm." 3-4 days later they were then asked to rate the "getting a shot in the arm" pain again on the FPS to determine test-retest reliability [121].
Method of administration	The paper copy instruments were presented to each patient by a research assistant (RA) and the patients either completed them with a pencil, told the RA what their ratings were, or pointed to the ratings they wished to report. In terms of literacy level, the VRS has no overt numeric features and is primarily verbal. The FPS both low verbal and numeric characteristics. The BRS is primarily numeric with low verbal features, and the GBS has both high verbal and numeric features [29]. In-person standard interview. No information provided about the literacy level, but part of the exclusion criteria was inability to read a line of 14-point font for other aspects of the reliability-validity testing (however, reading, writing, and expressive ability are not required for successful use of the tool). The second scale (administered to determine test-retest reliability) was given to the subject upon completion of the first, along with a SASE and instructions; they were each then contacted by telephone 2 weeks later and asked to recall the same pain, to mark the scale with the face corresponding to that pain, and to mail it back in the SASE [120]. For the FPS, patients were instructed to point to the face that corresponded with their present pain. On the NRS, they each selected the number that

Method of administration (continued)	reflected their pain intensity. For the VAS, each patient indicated the point on the line where their pain fell, that point was marked, and a measurement corresponding to a numerical measure of pain was then made. On the VDS, the response options were both read and shown to the patients, who then stated or pointed to the word that best described their pain. No information was provided about the required literacy level although instructions were provided in lay language and patients could direct the investigator how to mark their choice [121].
Time to administer	No information was provided about time to administer [29, 120, 121].
Instrument has been tested with the following populations	The instruments were tested with 90 patients at a subacute care facility in St. Louis, MO who were over 55 years of age, expected to stay for more than 7 days, reported pain at admission, and had no more than a moderate level (minimum MMSE score of 13) of cognitive impairment. 75 subjects completed week 1 (of these, 44% were cognitively unimpaired and 56% were impaired) and 51 completed week 2 (43.1% were unimpaired and 56.9% were impaired). There were no significant differences for demographic variables, mental status, depression, or primary diagnosis between the completers and the drop-outs [29]. A total of 168 volunteers, aged 65 or older with unimpaired vision, hearing, and cognition (i.e., CCSE score of 20 or more, able to hear instructions during initial screening, and able to read a line of 14-point font) [120]. 60 hospitalized older (age range = 55 to 87) adults who experienced acute pain since hospitalization for surgery or trauma, 24 of whom were African American, and the remainder were white [121].
Sensitivity to measure changes in status	"The current study did not demonstrate the sensitivity of the pain scales to interventions." However, to determine if the results were confounded by response perseveration, the constancy and standard deviations of the ratings given were examined. It was determined that there was little evidence of perseveration as pain ratings fluctuated for each visit and day among most of the patients [29]. Txt effects/changes in pain intensity due to intervention were not being measured [120]. Txt effects/changes in pain intensity due to intervention were not being measured. However, paired t-tests revealed a significant difference between the average ratings of worst and current pain, indicating that patients differentiate between various levels of pain and the FPS is able to detect those differences [121].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment [29, 120, 121].
Used for population estimates, facility-level estimates or individual level	Individual level [29, 120, 121].
Used for quality improvement, quality monitoring, quality measurement	No [29, 120, 121].

Validity/Reliability	
Internal consistency, list relevant statistics	Reliability was evaluated using intra-class coefficients (ICCs). The results across most categorizations/score types showed that the Box Scale performed the most consistently and had the highest reliability coefficient (the information presented in the article was in graphical form so exact numerical ICCs are not known); this was found for both mental status groups although the reliability levels for both the BS and the GBS, while still acceptable, were lower for the impaired group [29]. Good reproducibility over time was indicated by a Spearman rank correlation coefficient of 0.93 ($p = 0.01$) for the test-retest procedure [120]. Upon calculating the correlation between the first and the second ratings of pain for "getting a shot in the arm," as assessed with the FPS and with 19 patients (the others were discharged early and not available for retest), there was a positive, statistically significant relationship ($r = 0.102$, $p < 0.001$).
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Construct validity was assessed using factor analysis (the pain scales with the highest factor loadings were determined to have the greatest construct validity) and the relationship of pain ratings aggregated over time to individual retrospective ratings of pain for each scale. The highest average factor loading was found for the Box Scale, and for days 1-14 the BS and GBS had the highest loadings with most exceeding 0.80. In terms of correlation between retrospective daily or weekly ratings of least, usual, and worst pain and the actual daily or weekly/aggregated pain levels provided, the higher correlation values were consistently found for the BS than the other scales. This was true regardless of mental status although the correlation size was slightly lower for the cognitively impaired group [29]. Subjects were asked to rate their level agreement (on a 5-point scale from strongly agreed to strongly disagreed) between the FPS and a choice of 6 constructs (pain, sleepiness, sadness, sourness, boredom and anger), and results indicate stronger agreement with pain than any of the other constructs. It was noted, however, that 4 subjects said that they could not use the scale to communicate pain. To index the degree of agreement on a rank order of the faces from “no pain” to “most pain,” Kendall's W (coefficient of concordance) and its significance test were calculated. Kendall's W for the first assessment of rank-order was 0.96 ($p = 0.000$), indicating near-perfect agreement with the correct ordering of the faces. When such an assessment was made again 2 weeks later, the Kendall's W was similar at 0.95 ($p = 0.000$). Although all age groups demonstrated high degrees of agreement, the coefficients did decrease with increased age (65-74 years, $W = 0.98$; 75-84 years, $W = 0.96$; > 85 years, $W = 0.93$). Agreement variations were most frequently observed in the middle faces (3-5). In terms of interval assessment, the placement of some of the faces that were viewed and placed singly along the scale differed significantly from what was expected (this was not the case when the faces were viewed and placed along the scale simultaneously). In addition, the interval lengths for both single and simultaneous placements significantly differed from the expectation of truly equal intervals (more so for the former than the latter). Differences were also observed across age groups for the single placements, but not by gender. The Kendall's W for single placement along the ordinal scale was 0.89; for simultaneous placement, it was 0.96 [120]. A high degree of concurrent validity was found with the FPS as it had a strong correlation with the 3 other pain scales for ratings of current pain ($r = 0.81$ to 0.94, $p < 0.001$), the strongest of which was with the NRS.

Ceiling or floor effects	Neither ceiling nor floor effects are discussed in this article, but it was noted that only the VRS and FPS had errors (i.e., off-scale ratings, multiple ratings on a single scale, or rating ranges on a single scale) associated with them. Most of the VRS errors resulted from the fact that there was no option for any pain less than "mild" on the scale (thus indicating a possible floor effect). Likewise, many of the errors on the FPS were attributed to patients' feelings that their pain was less than that suggested by the first face on the scale [29].
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	The BRS, the instrument found to be the most valid and reliable in this study, is more similar to the pain assessment instrument used on the IRF-PAI than any of the other instruments explored in this study because it is a horizontal, numerical row/line bounded on the extreme endpoints by "no pain" and "worst pain." The numerical values are different (i.e., the IRF-PAI is in increments of 1 from 0-10, while the BRS is in increments of 5 from 0-99), and one is an 11-point scale while the other is a 21-point scale. Also, pain is not assessed daily, on a retrospective basis, or on a usual, least or worst basis on the IRF-PAI, which only reports the highest level of pain reported by the patient during the assessment period at admission and discharge [29]. The question asked using this scale in this study was in regard to remembered pain intensity rather than current pain intensity, and the scale itself differs from the NRS/VAS presently used on the IRF-PAI [120]. Patients in this study were asked about 3 different types of pain (i.e., worst, current, and "getting a shot in the arm") whereas the IRF-PAI only solicits information about current pain on an entirely different scale. The FPS might be of more use than the one currently on the IRF-PAI if it is believed that the pain measurement may be affected by fatigue, depleted physical or mental energy, literacy, or understanding of English [121].
How well does this instrument distinguish people compared with existing questions on IRF-PAI	These instruments do not so distinguish although the GBS has an affective component because it also measures the unpleasantness of the pain as well as the intensity/sensory component [29]. It is not clear that the FPS does distinguish people differently than the present IRF-PAI pain scale [120]. It is not clear that the FPS does distinguish people differently than the present IRF-PAI pain scale [121].
Limitations	Patients were older, admitted to a subacute care facility, did not have diagnoses of dementia of Alzheimer's disease, and reported pain at admission. The results cannot be separated from the methodology b/c the relationship between the actual pain reported and the retrospective ratings may be dependent on the multiple daily visits to elicit pain ratings which may create a heightened awareness of pain and increased familiarity with the scales. The order of administration of the scales also may be a limitation because it was never varied. Mental status was assessed the day after admission and may have changed over the 14-day course of the study. Not all pain measures were explored. Sensitivity of the scales to interventions was not demonstrated [29]. The sample studied here consisted of relatively highly educated, cognitively intact, noninstitutionalized individuals. A high percentage was women and white/European-American. Culture may play a role in facial representations of pain and is not accounted for in this study. Current pain was not assessed here [120]. Small sample that included only 2 cultural groups, with data collected from patients in one county hospital [121].

Other Comments	The data collected in this study suggest that the 21-point BRS is a reliable and valid measure of pain intensity for older patients, including those with mild-to-moderate cognitive impairment [29]. **Picture Scales** (*Strengths* = easy to administer and score; *Weaknesses* = limited number of response categories, no evidence about relative compliance rates, limited construct validity and relative sensitivity evidence, scores need to be statistically treated as ordinal data); **Descriptor Differential Scale** (*S* = may be more reliable because it has several items, estimates of consistency with measure completion are possible; *W* = some subjects may have difficulty understanding the measure, limited validity and sensitivity research, takes longer to complete than other simple scale measures) [135]. This instrument might be useful for elderly persons who have difficulty with language, conceptual thinking, and verbalization and, thus, may be unable to communicate their pain with traditional scales. *A copy of the FPS instrument is included with the article [120].* Upon asking the subjects to "identify the scale YOU would MOST like to use when telling the nurses how much pain you are having," most selected the FPS as their first choice. Of those who selected the FPS first, many ranked the NRS second and vice-versa. The difference in scale ranking preferences was statistically significant ($p < 0.001$), but the differences between preference rankings by gender or ethnicity were not [121].

Muscoloskeletal Form of the Medical Rehabilitation Follow-Along (MRFA) [130] (EE)	
Domain	Pain
Purpose (constructs measured, target population)	**Quality of daily living** (including physical function, **pain**, satisfaction, and emotional/psychological well-being) among **outpatient rehabilitation patients** [130].
Description	
Number and types of questions	30 questions (developed from the Functional Assessment Screening Questionnaire, the Oswestry Scale, the MPQ, and the Brief Symptom Questionnaire), consisting of word descriptor scales and at least one VAS (more detail can be provided when/if a copy of the instrument is obtained) [130].
Method of administration	Can be administered as an interview or written questionnaire, but in this study it was administered in questionnaire form at the rehab output clinic in NY and then a duplicate questionnaire was given to the patient and the completed form was mailed back to the clinic 1-7 days after the initial clinic visit [130].
Time to administer	7 to 16 minutes.
Instrument has been tested with the following populations	47 patients receiving outpatient rehabilitation services [130].
Sensitivity to measure changes in status	No information was provided in the article regarding analyses of sensitivity to change or to detect change over time [130].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment of txt outcomes, as well as to aid in screening and identification of functional problems to avert secondary complications and to monitor progress of rehab txt [130].
Used for population estimates, facility-level estimates or individual level	Individual level [130].
Used for quality improvement, quality monitoring, quality measurement	No [130].
Validity/Reliability	
Internal consistency, list relevant statistics	Using the ICC approach to reliability analysis, as well as computation of kappa values and the coefficient for test and retest scores, the test-retest reliability of the MRFA Musculoskeletal Form was examined. The ICC values for the items and subscales in the MRFA were generally high (range = 0.36—the only value below 0.60—and 0.97), and 9 of the 10 ICC values for the pain items were greater than 0.60 (the ICC for the visual analog pain scale was 0.97). The overall test-retest reliability for the 10 descriptor items evaluating pain was excellent, with an ICC of 0.83. The kappa statistic was consistent with the ICC, but the absolute values were lower. The analysis of method error values and accompanying coefficients of variation were computed for all values, and the coefficients ranged from 3% to 11% thus indicating very good stability between the 2 sets of scores

Internal consistency, list relevant statistics (continued)	for the MRFA. This suggests that the low ICC values reflect a lack of variability between the test and retest scores (the method error and coefficient of variation are not affected by raw scores which lack variability) [130].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Rasch analysis of the instruments included in the MRFA measure was performed to produce a measure with good clinical precision (reliability) on an interval level and adequate fit characteristics (quantitative validity). In addition, in regard to predictive validity, 47 subjects only completed the initial questionnaire (i.e., did not complete/mail in the retest) and so their scores and demographic characteristics were compared to those of the 47 subjects who completed both questionnaires, using descriptive statistics and independent sample t-tests. The range of t values = 0.22 to 1.63), and no statistically significant differences were revealed across the 2 groups of subjects. Based on description of the MRFA items, it seems that some may be similar to those used in the IRF-PAI (FIM) but the instrument examined in this study is used to assess outpatient rehab, as opposed to inpatient rehab, outcomes [130].
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	This can be assessed when a copy of the instrument is obtained, but the description of the VAS used in the MRFA does not appear to be the same as the scale used in the IRF-PAI as there is no mention of numerical ratings of pain intensity ("the VAS is marked by the patient...") [130].
How well does this instrument distinguish people compared with existing questions on IRF-PAI	The majority of subjects in this study suffered from low back pain or other orthopedic-related disabilities rather than a variety of different medical diagnoses. More research is needed to determine the MRFA's validity [130].
Limitations	Because this is not a unidimensional scale, more than pain intensity is assessed. The multidimensional information obtained with this instrument include: demographic and background medical information, general life satisfaction, basic functional skills, experience related to pain, ability to engage in strenuous activity, patients' feelings, and levels of distress [130].
Other Comments	*The Musculoskeletal Form of the MRFA is not included in this article [130].*

McGill Pain Questionnaire (MPQ) [45, 48] (F)	
Domain	Pain
Purpose (constructs measured, target population)	**Pain** [45]. **Chronic non-malignant pain** [48].
Description	
Number and types of questions	The Short-Form (SF) MPQ has 15 descriptors rated on an intensity scale with 4 categories from which to choose, a Present Pain Intensity (PPI) Index, and a VAS, for a total of 17 items [45]. 23 items (20 are verbal descriptors, ranging from 3-6 words from which the one word most accurately describing present pain is selected—the words are grouped into 4 dimensions of sensory, affective, evaluative and miscellaneous; 1 is a PPI numerical and VRS that runs from "No pain" at "0" to "Excruciating" at "5"; 1 item consists of 3 groups of verbal descriptors of 3 words each from which one group is chosen; and the last item is a drawing on which to indicate present internal and external pain location. The McGill Comprehensive Pain Questionnaire (MCPQ) incorporates the MPQ's pain description items as well as items intended to elicit descriptive, individual information from each patient. The VAS used here was a 10 cm horizontal line with "no pain" at one end and "pain as bad as it could possibly b" at the other [48].
Method of administration	In-person interview (the questionnaire was placed in front of the patient and the interviewer marked where the patient indicated; the patient marked the VAS) [45]. In-person interview [48].
Time to administer	2-5 minutes for the SF, with simple words and intensity rankings that were understood by all subjects; 5-10 minutes for the Long-Form (LF)-MPQ [45]. 15-30 minutes for instruction and admin of the MPQ, 5-10 minutes for scoring [47]. 60 to 90 minutes for the MCPQ, but no information was provided for the other 2 instruments [48].
Instrument has been tested with the following populations	40 post-surgical patients, 20 obstetrical patients, 10 physiotherapy for musculoskeletal patients, plus another 31 post-surgical patients, and 31 dental pain patients [45]. 30 individuals with rheumatic disease and chronic pain who were 21 years of age or older—the MPQ/MCPQ were administered to this group, and 30 patients at a 650-bed hospital who were admitted for general surgery and were experiencing acute pain, were not experiencing chronic pain, and were 21 years of age or older—the MPC and VAS were administered to this group pre- and post-surgery [48].
Sensitivity to measure changes in status	The SF-MPQ is sensitive to traditional clinical therapies (i.e., the mean intensities of pain significantly decreased across all the categories after pharmaceutical interventions) [45]. For PPI and the total number of words chosen from the descriptor scales, the MPQ did not discriminate between the chronic pain group and the acute pain group. However, the affective dimension discriminated significantly ($p < 0.05$) between the 2 groups ($t = 2.21$) [48].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment [45, 48].

Used for population estimates, facility-level estimates or individual level	Individual level [45, 48].
Used for quality improvement, quality monitoring, quality measurement	No [45, 48].
Validity/Reliability	
Internal consistency, list relevant statistics	The pain rating scores (sensory, affective, and total) obtained with both forms of the MPQs administered before and after therapeutic interventions were significantly correlated. For example, in the postsurgical pain category for the sensory scores, the r was 0.68 (P = 0.001) before txt and the r was 0.83 (P = 0.001) after txt [45]. Using Wilk's method, a discriminant analysis of the 4 MPQ dimensions revealed one canonical function (affect) that approached significance (p = 0.059) [48].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	To determine whether the order of presentation of the forms (i.e., long first and then short) affected the correlations obtained between them, a second study was done in which the forms were presented randomly. The high correlations were not affected by the order of presentation (e.g., the correlations of affective scores for the postsurgical pain category when the SF was presented first were r = 0.90, p = 0.001) [45]. The Pearson product moment correlation procedure (r = 0.58) indicated that the MPQ's PPI and the VAS are concurrently valid for measuring current pain intensity [48].
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	The VAS used on the SF-MPQ is similar to the pain intensity scale of the IRF-PAI, with the exception of the lack of 0-10 numbers and notches along the 10 cm line. The word descriptors and PPI are not found on the IRF-PAI [45]. The PPI of the MPQ and the VAS both ask about present pain intensity as does the pain item on the IRF-PAI, but the scales are different [48].
How well does this instrument distinguish people compared with existing questions on IRF-PAI	The SF-MPQ and the LF-MPQ were both able to show differences in the qualities pain characteristics for each pain category and the different effects of therapy on each quality (i.e., it may be capable of discriminating among different pain syndromes) [45]. The affective component of the MPQ distinguishes acute pain from chronic pain patients better than the other dimensions, but no information is provided about how it compares with the IRF-PAI [48].
Limitations	To determine the association between affect and depression or anxiety, further study is needed [48].
Other Comments	*The instrument (SF-MPQ) is included with this article [45]. The SF-MPQ, the LF-MPQ, VAS, VDS, and NRS are all included in this article [47].*

Chronic Pain Experience Instrument (CPEI) [131] (FF)	
Domain	Pain
Purpose (constructs measured, target population)	The **chronic pain experience/personal response to persistent, non-malignant pain** among chronic pain sufferers [131].
Description	
Number and types of questions	24-items, all using visual analogue scaling, on the CPEI (7 other instruments were administered as well); a second administration of the CPEI took place 2 weeks later to determine test-retest reliability [131].
Method of administration	PAPI, with no information about literacy level (although ability to read and speak English was required) [131].
Time to administer	Information about time required for completion was not provided [131].
Instrument has been tested with the following populations	160 individuals diagnosed with rheumatic disease who experienced pain for at least 3 months, were 21 or older, and could speak and read English [131].
Sensitivity to measure changes in status	Using a causal modeling design, staged model testing was performed through multiple regression analysis, revealing that all of the model's significant empirical relationships ($p \leq 0.05$) were in the directions hypothesized. As such, it was demonstrated that pain intensity ($\beta = -0.32$) directly influenced chronic pain experience ($\beta = -0.22$). In fact, pain intensity, as predicted, strongly influenced/was significantly related to several variables. Furthermore, it was shown that a person responded less well to chronic pain as intensity and depression increased [131].
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Assessment [131].
Used for population estimates, facility-level estimates or individual level	Individual level [131].
Used for quality improvement, quality monitoring, quality measurement	No [131].
Validity/Reliability	
Internal consistency, list relevant statistics	The computed coefficient alpha for the 24-item CPEI was 0.85. In terms of item-scale correlations, only 9 items met the criterion range of 0.50 to 0.70. Further analysis, which included inter-item correlations, item-scale correlations, Cronbach's coefficient alpha, and theta coefficient, were conducted to determine which of the 24-items contributed to the internal consistency of the CPEI, and resulted in the retention of 16 items. Based on the test-retest data for the 16-item instrument, the Pearson product-moment correlation procedure revealed an acceptable level of stability ($r = 0.77$). Furthermore, through exploratory factor analysis and the resulting rotated factor matrix, 3 underlying dimensions were identified

Internal consistency, list relevant statistics (continued)	(Distress, Perceived Effects on Functioning, Rest and Sleep), and internal consistency was further examined. The Cronbach's coefficient alpha for the 9-item Distress subscale was 0.84, the Spearman-Brown correlation coefficients for Perceived Functioning Effects and Rest/Sleep (both subscales having fewer than 5 items) were 0.79 and 0.67, respectively [131].
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	The 3 groupings revealed by the rotated factor matrix accounted for 54% of the common variance, and each dimension had eigenvalues greater than the set criterion of 1.00. The multiple regression analysis using a causal modeling design also revealed that situational anxiety ($\beta = 0.26$), pain description ($\beta = 0.50$), and situational depression ($\beta = 0.31$) were positively influenced by pain intensity. Situational anxiety ($\beta = -0.31$) was negatively influenced by duration of pain while dysphoria (a combination of depression, anxiety, and hostility) influenced pain intensity positively. Based on the magnitude of the significant chronic pain experience predictions, construct validity was estimated as moderate ($R^2 = 0.39$) [131].
Ceiling or floor effects	No ceiling or floor effects noted at this time.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	A copy of the instrument is not included with the article and so it is unclear whether any of the items contain the current IRF-PAI pain question. The IRF-PAI does not measure the multidimensional chronic pain experience, but rather assesses pain intensity using a unidimensional scale [131].
How well does this instrument distinguish people compared with existing questions on IRF-PAI	This instrument provides a more comprehensive approach to identifying patients' subjective responses to pain (e.g., "the CPEI measures the personal response to the pain's perceived effects, that is, frustration with ability to carry out responsibilities, interferences with how well activities are performed, and frustration with not being able to do what one wants to do") [131].
Limitations	This instrument was tested only with subjects experiencing chronic pain as a result of rheumatic disease—testing with other chronic pain etiologies is necessary. Data collection was simultaneous for all variables, calling for a causal modeling design and eliminating the possibility of making conclusions about the relationships of the variables over time [131].
Other Comments	The findings that pain intensity is positively related to dysphoria provide support for the hypothesis that personality is an important factor in pain perception. Pain intensity was negatively correlated with chronic pain experience and showed a positive relationship with pain description and situational depression. Other findings include an inverse relationship between pain duration and situational anxiety, but a strong positive one between pain intensity and situational anxiety as well as support for the belief that pain affects and is influenced by depression (i.e., how strongly one perceives pain intensity may be impacted by depression). A significant predictor of the chronic pain experience was shown to be situational depression. *A copy of the CPEI instrument is not included in the article* [131].

Physical Functioning

Physical Functioning

An initial search of PubMed for published literature regarding the measurement of physical functioning in medical rehabilitation from 1993 to 2002 yielded over 4,000 abstracts. Narrowing the search for measurement tools used for outcomes purposes yielded 17 articles. Because of some redundancy among the articles, 13 of these were examined closely for the purposes of this literature review. In the review of this literature and subsequent investigations, it becomes clear that there are numerous tools available to measure physical functioning outcomes. The measurement tools most frequently mentioned and used in studies found in the published literature include:

- Medical Outcomes Study Short Form 36 (SF-36) physical functioning scale
- Functional Independence Measure (FIM)
- Sickness Impact Profile (SIP)
- Nottingham Health Profile (NHP)
- Extended Activities of Daily Living (EADL)

Most often, these tools are administered in-person by a trained health professional, though some are also administered by observation of the individual (and some have been administered in both ways). Documentation in one study suggested that the FIM instrument could reasonably be self-administered. As an outcome measure, these scales are used to obtain baseline physical functioning at some point in time (generally admittance to a rehabilitation facility), then measure improvement in physical functioning over the course of the rehab episode. These five measures are likely the most frequently used by clinicians as they all have been successfully subjected to various tests of validity and reliability.

Additional measurement tools found in the literature include:

- Barthel Index
- Office of Population Censuses and Surveys (OPCS) disability scale
- Physical Activity Scale for the Elderly (PASE)
- Human Activity Profile (HAP)
- Sanford 7 day recall questionnaire (PAR)
- Assessment of Motor and Process Skills (AMPS)

While there are a number of tools available to measure physician functioning, it is clear from the literature that for a general population the SF-36 and EADL scales are dominant. These are cited in most review studies and have been used extensively by clinicians. That said, it is also clear that the FIM is the current standard in rehabilitation settings. The reason for dominance of the FIM appears to be that for a relatively short instrument (average completion time of about 6.1 minutes) the 7-point scale used for items in the FIM allows for the significant variation in results. This is particularly important for those individuals with more severe physical functioning limitations; ceiling effects are less likely and therefore identification of more extreme deficits and improvements more likely using this scale.

Nottingham Health Profile (NHP)	
Domain	Physical Functioning
Purpose (constructs measured, target population)	The purpose of the NHP is to measure generic health status and well-being. It was developed on a general population.
Description	
Number and types of questions	38 statements requiring "Yes" or "No" answers. There are six domains: energy, pain, emotional reactions, sleep, social isolation and physical mobility. Scores can range from 0–100 in each domain, a higher score reflecting a worse level of perceived well-being.
Method of administration	The NHP was developed as a self-assessment tool, though it has also been studied as administered by a health professional.
Time to administer	3 to 13 minutes, with a mean time of 6 minutes.
Instrument has been tested with the following populations	Literature suggests that the NHP has been used successfully with individuals with disabilities, but that there are floor effects/item non-response with severely disabled individuals.
Sensitivity to measure changes in status	Literature suggests validation of the NHP to measure changes in health status over time. Observed mean change has been reported as follows: NHP Overall—10.2; NHP Physical Functioning—8.6.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Literature suggests that the NHP has been used as a needs assessment instrument in rehabilitation settings.
Used for population estimates, facility-level estimates or individual level	Individual estimates, though population estimates (by 10 year age band) have also been reported to allow the NHP to be applied to different populations.
Used for quality improvement, quality monitoring, quality measurement	No evidence in literature reviewed.
Validity/Reliability	
Internal consistency, list relevant statistics	The mobility domain of the NHP has exhibited non-response problems as moderately to severely disabled individuals cannot perform the functions noted (particularly walking). In one study, this non-response was almost 50%. Test-Retest for NHP was reported as follows: NHP Overall: 0.95; NHP Physical Function: 0.76.
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Overlap with the motor portion of the FIM, though the NHP is more limited in eliciting only dichotomous (Yes or No) responses.

Note = Reference information for physical functioning domain begins on page 133.

Ceiling or floor effects	Overall, the NHP has been shown to be effected little by floor and ceiling effects. Defining the "floor" or "ceiling" effect as when 75% of responses are at the scale minimum (floor) or maximum (ceiling), in one study, 0 of 6 of the NHP domains showed a floor or ceiling effect. However, literature also suggests that the NHP is less effective with severely disabled populations, as they cannot perform some of the physical functions, and therefore cannot assess with a "yes" or "no" that they have difficulty. Therefore, the NHP is not sensitive to severely disabled.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	Would supplant much of the motor portion of the FIM, as included in the IRF-PAI. Both instruments (IRF-PAI and NHP) cover locomotion/mobility.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	The mobility portion of the NHP, because of the "yes" or "no" response format, would not be as sensitive to relatively small differences in individuals' physical functioning.
Limitations	No limitations noted at this time.
Other Comments	The pros of the NHP is that it is quick and easy to administer, and has a reasonable history of reliability and validity. It has been used successfully for many years to gauge generic health status, particularly in the U.K. Collection burden for the NHP may be lower than for the IRF-PAI. The cons include item response problems with severely disabled individuals, and less sensitivity to smaller differences in physical functioning.

SF-36 (Including Physical Function Scale)	
Domain	Physical Functioning
Purpose (constructs measured, target population)	The SF-36 is a measure of perceived health status, developed on a general population through the RAND health experiment.
Description	
Number and types of questions	36 items yield scores in eight domains (including physical function). The responses vary from dichotomous (Yes/No) to 6 point verbal rating scales. The questionnaire contains 10 questions related to physical functioning: 2 about social functioning, 4 about role limitations due to physical problems, 5 about mental health, 4 about vitality, 2 related to pain, 5 about general health perceptions, and 1 about change in health. Each domain or subscale score ranges from 0–100, with 100 representing the most desirable score.
Method of administration	The SF-36 can either be self-completed or completed by interview.
Time to administer	7 to 10 minutes, with a mean time of 9 minutes.
Instrument has been tested with the following populations	Literature suggests extensive use of the SF-36 with the elderly, and those with physical impairments. Some evidence that validity and reliability of the SF-36 is low for the cognitively impaired, even when the instrument is performed in face to face interview. Also, some evidence that the SF-36 is not as sensitive as some instruments (e.g., the OPCS) in measuring physical functioning—SF36 scores tend to cluster at the floor, while some other instruments result in scores more widely distributed for the same individuals.
Sensitivity to measure changes in status	Literature suggests validation of the SF-36 to measure changes in health status over time. Observed mean change has been reported as follows: SF-36 Overall—12.44; SF-36 Physical Functioning—15.48.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Literature suggests that the SF-36 has been used both as a needs assessment, case-mix, and outcomes tool. Evidence of use of the SF-36 as an outcome measurement before and after rehabilitation.
Used for population estimates, facility-level estimates or individual level	Individual and population estimates have been developed from the SF-36.
Used for quality improvement, quality monitoring, quality measurement	Literature suggests that the SF-36 has been used to measure change in outcomes before and after rehabilitation. Change scores in the various SF-36 domains is the most common outcome measurement.
Validity/Reliability	
Internal consistency, list relevant statistics	Item response rate has been show to be 100% in some studies. Test-Retest for SF-36 was reported as follows: SF-36 Overall: 0.85; SF-36 Physical Function: 0.74. In cognitively normal patients, Cronbach's alpha scores on the eight dimensions of the SF-36 ranged from 0.545 (social function) to 0.933 (bodily pain). A score of 0.860 was achieved for physical function. In cognitively impaired patients, Chronbach's alpha scores were: 0.413 (social function), 0.812 (physical function) and 0.861 (bodily pain).

Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Correlation (r = 0.528) between SF-36 physical function scale and motor FIM scale.
Ceiling or floor effects	Compared to other measurement tools (Barthel, NHP, EADL, OPCS, FIM), the SF-36 has been shown to be relatively unaffected by floor and ceiling effects; mean SF-36 scores on all scales except physical functioning were similar to those for the general population of similar age. Overall, the SF-36 has been shown to be effected little by floor and ceiling effects. Defining the "floor" or "ceiling" effect as when 75% of responses are at the scale minimum (floor) or maximum (ceiling), in one study, 1 of 8 of the SF-36 domains showed a floor or ceiling effect.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	Would supplant much of the motor portion of the FIM, as included in the IRF-PAI. Both instruments (IRF-PAI and SF-36) cover locomotion/mobility.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	There is some evidence that the questions and scaling of the SF-36 do not permit as much differentiation of physical limitations, particularly among the severely disabled.
Limitations	No limitations noted at this time.
Other Comments	The pros of the SF-36 is that it is very well know, and has been heavily tested, It has been used in many clinical settings. The cons include some evidence of less sensitivity (possible floor) effect in the severely disabled.

Sickness Impact Profile (SIP)	
Domain	Physical Functioning
Purpose (constructs measured, target population)	The SIP is a generic measurement of health status, including physical functioning.
Description	
Number and types of questions	136 items requiring respondents to identify illness behaviors. Scores are summarized in 12 subscales, which are then combined into the Physical Dimension and Psychosocial Dimensions. All 12 subscales are combined to give the final scores. In its original form, higher scores indicate worse health.
Method of administration	The SIP can be self-administered or completed by interview.
Time to administer	Up to 30 minutes.
Instrument has been tested with the following populations	Literature suggests that the SIP has been used as a generic measure of health status.
Sensitivity to measure changes in status	Literature suggests the SIP is sensitive to changes in health status over time. Observed mean change has been reported as follows: SIP Overall—5.24; SIP Physical Functioning—3.79.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Literature suggests that the SIP has been used to measure outcomes of care and individual patient progress. However, some studies have shown the SIP to be not as sensitive to change when used as a disease specific measure.
Used for population estimates, facility-level estimates or individual level	Individual and population estimates have been developed from the SIP.
Used for quality improvement, quality monitoring, quality measurement	The SIP has been studied particularly in patients with low back pain and other musculoskeletal disorders, and in the elderly for self-assessment of health status and quality of life.
Validity/Reliability	
Internal consistency, list relevant statistics	Test-retest for SIP is as follows: SIP Overall: 0.93; SIP Physical Function: 0.94.
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	Patient's advancing age was independently associated with higher scores (lower perceived functioning) on the physical dimension of the SIP.
Ceiling or floor effects	
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	Would supplant much of the motor portion of the FIM, as included in the IRF-PAI. Both instruments (IRF-PAI and SIP) cover ambulation/mobility.

How well does this instrument distinguish people compared with existing questions on IRF-PAI	The ambulation and mobility portions of the SIP, because of the "yes" or "no" response format, would not be as sensitive to relatively small differences in individuals' physical functioning.
Limitations	No limitations noted at this time.
Other Comments	The pros of the SIP include it's relatively wide use to measure general health status. The cons include it's relative time to complete; at 30 minutes, it places a much higher collection burden on individuals and facilities.

Extended Activities of Daily Living (EADL)	
Domain	Physical Functioning
Purpose (constructs measured, target population)	The EADL is widely used as a measure of disability.
Description	
Number and types of questions	22 items thought to be important for daily living at home (grouped into four categories: mobility, kitchen, domestic and leisure). The respondents are asked whether they perform these activities.
Method of administration	The EADL can be self-administered or completed by interview.
Time to administer	Between 1 and 9 minutes, with a mean time of 4.3 minutes.
Instrument has been tested with the following populations	Literature suggests extensive use with the elderly.
Sensitivity to measure changes in status	Literature suggests that the EADL can be used to evaluate patient progress over time.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Literature suggests that the EADL has been used to measure outcomes of care, individual patient progress, and as a risk adjustment tool.
Used for population estimates, facility-level estimates or individual level	Individual and population estimates have been developed from the EADL scale.
Used for quality improvement, quality monitoring, quality measurement	The EADL scale has been used to measure overall health status and quality of life in elderly patients.
Validity/Reliability	
Internal consistency, list relevant statistics	Overall item non-completion has been reported at a relatively high 11% (completion rate of 89%) The items for taking a drink from one room to another and managing one's own garden have the highest proportions of "not relevant" responses. Test-retest reliability (Kappa coefficient) for mobility have been reported as excellent for all 6 items in this subscale, ranging from K = 0.83 to K = 1.00.
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	EADL mobility score correlates well with the OPCS locomotion and FIM locomotion items.
Ceiling or floor effects	Compared to other measurement tools (Barthel, NHP, SF-36, OPCS, FIM), the EADL has been shown to be the most highly affected by floor and ceiling effects. Overall, the SF-36 has been shown to be affected by floor and ceiling effects. Defining the "floor" or "ceiling" effect as when 75% of responses are at the scale minimum (floor) or maximum (ceiling), in one study, 8 of 22 of the EADL domains showed a floor or ceiling effect.

IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	Would supplant much of the motor portion of the FIM, as included in the IRF-PAI. Both instruments (IRF-PAI and EADL cover locomotion/mobility.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	There is some evidence that some items in the EADL, because they are focused on independent living, are not relevant to the disabled. Therefore, the EADL is not effective at differentiation among the severely disabled.
Limitations	No limitations noted at this time.
Other Comments	The pros of the EADL scale include its wide use in measuring independent functioning and it's relatively quick completion time. The cons include the high level of item non-response and significant floor/ceiling effects.

Pre-Morbid Functioning

Pre-Morbid Functioning

The initial PubMed search for pre-morbid functioning was more successful when key word searches were expanded to include the "psycho-social." Articles on both long term (i.e. family functioning and support) and short term (i.e., pre-injury employment status, income and education) predictors have been reviewed. Thus far, 61 articles on pre-morbid functioning have been input in the database.

The literature describing predictors of successful rehabilitation includes studies of stroke, heart attack, traumatic brain injury, and spinal cord injury outcomes - including return to work or community activities, satisfactory home life, and survival. Psychologists made significant contributions to this literature, finding that mental states such as depression, emotional reactivity, debilitating beliefs, and illness behaviors such as acceptance of disability (fatalism) or hyperchondria can significantly impact recovery. Age, cognitive functioning, physical functioning, income level, and disease severity are consistently significant independent predictors of outcomes. Others, such as simple indicators of marital status or living alone, are not always significant predictors when additional variables, such as social or emotional support, social integration, and family functioning, are included as covariates. Thus their independent contribution to outcomes is questionable. For example, one study found that living alone was not a significant predictor of mortality in elderly persons when social activities, depression, severity of illness and complicating morbidity factors were also included as covariates. Another study of the elderly found that living alone was a significant predictor of stroke survival, but this study did not control for other social or emotional factors. Another study found that heart attack survival was independently predicted by social ties including emotional support and social network structures, whereas depression and marital status/living alone had no significant independent predictive ability. Psychologists have put forward models of outcomes in which psychosocial variables intervene with other person-specific and physical factors to influence outcomes.

The literature is convincing regarding the importance of social involvement and emotional support as independent predictors of outcomes. Two areas that are frequently measured are family functioning and social integration. The measures constructed for these areas often overlap in that factors describing emotional support are common to both. There are many studies, containing a variety of measures for these areas, ranging from single or multiple variables used as covariates and untested indices to fully-developed instruments with demonstrated psychometric properties. Several of these more fully developed instruments include the following:

- McMaster Family Assessment Device (FAD) (60 items)
- Social Support Inventory for Stroke Survivors (SSISS) (more than 10 items)
- Lubben Social Network Scale (LSNS) (10 items)
- Debilitating Beliefs Scale (DBS) (11 items)

The McMaster FAD is probably too long to be useful in our current scope of work, but it has been used often in the literature and extensively validated. The SSISS is likely too specific to a subset of the rehabilitation population to be useful in our work. The Lubben SNS has promise

because it was designed specifically for use with elderly populations, is short, and easy to administer and score. The DBS is short and easy to administer but has only been validated for heart attack, and cannot be used for the cognitively impaired.

These four instruments were chosen for inclusion in the "top 4" Matrix developed for CMS for several reasons. All measures are appropriate as individual descriptive measures for the elderly in general, or for the elderly post heart-attack. The measures are short and easy to administer, easy to comprehend for those of low literacy, and have been extensively evaluated and validated in the literature. Each measure can be used for both predicting outcomes and for risk assessment (upon admission), to determine how capable the person may be of getting help for or engaging in activities of daily living post discharge. None of these measures have been used previously as quality measures, so any one could be a valuable contribution. Margaret Stineman, Ph.D., a consultant to CMS on this project, is particularly concerned that current quality measures do not capture pre-morbid functioning aspects such as these, that could predict success in rehab.

Lubben Social Network Scale [36]	
Domain	Pre-Morbid Functioning
Purpose (constructs measured, target population)	The scale measures the nature of social networks, including number, frequency, closeness, reciprocity, and living arrangements, targeted for the elderly.
Description	
Number and types of questions	10 ratings (1-5) each with equal weight; maximum value for scale is 50.
Method of administration	Could be administered in person or as a paper-and-pencil interview (PAPI) survey; questions are straightforward, simple, easy to comprehend.
Time to administer	5 minutes to administer, 2 minutes to score.
Instrument has been tested with the following populations	The measure was designed for and tested on elderly population.
Sensitivity to measure changes in status	Quite sensitive—values range from 1 to 5 points for each rank, with 5 rankings possible.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Measure can be used to predict outcomes (see concurrent validity), and for risk assessment (upon admission) to determine how capable the person may be of getting help for activities of daily living (ADL) post discharge.
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	No.
Validity/Reliability	
Internal consistency, list relevant statistics	Internal consistency established using intercorrelation among the 10 response items, using Cronbach's Alpha (Alpha = 0.70); Concurrent validity using regression of LSNS on three health indicators (LOS in hospital following surgery, mental health, healthy practices). The regression coefficients on the three health measures were of the expected sign and statistically significant, even when other control variables (age, self-reported health at survey, any hospitalizations in the past 6 months) were included in the model.
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	IRF-PAI includes living arrangement as an admission question; this single question has less face validity than the Lubben scale, as living alone could suggest either more or less need for social support.

Note = Reference information for pre-morbid functioning domain begins on page 135.

Ceiling or floor effects	Because these questions refer to existing social interaction and they are tailored to the elderly population, there is a ceiling effect that would render the scale less useful in a younger population. (Marital status and participation in clubs and organizations were dropped from the Berkman-Syme Network Index when creating the Lubben Scale).
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	One question in the scale is on the IRF-PAI; additional question(s) would increase the face validity of the single question now included. There is a social interaction component in the FIM which may capture some of the same information as the Lubben scale (Item P under Social Cognition).
How well does this instrument distinguish people compared with existing questions on IRF-PAI	The existing IRF-PAI question 17: pre-hospital living with: 1-alone; 2-family/relatives; 3-friends; 4-attendants; 5-other does not have an implied ranking as in the Lubben scale: 0-live alone; 1-live with unrelated individuals (paid help); 4-live with other relatives or friends; 5-live with spouse. Adding another question from the 'confidant relationships' or 'helping others' categories of the Lubben scale would contribute information about self-sufficiency and capability. With the addition of these components, the Lubben scale addresses the fact that social structures and organizations can have both positive (supportive) and negative effects (possibly discourage self-help).
Limitations	No limitations noted at this time.
Other Comments	We have a copy of this instrument. Nunnally (1978) established Cronbach's Alpha of at least 0.70 as a threshold criterion for reliability of a research instrument [28].

McMaster Family Assessment Device (FAD)—60 item—General Functioning Subscale (GF) [3, 13]	
Domain	Pre-Morbid Functioning
Purpose (constructs measured, target population)	This subscale measures general family functioning; [13] study tests the measure in a population of elderly stroke survivors to see whether family functioning improves adherence to stroke-care information. In [3] this subscale correlates well with the larger FAD and is seen as a good proxy for the longer 60-item FAD.
Description	
Number and types of questions	12 items in this subscale of the 60-item FAD, see p. 102 of [13] for a list of the items; PAPI, closed-ended items.
Method of administration	PAPI survey to be filled out by the patient's family; I have not seen the actual questions.
Time to administer	The longer 53-item test takes about 20 minutes, so this shorter subscale should take about 5 minutes.
Instrument has been tested with the following populations	Tested with children and with elderly; thought to be especially useful when assessing behavior in emotional circumstances (i.e., stroke care following rehabilitation), which can lead to depression, negatively impacting outcomes, and rehabilitation.
Sensitivity to measure changes in status	Sensitivity to measure change will be assessed when additional information and copy of instrument is obtained.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Measure can be used to predict outcomes (see concurrent validity) and for risk assessment (upon admission) to determine how capable the person may be of getting help for ADL post discharge.
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	No.
Validity/Reliability	
Internal consistency, list relevant statistics	[3] in this study on a population of children, internal consistency: Cronbach's Alpha = 0.86 ; Gutman split-half correlation coefficient = 0.83; Concurrent validity: using Pearson correlations, the study found consistency with the hypothesized signs for correlations with other continuous family variables (marital disharmony; parent's general health; SES/income), and using t-tests, the study found consistency with the hypothesized signs for correlations with other dichotomous family variables (parent(s) arrested; alcohol source of family tension; parent(s) ever treated for "nerves"; marital violence; parental separation; family structure). [13] Construct validity: discriminative ability was assessed using Chi-Square categorical test; the GF subscale was found to be stat significantly different between high and low "adherence to stroke care information" categories (pval < 0.05). Unknown source: construct validity: using 178 older couples aged 60-69, the FAD correlated significantly (r = 0.53) with the Locke-Wallace Marital Adjustment Test scores and was better than the Locke-Wallace at predicting Philadelphia Geriatric Morale Scale scores.

Construct validity & description of relationships with existing IRF-PAI (FIM) questions	[3] Study notes that construct validity is achieved when a body of information from a variety of studies yields consistent information about the measure—this study used a population of children; [13] study finds additional evidence for construct validity, in the context of elderly stroke survivors.
Ceiling or floor effects	Seems robust to ceiling and floor because it assesses families, not individuals—additional comments about ceiling and flooring effects will be added when instrument is obtained.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	Nothing like this is on the IRF-PAI.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	No comments noted at this time.
Other Comments	We do not have a copy of the FAD or the GF subscale, at this time.

Illness Behavior Assessment Schedule (IBAS) [6, 7, 41, 42]	
Domain	Pre-Morbid Functioning
Purpose (constructs measured, target population)	This questionnaire attempts to measure abnormal illness behavior (AIB), especially hypochondriacal reaction, which has been shown to be associated with impeded rehabilitation in elderly stroke and chronic back pain patients. The questions would have to be administered by an interviewer; questions are not directed at the individual, they're directed toward the psychologist making the assessment. The goal of one study [6] was to see whether the Illness Behavior Questionnaire (IBQ) could be used to identify which patients developed AIB during rehabilitation (between admission and discharge) for stroke.
Description	
Number and types of questions	19 questions, PAPI, closed-ended items.
Method of administration	To be completed by a professional psychologist/ interviewer.
Time to administer	10 minutes.
Instrument has been tested with the following populations	Has been tested with elderly stroke patients in a rehabilitation setting. The first six questions assess whether the patient recalls having received an explanation concerning their health status, which could be used as a cognitive screen.
Sensitivity to measure changes in status	Not very—questions are all couched in the second person, written for subjective assessment by the interviewer.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Measure can be used to predict outcomes (see concurrent validity) but not for risk assessment (upon admission)—patients who develop AIB often develop it during rehabilitation (there is little direct evidence that it is a predisposing personality factor). AIB has been associated in the psychology literature with failure to recover or rehabilitate as well as expected, given the severity of the morbidity and complicating factors. It has been found to be a key determinant of long-term disability, in studies of stroke patients and patients with chronic back pain [6, 7, 41, 42].
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	No.
Validity/Reliability	
Internal consistency, list relevant statistics	Concurrent validity: in regression analysis, AIB is found to be a strong predictor of (positively correlated with) functional competence and performance (ADL) at discharge and 12 months later [7]. Kappa coefficients were used to assess inter-rater reliability [42]. Kappa scores were low when two different psychiatrists interviewed the same patient 24 hours apart, but agreement improved considerably when the second rater evaluated the subject based on a video taping of the first rater's interview.

Construct validity & description of relationships with existing IRF-PAI (FIM) questions	No comparable questions are in the IRF-PAI. One study [6] provides construct validity to the clustering methodology used to determine AIB versus non-AIB status—the poor scores of the patients in the AIB group followed the expected pattern for patients with AIB. Another study [42] assessed construct validity: discriminative ability was assessed by comparison of means t-tests across groups rated as AIB and non-AIB based on the IBAS; mean scores were then calculated for each group on the IBQ items. If the IBQ and IBAS were concurrently valid, we would expect that the discrimination into groups using the IBAS would yield significant differences for scores based on the IQB across groups. Significant differences were found for 4 of the 6 subscales: disease conviction, psychological/somatic, affective disturbance, and denial.
Ceiling or floor effects	No obvious ceiling or floor effects.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	Nothing like this is on the IRF-PAI. Neuroticism is a human personality characteristic shown to be correlated with depressive symptoms following stroke [7], and neuroticism may be associated with greater vulnerability to AIB. The IRF-PAI does not capture depression or AIB, and these may be related to each other and to the prognosis for success in rehabilitation.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	No comments noted at this time.
Other Comments	We have a copy of the 19-item IBAS. This instrument is of questionable practical value because it requires a psychologist to conduct the assessment. IBAS was developed as a basis for a systematic and standardized approach to illness behavior. Its precursor was the IBQ, a self-report instrument whose use has demonstrated the need for a standardized interview method for gathering the data. The IBAS seeks to provide such a standardized method, and thus to improve the reliability and validity of the measures of illness behavior. The IBAS is predicated on the assumption that a subject's responses at the time of a single interview will reflect current illness behavior well, obviating the need for additional interviews. The IBAS was found to have good inter-rater reliability in some situations and good construct and concurrent validity.

Debilitating Beliefs Scale (DBS) [37]	
Domain	Pre-Morbid Functioning
Purpose (constructs measured, target population)	The purpose of this measure is to identify personality and cognitive variables that predict poor adjustment following myocardial infarction.
Description	
Number and types of questions	There are 11 items in the DBS scale. To answer each question, patients are told to circle the appropriate number (1-6) from "strongly disagree" to "strongly agree."
Method of administration	Could be administered in person or as a PAPI survey; questions are straightforward, simple, easy to comprehend.
Time to administer	5 minutes.
Instrument has been tested with the following populations	Tested in Scandinavian men following heart attack. Not suitable for the cognitively impaired.
Sensitivity to measure changes in status	The measure of debilitating beliefs at admission is found to be a reliable and valid predictor of emotional distress in the adjustment period post-discharge. Debilitating beliefs are found to be very stable over time—suggesting that these beliefs are rather entrenched; these don't change during the experience of hospitalization. Similar persistence is noted for emotional distress.
Application/Relevance	
Used for case mix, outcome prediction, admission assessment	Emotional reactivity and debilitating beliefs were found to be strong predictors of psychosocial adjustment after heart attack; debilitating beliefs had a strong negative relationship with rehabilitation outcomes. These findings suggest that an assessment of the patient's interpretation of the illness at admission can be used to predict rehabilitation outcome.
Used for population estimates, facility-level estimates or individual level	Individual level.
Used for quality improvement, quality monitoring, quality measurement	No.
Validity/Reliability	
Internal consistency, list relevant statistics	For the DBS Cronbach's Alpha = 0.86 suggesting good internal consistency and reliability: [37] cites results from a previous pilot study. For Emotional Reactivity (ER), not a component of the DBS, Cronbach's Alphas = 0.81.
Construct validity & description of relationships with existing IRF-PAI (FIM) questions	No comparable questions are in the IRF-PAI. Concurrent validity: path analysis in this study [37] finds that ER and debilitating beliefs account for 56% of the variance in emotional distress at follow-up, which is consistent with what we would expect if debilitating beliefs is a significant predictor of adjustment to MI. Concurrent validity: [37] cites results from a previous pilot study where regression coefficients from a regression of debilitating beliefs on other outcome measures (number

Construct validity & description of relationships with existing IRF-PAI (FIM) questions (continued)	working hours, participation in social events, ambulation/ independence, perceived health status, and acceptance of disability) had expected signs, which lends concurrent validity to the debilitating beliefs measure.
Ceiling or floor effects	Aimed specifically at the population recovering from heart attack.
IRF-PAI	
Instrument contains questions that are currently in the IRF-PAI or would replace existing questions	Nothing like this is on the IRF-PAI. This study [37] argues that debilitating beliefs are a predisposed trait that can be assessed objectively at admission to rehabilitation.
How well does this instrument distinguish people compared with existing questions on IRF-PAI	No comments noted at this time.
Other Comments	We have a copy of this instrument. This instrument is targeted to a specific population recovering from heart attack.

References

COGNITIVE FUNCTION

1 Hall, K.M., Bushnik, T., Lakisic-Kazazic, B., Wright, J., and Cantagallo, A.: Assessing traumatic brain injury outcome measures for long-term follow-up of community-based individuals. Archives of Physical Medicine and Rehabilitation 82(83):367-374, Mar. 2001.

2 Heruti, R.J., Lusky, A., Danker, R., Ring, H., Dolgopiat, M., Barell, V., Levenkrohn, S., and Adunsky, A.: Rehabilitation outcome of elderly patients after a first stroke: Effect of cognitive status at admission on the functional outcome. Archives of Physical Medicine and Rehabilitation 83(82):742-749, Jun. 2002.

3 Landi, F., Bernabei, R., Russo, A., Zuccala, G., Onder, G., Carosella, L., Cesari, M., and Cocchi, A.: Predictors of rehabilitation outcomes in frail patients treated in a geriatric hospital. Journal of the American Geriatrics Society 50(4):679-684, Apr. 2002.

4 Landi, F., Tua, E., Onder, G., Carrara, B., Sgadari, A., Rinaldi, C., Gambassi, G., Lattanzio, F., and Bernabei, R.: Minimum data set for home care: A valid instrument to assess frail older people living in the community. Medical Care 38(12):1184-1190, Dec. 2000.

5 Gruber-Baldini, A.L., Zimmerman, S.I., Mortimore, E., and Magaziner, J.: The validity of the minimum data set in measuring the cognitive impairment of persons admitted to nursing homes. Journal of the American Geriatrics Society 48(12):1601-1606, Dec. 2000.

6 Snowden, M., McCormick, W., Russo, J., Srebnik, D., Comtois, K., Bowen, J., Teri, L., and Larson, E.B.: Validity and responsiveness of the Minimum Data Set. Journal of the American Geriatrics Society 47(8):1000-1004, Aug. 1999.

7 Lawton, M.P., Casten, R., Parmalee, P.A., Van Haitsma, K., Corn, J., and Kleban, M.H.: Psychometric characteristics of the Minimum Data Set II: Validity. Journal of the American Geriatrics Society 46(6):736-744, Jun. 1998.

8 Frederiksen, K., Tariot, P., and DeJonghe, E.: Minimum Data Set Plus (MDS+) scores compared with scores from five rating scales. Journal of the American Geriatrics Society 44(3):305-309, Mar. 1996.

9 Cullum, C.M., Saine, K., Chan, L.D., Martin-Cook, K., Gray, K.F., and Weiner, M.F.: Performance-based instrument to assess functional capacity in dementia: The Texas Functional Living Scale. Neuropsychiatry, Neuropsychology, & Behavioral Neurology 14(2):103-108, Apr.-Jun. 2001.

10 Park, J.H., Ko, H.J., Kim, J.B., Choi, S.Y., Cho, H.C., and Lee, S.K.: Reliability and validity of the cognitive impairment diagnosing instrument (CIDI) in the elderly. Journal of Korean Medical Science 10(4):287-293, Aug. 1995.

11 Ford, G.R., Haley, W.E., Thrower, S.L., West, C.A., and Harrell, L.E.: Utility of the Mini-Mental State Exam scores in predicting functional impairment among white and African American dementia patients. Journals of Gerontology Series A—Biological Sciences & Medical Sciences 51(4):M185-M188, Jul. 1996.

12 Schmand, B., Lindeboom, J., Hooijer, C., and Jonker, C.: Relation between education and dementia: the role of test bias revisited. Journal of Neurology, Neurosurgery & Psychiatry 59(2):170-174, Aug. 1995.

13 Ihl, R., Frolich, L., Dierks, T., Martin, E.M., and Maurer, K.: Differential validity of psychometric tests in dementia of the Alzheimer type. Psychiatry Research 44(2):93-106, Nov. 1992.

14 Tombaugh, T.N., and McIntyre, N.J.: The mini-mental state examination: A comprehensive review. Journal of the American Geriatrics Society 40(9):922-935, Sep. 1992.

15 Folstein, M.F., Folstein, S.F., and McHugh, P.R.: "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 12(3):189-198, Nov. 1975.

16 Agrell, B., and Dehlin, O.: Mini mental state examination in geriatric stroke patients. Validity differences between subgroups of patients, and relationships to somatic and mental variables. Aging—Clinical & Experimental Research 12(6):439-444, Dec. 2000.

17 Marin, D.B., Flynn, S., Mare, M., Lantz, M., Hsu, M.A., Laurans, M., Paerdes, M., Shreve, T., Zaklad, G.R., and Mohs, R.C.: Reliability and validity of a chronic care facility adaptation of the Clinical Dementia Rating Scale. International Journal of Geriatric Psychiatry 16(8):745-750, Aug. 2001.

18 Brown, G.G., Rahill, A.A., Gorell, J.M., McDonald, C., Brown, S.J., Sillanpaa, M., and Sults, C.: Validity of the Dementia Rating Scale in assessing cognitive function in Parkinson's disease. Journal of Geriatric Psychiatry & Neurology 12(4):180-188, Winter 1999.

19 Fillenbaum, G.G., Peterson, B., and Morris, J.C.: Estimating the validity of the clinical Dementia Rating Scale: the CERAD experience. Consortium to Establish a Registry for Alzheimer's Disease. Aging—Clinical & Experimental Research 8(6):379-385, 1996.

20 Stuss, D.T., Meiran, N., Guzman, D.A., Lafleche, G., and Willmer, J.: Do long tests yield a more accurate diagnosis of dementia than short tests? A comparison of 5 neuropsychological tests. Archives of Neurology 53(10):1033-1039, Oct. 1996.

21 Green, R.C., Woodard, J.L., and Green, J.: Validity of the Mattis Dementia Rating Scale for detection of cognitive impairment in the elderly. Journal of Neuropsychiatry & Clinical Neurosciences 7(3):357-360, Summer 1995.

22 Bobholz, J.H., and Brandt, J.: Assessment of cognitive impairment: Relationship of the Dementia Rating Scale to the Mini-Mental State Examination. Journal of Geriatric Psychiatry & Neurology 6(4):210-213, Oct.-Dec. 1993.

23 Helmes, E., Merskey, H., Hachinski, V.C., and Wands, K.: An examination of psychometric properties of the extended scale for dementia in three different populations. Alzheimer Disease & Associated Disorders 6(4):236-246, Winter 1992.

24 Heun, R., Papassotiropoulous, A., and Jennssen, F.: The validity of psychometric instruments for detection of dementia in the elderly general population. International Journal of Geriatric Psychiatry 13(6):368-380, Jun. 1998.

25 van Belle, G., Uhlmann, R.F., Hughes, J.P., and Larson, E.B.: Reliability of estimates of changes in mental status test performance in senile dementia of the Alzheimer type. Journal of Clinical Epidemiology 43(6):589-595, 1990.

26 Fillenbaum, G.G., Heyman, A., Wilkinson, W.E., and Haynes, C.S.: Comparison of two screening tests in Alzheimer's disease. The correlation and reliability of the Mini-Mental State Examination and the modified Blessed Test. Archives of Neurology 44(9):924-927, Sep. 1987.

27 Blessed, G., Tomlinson, B.E., and Roth, M.: The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. British Journal of Psychiatry 114(512):797-811, Jul. 1968.

28 Hobart, J.C., Lamping, D.L., Freeman, J.A., Langdon, D.W., McLellan, D.L., Greenwood, R.J., and Thompson, A.J.: Evidence-based measurement: which disability scale for neurologic rehabilitation? Neurology 57(4):639-644, Aug. 28, 2001.

29 Hawley, C.A., Taylor, R., Hellawell, D.J., and Pentland, B.: Use of the functional assessment measure (FIM+FAM) in head injury rehabilitation: A psychometric analysis. Journal of Neurology, Neurosurgery & Psychiatry 67(6):749-754, Dec. 1999.

30 Donaghy, S., and Wass, P.J.: Interrater reliability of the Functional Assessment Measure in a brain injury rehabilitation program. Archives of Physical Medicine and Rehabilitation 79(10):1231-1236, Oct. 1998.

31 Alcott, D., Dixon, K., and Swann, R.: The reliability of the items of the Functional Assessment Measure (FAM): Differences in abstractness between FAM items. Disability & Rehabilitation 19(9):355-358, Sep. 1997.

32 McPherson, K.M., Pentland, B., Cudmore, S.F., and Prescott, R.J.: An inter-rater reliability study of the Functional Assessment Measure (FIM + FAM). Disability & Rehabilitation 18(7):341-347, Jul. 1996.

33 Seymour, D.G., Ball, A.E., Russell, E.M., Primrose, W.R., Garratt, A.M., and Crawford, J.R.: Problems in using health survey questionnaires in older patients with physical disabilities. The reliability and validity of the SF-36 and the effect of cognitive impairment. Journal of Evaluation in Clinical Practice 7(4):411-418, Nov. 2001.

34 Voll, R., Krumm, B., and Schweisthal, B.: Functional independence measure (FIM) assessing outcome in medical rehabilitation of neurologically ill adolescents. International Journal of Rehabilitation Research 24(2):123-131, Jun. 2001.

35 Paolinelli, C., Gonzales, P., Doniez, M.E., Donoso, T., and Salinas, V.: Instrument for the functional evaluation of disability rehabilitation. Study on the reliability and clinical experience with the use of functional independence measure. Revista Medica de Chile 129(1):23-31, Jan. 2001.

36 Ottenbacher, K.J., Hsu, Y., Granger, C.V., and Fiedler, R.C.: The reliability of the functional independence measure: A quantitative review. Archives of Physical Medicine and Rehabilitation 77(12):1226-1232, Dec. 1996.

37 Stineman, M.G., Shea, J.A., Jette, A., Tassoni, C.J., Ottenbacher, K.J., Fiedler, R., and Granger, C.V.: The Functional Independence Measure: Tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. Archives of Physical Medicine and Rehabilitation 77(11):1101-1108, Nov. 1996.

38 Pollak, N., Rheault, W., and Stoecker, J.L.: Reliability and validity of the FIM for persons aged 80 years and above from a multilevel continuing care retirement community. Archives of Physical Medicine and Rehabilitation 77(10):1056-1061, Oct. 1996.

39 Yip, J.Y., Wilber, K.H., Myrtle, R.C., and Grazman, D.N.: Comparison of older adult subject and proxy responses on the SF-36 health-related quality of life instrument. Aging & Mental Health 5(2):136-142, May. 2001.

40 Riesma, R.P., Forbes, C.A., Glanville, J.M., Eastwood, A.J., and Kleijnen, J.: General health status measures for people with cognitive impairment: learning disability and acquired brain injury. Health Technology Assessment 5(6):1-100, 2001.

41 Zwecker, M., Levernkrohn, S., Fleisig, G., Ohry, A., and Adunsky, A.: Mini-Mental State Examination, cognitive FIM instrument, and the Loewenstein Occupational Therapy Cognitive Assessment: Relation to functional outcome of stroke patients. Archives of Physical Medicine and Rehabilitation 83(3):342-345, Mar. 2002.

42 Katz, N., Itzkovich, M., Averbuch, S., and Elazar, B.: Loewenstein Occupational Therapy Cognitive Assessment (LOTCA) battery for brain-injured patients: Reliability and validity. American Journal of Occupational Therapy 43(3):184-192, Mar. 1989.

43 Sinoff, G., and Ore, L.: The Barthel activities of daily living index: Self-reporting versus actual performance in the old-old (>or=75 years). Journal of the American Geriatrics Society 45(7):832-836, Jul. 1997.

44 Neese, L.E., Caroselli, J.S., Klaas, P., High, W.M. Jr., Becker, L.J., and Scheibel, R.S.: Neuropsychological assessment and the Disability Rating Scale (DRS): A concurrent validity study. Brain Injury 14(8):719-724, Aug. 2000.

45 Gouvier, W.D., Blanton, P.D., LaPorte, K.K., and Nepomuceno, C.: Reliability and validity of the Disability Rating Scale and the Levels of Cognitive Functioning Scale in monitoring recovery from severe head injury. Archives of Physical Medicine and Rehabilitation 68(2):94-77, Feb. 1987.

46 Rappaport, M., Hall, K.M., Hopkins, K., Belleza, T., and Cope, D.N.: Disability Rating Scale for severe head trauma: Coma to community. Archives of Physical Medicine and Rehabilitation 63(3):118-123, Mar. 1982.

47 Brown, S.A., McCauley, S.R., Levin, H.S., Boake, C., Goldfader, P.R., McCormick, S.D., Rockers, D., Butters, M., Kareken, D.A., Gostnell, D., and Clifton G.L.: Factor analysis of an outcome interview for use in clinical trials of traumatically brain-injured patients: A preliminary study. American Journal of Physical Medicine & Rehabilitation 80(3):196-205, Mar. 2001.

48 Satz, P., Zaucha, K., Forney, D.L., McCleary, C., Asarnow, R.F., Light, R., Levin, H., Kelly, D., Bergsneider, M., Hovda, D., Martin, N., Caron, M.J., Namerow, N., and Becker, D.: Neuropsychological, psychosocial and vocational correlates of the Glasgow Outcome Scale at 6 months post-injury: A study of moderate to severe traumatic brain injury patients. Brain Injury 12(7):555-567, Jul. 1998.

49 Drane, D.L., Yuspeh, R.L., Huthwaite, J.S., Klingler, L.K., Foster, L.M., Mrazik, M., and Axelrod, B.N.: Healthy older adult performance on a modified version of the Cognistat (NCSE): Demographic issues and preliminary normative data. Journal of Clinical & Experimental Neuropsychology 25(1):133-144, Feb. 2003.

50 Wallace, J.J., Caroselli, J.S., Scheibel, R.S., and High, W.M. Jr.: Predictive validity of the Neurobehavioral Cognitive Status Examination (NCSE) in a post-acute rehabilitation setting. Brain Injury 14(1):63-69, Jan. 2000.

51 Marcotte, T.D., van Gorp, W., Hinkin, C.H., and Osato, S.: Concurrent validity of the Neurobehavioral Cognitive Status Exam subtests. Journal of Clinical & Experimental Neuropsychology 19(3):386-395, Jun. 1997.

52 Toedter, L.J., Schall, R.R., Reese, C.A., Hyland, D.T., Berk, S.N., and Dunn, D.S.: Psychological measures: Reliability in the assessment of stroke patients. Archives of Physical Medicine and Rehabilitation 76(8):719-725, Aug. 1995.

53 Mitrushina, M., Abara, J., and Blumefeld, A.: Aspects of validity and reliability of the Neurobehavioral Cognitive Status Examination (NCSE) in assessment of psychiatric patients. Journal of Psychiatric Research 28(1):85-95, Jan.-Feb. 1994.

54 Lamarre, C.J., and Patten, S.B.: A clinical evaluation of the Neurobehavioral Cognitive Status Examination in a general psychiatric inpatient population. Journal of Psychiatry & Neuroscience 19(2):103-108, Mar. 1994.

55 Logue, P.E., Tupler, L.A., D'Amico, C., and Schmitt, F.A.: The Neurobehavioral Cognitive Status Examination: Psychometric properties in use with psychiatric inpatients. Journal of Clinical Psychology 49(1):80-89, Jan. 1993.

56 Schwann, L.H., Van Dyke, C., Kiernan, R.J., Merrin, E.L., and Mueller, J.: The Neurobehavioral Cognitive Status Examination: Comparison with the Cognitive Capacity Screening Examination and the Mini-Mental State Examination in a neurosurgical population. Annals of Internal Medicine 107(4):486-491, Oct. 1987.

57 Hopp, G.A., Dixon, R.A., Grut, M., and Backman, L.: Longitudinal and psychometric profiles of two cognitive status tests in very old adults. Journal of Clinical Psychology 53(7):673-686, Nov. 1997.

58 Mate-Kole, C.C., Major, A., Lenzer, I., and Connolly, J.F.: Validation of the Quick Cognitive Screening Test. Archives of Physical Medicine and Rehabilitation 75(8):867-875, Aug. 1994.

59 Zhang, L., Abreu, B.C., Gonzales, V., Seale, G., Masel, B., and Ottenbacher, K.J.: Comparison of the Community Integration Questionnaire, the Craig Handicap Assessment and Reporting Technique, and the Disability Rating Scale in traumatic brain injury. Journal of Head Trauma Rehabilitation 17(6):497-509, Dec. 2002.

60 Wood-Dauphinee, S., Exner, G., Bostanci, B., Exner, G., Glass, C., Jocheim, K.A., Kluger, P., Koller, M., Krishnan, K.R., Post, M.W., Ragnarsson, K.T., Rommel, T., Zitnay, G., and SCI Consensus Group: Quality of life in patients with spinal cord injury—basic issues, assessment, and recommendations. Restorative Neurology & Neuroscience 20(3-4):135-149, 2002.

61 Whiteneck, G.G., Charlifue, S.W., Gerhart, K.A., Overholser, J.D., and Richardson, G.N.: Quantifying handicap: A new measure of long-term rehabilitation outcomes. Archives of Physical Medicine and Rehabilitation 73(6):519-526, Jun. 1992.

62 Kreutzer, J.S., Marwitz, J.H., Seel, R., and Serio, C.D.: Validation of the neurobehavioral functioning inventory for adults with traumatic brain injury. Archives of Physical Medicine and Rehabilitation 77(2):116-124, Feb. 1996.

63 Alderson, A.L., and Novack, T.A.: Reliable serial measurement of cognitive processes in rehabilitation: The Cognitive Log. Archives of Physical Medicine and Rehabilitation 84(5):668-672, May, 2003.

64 Tsuji, T., Liu, M., Sonoda, S., Domen, K., Tsujiuchi, K., and Chino, N.: Newly developed short behavior scale for use in stroke outcome research. American Journal of Physical Medicine & Rehabilitation 77(5):376-381, Sep.-Oct. 1998.

65 Boss, B.J., Barlow, D., McFarland, S.M., and Sasser, L.: A self-care assessment tool (SCAT) for persons with a spinal cord injury: An expanded abstract. AXONE 17(3):66-77, Mar 1996.

66 Johnston, M.V., Findley, T.W., DeLuca, J., and Katz, R.T.: Research in physical medicine and rehabilitation. XII. Measurement tools with application to brain injury. American Journal of Physical Medicine & Rehabilitation 70(1):40-56, Feb. 1991.

67 Royall, D.R., Mahurin, R.K., and Gray, K.F.: Bedside assessment of executive cognitive impairment: The executive interview. Journal of the American Geriatrics Society 40(12):1221-1226, Dec. 1992.

68 Royall, D.R., Lauterbach, E.C., Cummings, J.L., Reeve, A., Rummans, T.A., Kaufer, D.I., LaFrance, W.C. Jr., and Coffey, C.E.: Executive control function: A review of its promise and challenges for clinical research. A report from the Committee on Research of the American Neuropsychiatric Association. Journal of Neuropsychiatry & Clinical Neurosciences 14(4):377-405, Fall 2002.

69 Kelly, C., Sharkey, V., Morrison, G., Allardyce, J., and McCreadie, R.G.: Nithsdale Schizophrenia Surveys. 20. Cognitive function in a catchment-area-based population of patients with schizophrenia. British Journal of Psychiatry 177:348-353, Oct. 2000.

70 Royall, D.R., Mahurin, R.K., and Cornell, J.: Effect of depression on dementia presentation: Qualitative assessment with the Qualitative Evaluation of Dementia (QED). Journal of Geriatric Psychiatry & Neurology 8(1):4-11, Jan. 1995.

71 Villardita, C., and Lomeo, C.: Alzheimer's disease: Correlational analysis of three screening tests and three behavioral scales. Acta Neurologica Scandinavica 86(6):603-608, Dec. 1992.

72 Foreman, M.D.: Reliability and validity of mental status questionnaires in elderly hospitalized patients. Nursing Research 36(4):216-220, Jul.-Aug. 1987.

73 Nelson, A., Fogel, B.S., and Faust, D.: Bedside cognitive screening instruments. A critical assessment. Journal of Nervous & Mental Disease 174(2):73-83, Feb. 1986.

74 Dalton, J.E., Pederson, S.L., Blom, B.E., and Holmes, N.R.: Diagnostic errors using the Short Portable Mental Status Questionnaire with a mixed clinical population. Journal of Gerontology 42(5):512-514, Sep. 1987.

75 Erkinjuntti, T., Sulkava, R., Wilkstrom, J., and Autio, L.: Short Portable Mental Status Questionnaire as a screening test for dementia and delirium among the elderly. Journal of the American Geriatrics Society 35(5):412-416, May 1987.

76 Wolber, G., Romaniuk, M., Eastman, E., and Robinson, C.: Validity of the Short Portable Mental Status Questionnaire with elderly psychiatric patients. Journal of Consulting & Clinical Psychology 52(4):712-713, Aug. 1984.

77 Smyer, M.A., Hofland, B.F., and Jonas, E.A.: Validity study of the short portable mental status questionnaire for the elderly. Journal of the American Geriatrics Society 27(6):263-269, Jun. 1979.

78 Fogel, B.S., Brock, D., Goldscheider, G., and Royall, D.: Cognitive Dysfunction and the Need for Long-Term Care: Implications for Public Policy. AARP Public Policy Institute, April 1994.

79 Patrick, L., Leber, M., and Johnston, S.: Aspects of cognitive status as predictors of mobility following geriatric rehabilitation. Aging—Clinical & Experimental Research 8(5):328-333, 1996.

80 Royall, D.: Back to the future of mental capacity assessment. Journal of the American Geriatrics Society 50(11):1884-1885, Nov. 2002.

81 Heruti, R.J., Lusky, A., Barell, V., Ohry, A., and Adunsky, A.: Cognitive status at admission: Does it affect the rehabilitation outcome of elderly patients with hip fracture? Archives of Physical Medicine and Rehabilitation 80:432-436, April 1999.

82 Resnick, B., and Daly, M.P.: The effect of cognitive status on outcomes following rehabilitation. Family Medicine; Special Series: Evaluating Functional Status in Older Persons 29(6):400-405, June 1997.

83 Royall, D.R., Chiodo, L.K., and Polk, M.J.: Correlates of disability among elderly retirees with "subclinical" cognitive impairment. Journal of Gerontology 55A(9):M541-M546, 2000.

84 Royall, D.R., Cabello, M., and Polk, M.J.: Executive dyscontrol: An important factor affecting the level of care received by older retirees. Journal of the American Geriatrics Society 46(12):1519-1524, Dec. 1998.

85 Royall, D.R., Corder, J., and Polk, M.: Executive control and the comprehension of medical information by elderly retirees. Experimental Aging Research 23:301-313, 1997.

86 Dodds, T.A., Martin, D.P., Stolov, W.C., and Deyo, R.A.: A validation of the functional independence measurement and its performance among rehabilitation patients. Archives of Physical Medicine and Rehabilitation 74(5):531-536, May 1993.

87 Loewenstein, D., and Mogosky, B.: Ch. 19 The functional assessment of the older patient. In Handbook of Assessment in Clinical Gerontology (Peter A. Lictenberg, Ed.) John Wiley & Sons. pp. 529-554, 1999.

88 MacNeil, S., and Lichtenberg, P.: Ch. 15 Screening instruments and brief batteries for assessment of dementia. In Handbook of Assessment in Clinical Gerontology (Peter A. Lictenberg, Ed.) John Wiley & Sons. pp. 417-441, 1999.

89 Ylvisaker, M., Hanks, R., and Johnson-Green, D.: Rehabilitation of children and adults with cognitive communication disorders after brain injury. ASHA Supplement 23:59-72, 2003.

90 Brown, P.D., and Ricker, J.H.: Evaluating and treating communication and cognitive disorders: Approaches to referral and collaboration for speech-language pathology and clinical neuropsychology. Technical Report. ASHA Supplement 23:47-57, 2003.

91 Goldstein, B., and Hammond, M.: Physical medicine and rehabilitation. Journal of the American Medical Association 277(23):1891-1892, Jun. 18, 1997.

92 Klapow, J.C., Evans, J., Patterson, T.L., Heaton, R.K., Koch, W.L., and Jeste, D.V.: Direct assessment of functional status in older patients with schizophrenia. American Journal of Psychiatry 154(7):1022-1024, Jul. 1997.

93 Baines, K.A., Heeringa, H.M., and Martin, A.W.: Assessment of Language-Related Functional Activities. PRO-ED, 1999.

94 Ross-Swain, D., and Fogle, P.T.: Ross Information Processing Assessment—Geriatric. PRO-ED, 1996.

95 Royall, D.R., Cordes, J.A., and Polk, M.: CLOX: An executive clock drawing task. Journal of Neurology, Neurosurgery & Psychiatry 64(5):588-594, May 1998.

96 Royall, D.R.: Personal correspondence/letter, dated 28 July 1998. University of TX, Health Science Center at San Antonio, July 1998.

DEPRESSION

1 Brink, T.L., Yesavage, J.A., Lum, O. et al.: Screening tests for geriatric depression. Clinical Gerontologist 1(1):37-43, 1982.

2 D'Ath, P., Katona, P., Mullan, E. et al.: Screening detection and management of depression in elderly primary care attenders. Family Practice 11(3):260-266, 1994.

3 Mossey, J.M., Mutran, E., Knott, K. et al.: Determinants of recovery 12 months after hip fracture: The importance of psychological factors. American Journal of Public Health 79(3):279-286, 1989.

4 Hosking, S.G., Marsh, N.V., and Friedman, P.J.: Poststroke depression: Prevalence, course, and associated factors. Neuropsychology Review 6(3):107-133, 1996.

5 Williams, J.B.W.: A structured interview guide for the hamilton depression rating scale. Archives of General Psychiatry 45:742-747, 1998.

6 Guilmette, T.J., Snow, M., Grace, J. et al.: Emotional dysfunction in a geriatric population: Staff observations and patients' reports. Archives of Physical Medicine and Rehabilitation 73:587-593, 1992.

7 Lyness, J.M., Cox, C., Curry, J. et al.: Older age and the underreporting of depressive symptoms. Journal of the American Geriatrics Society 43:216-221, 1995.

8 Radloff, L.S.: The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement 1(3):385-401, 1977.

9 Sheikh, J.I., and Yesavage, J.A.: Geraiatric Depression Scale (GDS) recent evidence and development of a shorter version. Clinical Gerontology, 165-173, 1986.

10 Davidson, H., Feldman, P.H., and Crawford, S.: Measuring depressive symptoms in the frail elderly. Journal of Gerontology: Psychological Sciences 49(4):159-164, 1994.

11 Shinar, D., Gross, C.R., Price, T.R. et al.: Screening for depression in stroke patients: The reliability and validity of the CES-D scale. Stroke 17(2):241-245, 1986.

12 Mahoney, J., Drinka, T.J., Abler, R., et al.: Screening for depression: Single question versus GDS. Journal of the American Geriatrics Society 42:1006-1008, 1994.

13 Gerety, M.B., Williams, J.W., Murlow, C.D., et al.: Performance of case-finding tools for depression in the nursing home: Influence of clinical and functional characteristics and selection of optimal threshold scores. Journal of the American Geriatrics Society 42:1103-1109, 1994.

14 Parikh, R.M., Robinson, R.G., Lipsey, J.R. et al.: The impact of poststroke depression on recovery in activities of daily living over a 2-year follow-up. Archives of Neurology 47:785-789, 1990.

15 Lyness, J.M., Noel, T.K., Cox, C. et al.: Screening for depression in elderly primary care patients. Archives of Internal Medicine 157:449-454, 1997.

16 Yesavage, J.A., and Brink, T.L.: Development and validation of a geriatric depression screening scale: A preliminary report. Journal of Psychiatric Research 17(1):37-49, 1983.

17 Lichtenberg, P.A.: The DOUR Project: A Program of depression research in geriatric rehabilitation minority inpatients. Rehabilitation Psychology 42(2):103-114, 1997.

18 Robinson, R.G., Starr, L.B., Kubos, K.L. et al.: A two-year longitudinal study of post-stroke mood disorders: Findings during the initial evaluation. Stroke 14(5):736-741, 1983.

19 Ng, K.C., Chan, K.L., and Straughan, P.T.: A study of post-stroke depression in a rehabilitative center. Acta Psychiatrica Scandinavica 92:75-79, 1995.

20 Gillen, R., Tennen, H., McKee, T.E. et al.: Depressive symptoms and history of depression predict rehabilitation efficiency in stroke patients. Archives of Physical Medicine and Rehabilitation 82:1645-1649, 2001.

21 Arfken, C.L., Lichtenberg, P.A., and Tancer, M.E.: Cognitive impairment and depression predict mortality in medically ill older adults. Journal of Gerontology: Medical Sciences 54A(3):152-156, 1999.

22 Burke, W.J., Nitcher, R.L., Roccaforte, W.H. et al.: A prospective evaluation of the geriatric depression scale in an outpatient geriatric assessment center. Journal of the American Geriatrics Society 40:1227-1230, 1992.

23 NIH Consensus Development Panel on Depression in Late Life. Diagnosis and treatment of depression in late life. JAMA 268(8):1018-1024, 1992.

24 Lesher, E.L., and Berryhill, J.S.: Validation of the Geriatric Depression Scale- Short Form Among Inpatients. Journal of Clinical Psychology 50(2):256-260, 1994.

25 Garrard, J., Rolnick, S.J., Nitz, N.M. et al.: Clinical detection of depression among community-based elderly people with self-reported symptoms of depression. Journal of Gerontology: Medical Sciences 53A(2):92-101, 1998.

26 Van Marwijk, H.W., Wallace, P., De Bock, G.H. et al.: Evaluation of the feasibility, reliability, and diagnostic value of shortened versions of the Geriatric Depression Scale. British Journal of General Practice 45:195-199, 1995.

27 Hoyl, M.D., Alessi, C.A., Harker, J.O. et al.: Development and testing of a five-item version of the Geriatric Depression Scale. Journal of the American Geriatrics Society 47:873-878, 1999.

28 Alden, D., Austin, C., and Sturgeon, R.: A correlation between the Geriatric Depression Scale long and short forms. Journal of Gerontology: Psychological Sciences 44(4):124-125, 1989.

29 Jackson, R., and Baldwin, B.: Detecting depression in elderly medically ill patients: The use of the Geriatric Depression Scale compared with medical and nursing observations. Age and Ageing 22:349-353, 1993.

30 Epstein, L.J.: Symposium on age differentiation in depressive Illness: Depression in the elderly. Journal of Gerontology 31(3):278-282, 1976.

31 Wall, J.R., Lichtenberg, P.A., Macneill, S.E. et al.: Depression detection in geriatric rehabilitation: Geriatric Depression Scale Short Form vs. Long Form. Clinical Gerontologist 20(3):13-21, 1999.

32 Schubert, D.S., Burns, R., and Paras, W.: Increase of medical hospital length of stay by depression in stroke and amputation patients: A pilot study. Psychotherapy and Psychosomatics 57:61-66, 1992.

33 Gillen, R., Eberhardt, T.L., and Tennen, H.: Screening for depression in stroke: relationship to rehabilitation efficiency. Journal of Stroke and Cerebrovascular Diseases 8(5):300-306, 1999.

34 Parmelee, P.A., Katz, I.R., and Lawton, M.P.: Depression among institutionalized aged: assessment and prevalence estimation. Journal of Gerontology: Medical Sciences 44(1):22-29, 1989.

35 Robinson, J., Gruman, C., Gaztambide, S. et al.: Screening for depression in middle-aged and older Puerto Rican primary care patients. Journal of Gerontology: Bio Sciences and Medical Sciences 57(5):308-314, 2002.

PAIN

1 Chapman, C.R., Casey, K.L., Dubner, R. et al.: Pain measurement: An overview. Pain 22(1):1-31, 1984.

2 Williams, C.: Outcome assessment in chronic non-cancer pain treatment. Acta Anesthesiologica Scandinavica 45(9):1076-1079, 2001.

3 Caraceni, A., Cherny, N., Fainsinger, R. et al.: Pain measurement tools and methods in clinical research in palliative care: Recommendations of an Expert Working Group of the European Association of Palliative Care. Journal of Pain and Symptom Management 23(3):239-255, 2002.

4 Elsayam, A., Driver, L., and Bruera, E.: The MD Anderson Symptom Control and Palliative Care Handbook, 2nd edition. Houston, TX. University of Texas MD Anderson Cancer Center, 2002.

5 Ingham, J., and Portenoy, R.K.: Oxford Textbook of Palliative Medicine (eds. Doyle, D., Hanks, G.W.C., and MacDonald, N.). New York. Oxford University Press. pp. 209-219, 1993.

6 Turk, D.C., Monarch, E.S., and Williams, A.D.: Cancer patients in pain: Considerations for assessing the whole person. Hematology/Oncology Clinics of North America 16:511-525, 2002.

7 Jensen, M., and McFarland, C.A.: Increasing the reliability and validity of pain intensity measurement in chronic pain patients. Pain 55:195-203, 1993.

8 Jensen, M., Turner, L.R., Turner, J.A., and Romano, J.M.: The use of multiple-item scales for pain intensity measurement in chronic pain patients. Pain 67:35-40, 1996.

9 Jensen, M.P., Turner, J.A., Romano, J.M., and Fisher, L.D.: Comparative reliability and validity of chronic pain intensity measures. Pain 83:157-162, 1999.

10 Smith, W.B., and Safer, M.A.: Effects of present pain level on recall of chronic pain and medication use. Pain 55:355-361, 1993.

11 Linton, S.J., and Gotestam, K.G.: A clinical comparison of two pain scales: Correlation, remembering chronic pain, and a measure of compliance. Pain 17:57-65, 1983.

12 Linton, S.J., and Melin, L.: The accuracy of remembering chronic pain. Pain 13:281-285, 1982.

13 Feine, J.S., Lavigne, G.J., Thuan Dao, T.T. et al.: Memories of chronic pain and perceptions of relief. Pain 77:137-141, 1998.

14 Jensen, M.P., Turner, J.A., and Romano, J.M.: What is the maximum number of levels needed in pain intensity measurement? Pain 58:387-392, 1994.

15 Briggs, M., and Closs, J.S.: A descriptive study of the use of Visual Analogue Scales and Verbal Rating Scales for the assessment of postoperative pain in orthopedic patients. Journal of Pain and Symptom Management 18(6):438-446, 1999.

16 Jensen, M.P., Karoly, P., and Braver, S.: The measurement of clinical pain intensity: A comparison of six methods. Pain 27:117-126, 1986.

17 Jensen, M.P., Turner, J.A., Romano, J.M., and Fisher, L.D.: Comparative reliability and validity of chronic pain intensity measures. Pain 83:57-162, 1999.

18 Jensen, M.P., Chen, C., and Brugger, A.M.: Postsurgical pain outcome assessment. Pain 99:101-109, 2002.

19 Heft, M.W., Gracely, R.H., Dubner, R., and McGrath, P.A.: A validation model for verbal descriptor scaling of human clinical pain. Pain 9:363-373, 1980.

20 Price, D.D., McGrath, P.A., Rafii, A., and Buckingham, B.: The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 17:45-56, 1983.

21 Price, D.D., Bush, F.M., Long, S., and Harkins, S.W.: A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain 56:217-226, 1994.

22 Ohnhaus, E.E., and Adler, R.: Methodological problems in the measurement of pain: A comparison between the verbal rating scale and the visual analogue scale. Pain 1:379-384, 1975.

23 Joos, E., Peretz, A., Beguin, S. et al.: Comment on reliability and reproducibility of visual analogue scale and numeric rating scale for therapeutic evaluation of pain in rheumatic patients. Journal of Rheumatology 18(8):1269-1270, 1991.

24 Carlsson, A.M.: Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. Pain 16:87-101, 1983.

25 Downie, W.W., Leatham, P.A., Rhind, V.M. et al.: Studies with pain rating scales. Annals of Rheumatic Diseases 37:378-381, 1978.

26 Sriwatanakul, K., Kelvie, W., Lasagna, L. et al.: Studies with different types of visual analog scales for measurement of pain. Clinical Pharmacology and Therapeutics 34(2):234-239, 1983.

27 Joyce, C.R.B., Zutshi, D.W., Hrubes, V., Mason, R.M.: Comparison of fixed interval and visual analogue scales for rating chronic pain. European Journal of Clinical Pharmacology 8:415-420, 1975.

28 Chouniere, M., and Amsel, R.: A visual analogue thermometer for measuring pain intensity. Journal of Pain and Symptom Management 11(5):299-311, 1996.

29 Chibnail, J.T., and Tait, R.C.: Pain assessment in cognitively impaired and unimpaired older adults: A comparison of 4 scales. Pain 92:173-186, 2001.

30 Lundeberg, T., Lund, I., Dahlin, L. et al.: Reliability and responsiveness of three different pain assessments. Journal of Rehabilitative Medicine 33:279-283, 2001.

31 Doctor, J.N., Slater, M.A., and Atkinson, J.H.: The descriptor differential scale of pain intensity: An evaluation of item and scale properties. Pain 61:251-260, 1995.

32 Chambers, C.T., Giesbrecht, K., Craig, K.D. et al.: A comparison of faces scales for the measurement of pediatric pain: Children's and parent's ratings. Pain 83:25-35, 1999.

33 Bieri, D., Reeve, R.A., Champion, G.D. et al.: The faces pain scale for the self-assessment of the severity of pain experienced by children: Development, initial validation, and preliminary investigation for ratio scale properties. Pain 41:139-150, 1990.

34 Hicks, C.L., von Baeyer, C.L., Spafford, P.A. et al.: The faces pain scale - revised: Toward a common metric in pediatric pain measurement. Pain 93:173-183, 2001.

35 Wong, D.L., and Baker, C.M.: Pain in children: Comparison of assessment scales. Pediatric Nursing 14(1):9-17, 1988.

36 Abu-Saad, H.H., Pool, H., and Tulkens, B.: Further validity testing of the Abu-Saad Paediatric Pain Assessment Tool. Journal of Advanced Nursing 19:1063-1071, 1994.

37 Gauvain-Piquard, A., Rodary, C., Rezvani, A., and Serbouti, S.: The development of the DEGR: A scale to assess pain in young children with cancer. European Journal of Pain 3: 165-176, 1999.

38 Foster, R.L., and Varni, J.W.: Measuring the quality of children's postoperative pain management: Initial validation of the child/parent Total Quality Pain Management (TQPM) instruments. Journal of Pain and Symptom Management 23(3):201-210, 2002.

39 Bucholz, M., Karl, H.W., Pomietto, M., and Lynn, A.: Pain scores in infants: A modified infant pain scale versus visual analogue. Journal of Pain and Symptom Management 15(2):117-124, 1998.

40 Melzack, R.: The McGill Pain Questionnaire: Major Properties and Scoring Methods. Pain 1:277-299, 1975.

41 Graham, C., Bond, S.S., Gerkovich, M.M., and Cook, M.R.: Use of the McGill Pain Questionnaire in the assessment of cancer pain: Replicability and consistency. Pain 8: 377-387, 1980.

42 Kremer, E.F., Atkinson, Jr., J.H., and Ignezi, R.J.: Pain measurement: the affective dimensional measure of the McGill Pain Questionnaire with a cancer pain population. Pain 12:153-163, 1982.

43 Holroyd, K.A., Holm, J.E., Keefe, F.J. et al.: A multi-center evaluation of the McGill Pain Questionnaire: Results from more than 1700 chronic pain patients. Pain 48:301-311, 1992.

44 Kremer, E., and Atkinson, Jr., J.H.: Pain measurement: Construct validity of the affective dimension of the McGill Pain Questionnaire with chronic benign patients. Pain 11:93-100, 1981.

45 Melzack, R.: The short-form McGill Pain Questionnaire. Pain 30:191-197, 1987.

46 Wright, K.D., Asmundson, G.J.G., and McCreary, D.R.: Factorial validity of the short-form McGill Pain Questionnaire (SF-MPQ). European Journal of Pain 5:279-284, 2001.

47 Flaherty, S.A.: Pain measurement tools for clinical practice and research. Journal of the American Association of Nurse Anesthetists 64:2133-140, 1996.

48 Davis, G.C.: The clinical assessment of chronic pain in rheumatic disease: Evaluating the use of two instruments. Journal of Advanced Nursing 14:397-402, 1989.

49 Scrimshaw, S.V., and Maher, C.: Responsiveness of visual analogue and McGill Pain Scale Measures. Journal of Manipulative and Physiological Characteristics 24(8):501-504, 2001.

50 Seymour, D.G., Ball, A.E., Russell, E.M. et al.: Problems in using health survey questionnaires in older patients with physical disabilities. The reliability and validity of the SF-36 and the effect of cognitive impairment. Journal of Evaluation in Clinical Practice 7(4):411-418, 2001.

51 Andersen, E.M., Gravitt, G.W., Aydelotte, M.E. et al.: Limitations of the SF-36 in a sample of nursing home residents. Age and Aging 28; 562-566, 1999.

52 Dorman, P., Slattery, J., Farrell, B. et al.: Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 Questionnaires after stroke. Stroke 29:63-68, 1998.

53 O'Mahoney, P.G., Rodgers, H., Thomson, R.G. et al.: Is the SF-36 suitable for assessing health status of older stroke patients? Age and Aging 27(1):19-23, 1998.

54 Stadnyk, K., Calder, J., and Rockwook, K.: Testing the measurement of properties of the Short Form-36 Health Survey in a frail elderly population. Journal of Clinical Epidemiology 51(10):827-835, 1998.

55 Kiebzak, G.M., Pierson, L.M., Campbell, M. et al.: Use of the SF36 general health status survey to document health-related quality of life in patients with coronary artery disease: Effect of disease and response to coronary artery bypass graft surgery. Heart and Lung 31(3):207-213, 2002.

56 Franks, P.J., Winterberg, H., and Moffatt, C.J.: Health-related quality of life and pressure ulceration assessment in patients treated in the community. Wound Repair and Regeneration 10:133-140, 2002.

57 Pell, J.P.: Impact of intermittent claudication on quality of life. European Journal of Vascular and Endovascular Surgery 9: 469-472, 1995.

58 Bullinger, M., Alonso, J., Apolone, G. et al.: Translating health status questionnaires and evaluating their quality: The IQOLA project approach. Journal of Clinical Epidemiology 51(11):913-923, 1998.

59 Svensson, E. Construction of a single global scale for muli-item assessments of the same variable. Statistics in Medicine 20:3831-3846, 2001.

60 Ford, M.D., Havstad, S.L., and Kart, C.S.: Assessing the reliability of the EORTC QLQ-C30 in a sample of older African American and Caucasian adults. Quality of Life Research 10:533-541, 2001.

61 Chang, V.T., Hwang, S.S., and Feuerman, M.: Validation of the Edmonton Symptom Assessment Scale. Cancer 88(9):2164-2171, 2000.

62 Husted, J., Gladman, D.D., Farewell, V.T. et al.: Validation of the revised and expanded version of the Arthritis Impact Measurement Scales for patients with psoriatic arthritis. Journal of Rheumatology 23(6):1015-1019, 1996.

63 Weiner, D., Pieper, McConnell, D. et al.: Pain measurement in elders with chronic low back pain: Traditional and alternative approaches. Pain 67:461-467, 1996.

64 Riemsma, R.P., Taal, E., Rasker, J.J. et al.: Evaluation of a Dutch Version of the Aims2 for Patients with Rheumatoid Arthritis. British Journal of Rheumatology 35:755-760, 1996.

65 Ryser, L., Wright, B.D., Aeschlimann, A. et al.: A new look at the Western Ontario and McMaster Universities Osteoarthritis Index Using Rasch Analysis. Arthritis Care and Research 12(5):331-335, 1999.

66 Angst, F., Aeschlimann, A., Steiner, W., and Stucki, G.: Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Annals of Rheumatic Diseases 60:834-840, 2001.

67 Brown, N., Melville, M., Gray, D. et al.: Comparison of the SF-36 health survey questionnaire with the Nottingham Health Profile in long-term survivors of a myocardial infarction. Journal of Public Health Medicine 22(2):167-175, 2000.

68 Kersten, P., Mullee, M.A., Smith, J.A.E. et al.: Generic health status measures are unsuitable for measuring health status in severely disabled people. Clinical Rehabilitation 13:219-228, 1999.

69 van der Windt, D.A.W.M., van der Heijden, G.J.M.G., de Winter, A.F. et al.: The responsiveness of the Shoulder Disability Questionnaire. Annals of Rheumatic Diseases 57:82-87, 1998.

70 Fritz, J.M., and Irrgang, J.J.: A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. Physical Therapy 81(2):776-788, 2001.

71 Reneman, M.F., Jorritsma, W., Schellekens, J.M.H. et al.: Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. Journal of Occupational Rehabilitation 12(3):119-129, 2002.

72 Garratt, A.M., Moffett, J.K., and Farrin, A.J.: Responsiveness of generic and specific measures of health outcome in low back pain. Spine 26(1):71-77, 2001.

73 Jacob, T., Baras, M., Zeev, A. et al.: Low back pain: Reliability of a set of pain measurement tools. Archives of Physical Medicine and Rehabilitation 82:735-742, 2001.

74 Lingard, E.A., Katz, J.N., Wright, R.J. et al.: Validity and responsiveness of the Knee Society clinical rating system in comparison with the AF-36 and WOMAC. The Journal of Bone and Joint Surgery 83-A(12):1856-1864, 2001.

75 Laprade, J.A., and Culham, E.G.: A self-administered pain severity scale for patellofemoral pain syndrome. Clinical Rehabilitation 16:780-788, 2002.

76 Echteld, M.A., van Elderen, T.M.T., van der Kamp, L.J.T.H.: How goal disturbance, coping, and chest pain relate to quality of life: A study among patients waiting for PTCA. Quality of Life Research 10:487-501, 2001.

77 Bolton, J.E., and Humphreys, B.K.: The Bournemouth Questionnaire: A Short-Form Comprehensive Outcome Measure. II. Psychometric Properties in Neck Pain Patients. Journal of Manipulative and Physiological Therapeutics 25(3):141-148, 2002.

78 Gliklich, R.E., Goldsmith, T.A., and Funk, G.F.: Are head and neck specific quality of life measures necessary? Head & Neck 19:474-480, 1997.

79 Rogers, S.N., Lowe, D., Brown, J.S. et al.: A comparison between the University of Wahsington Head and Neck Disease-Specific Measure and the Medical Short Form 36, EORTC QOQ-C33 and EORTC Head and Neck 35. Oral Oncology 34:361-372, 1998.

80 Stewart, W.F., Lipton, R.B., Simon, D. et al.: Validity of an illness severity measure for headache in a population sample of migraine sufferers. Pain 79:291-301, 1999.

81 Stewart, W.F., Lipton, R.B., Simon, D. et al.: Reliability of an illness severity measure for headache in a population sample of migraine sufferers. Cephalalgia 18:44-51, 1998.

82 Wolkenstein, P., Zeller, J., Revuz, J. et al.: Quality-of-life impairment in Neurofibromatosis Type 1. Archives of Dermatology 137:1421-1425, 2001.

83 Skevington, S.M.: Investigating the relationship between pain and discomfort and the quality of life, using the WHOQOL. Pain 76:395-406, 1998.

84 Cardenas, D.D., Turner, J.A., Warms, C.A. et al.: Classification of chronic pain associated with spinal cord injuries. Archives of Physical Medicine and Rehabilitation 83:1708-1714, 2002.

85 Smith, B.H., Penny, K.I., Purves, A.M. et al.: The Chronic Pain Grade questionnaire: validation and reliability in postal research. Pain 71:141-147, 1997.

86 Elliott, A.M., Smith, B.H., Smith, W.C. et al.: Changes in chronic pain severity overtime: the Chronic Pain Grade as a valid measure. Pain 88:303-308, 2000.

87 Tyler, E.J., Jensen, M.P., Engel, J.M. et al.: The reliability and validity of pain interference measures in persons with cerebral palsy. Archives of Physical Medicine and Rehabilitation 83: 236-239, 2002.

88 Buchi, S., Buddeberg, C., Klaghofer, R. et al.: Preliminary validation of PRISM (Pictorial Representation of Illness and Self Measure): A brief method to assess suffering. Psychotherapy and Psychosomatics 71:333-341, 2002.

89 Ferrell, B.A., Stein, W.M., and Beck, J.C.: The geriatric pain measure: validity, reliability, and factor analysis. Journal of the American Geriatrics Society 48(12):1669-1673, 2000.

90 Evers, A.W.M., Taal, Kraaimatt, F.W. et al.: A comparison of two recently developed health status instruments for patients with arthritis: DUTCH-AIMS2 and IRGL. British Journal of Rheumatology 37:157-164, 1998.

91 Vallerand, A.H.: Development and testing of the inventory of functional status—chronic pain. Journal of Pain and Symptom Management 15(2):125-133, 1998.

92 Davis, G.C.: The development of the Pain Management Inventory for patients with arthritis. Journal of Advanced Nursing 24:236-243, 1996.

93 Fishman, B., Pasternak, S., Wallenstein, S.L. et al.: The Memorial Pain Assessment Card. Cancer 60:1151-1158, 1987.

94 Gaston-Johansson, F.: Measurement of pain: The psychometric properties of the pain-o-meter, a simple, inexpensive pain assessment tool that could change health care practices. Journal of Pain and Symptom Management 12(3):172-181, 1996.

95 Koho, P., Aho, S., Watson, P. et al.: Assessment of chronic pain behaviour: Reliability of the method and its relationship with perceived disability, physical impairment and function. Journal of Rehabilitative Medicine 33:128-132, 2001.

96 Tucker, C.L., Slifer, K.J., Dahlquist, L.M. et al.: Reliabilty and validity of the Brief Behavioral Distress Scale: A measure of children's distress during invasive medical procedures. Journal of Pediatric Psychology 26(8):513-523, 2001.

97 Weiner, D., Peterson, B., and Keefe, F.: Evaluating persistent pain in long term care residents: What role for pain maps? Pain 76:249-257, 1998.

98 Fisher, S.E., Burgio, L.D., Thorn, B.E. et al.: Pain assessment and management in cognitively impaired nursing home residents: Association of certified nursing assistant pain report, minimum data set report, and analgesic medication use. Journal of the American Geriatrics Society 50(1):152-156, 2002.

99 Anderson, D.L.: Development of an instrument to measure pain in rheumatoid arthritis: Rheumatoid Arthritis Pain Scale (RAPS). Arthritis Care and Research 45(4):317-323, 2001.

100 Ferraz, M.B., Quaresma, M.R., Aquino, L.R.L. et al.: Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. Journal of Rheumatology 17(8):1022-1024, 1990.

101 Gracely, R.H., and Dubner, R.: Reliability and validity of verbal descriptor scales of painfulness. Pain 29:175-185, 1987.

102 Wilson, G.A.M., and Doyle, E.: Validation of three paediatric pain scores for use by parents. Anaesthesia 51:1005-1007, 1996.

103 Coulton, C.J., Zborowsky, E., Lipton, J. et al.: Assessment of the reliability and validity of the arthritis impact measurement scales for children with juvenile arthritis. Arthritis and Rheumatism 30(7):819-824, 1987.

104 Boeckstyns, M.E.H., and Backer, M.: Reliability and validity of the evaluation of pain in patients with total knee replacement. Pain 38:29-33, 1989.

105 Williams, J.W., Holleman, D.R., and Simel, D.L.: Measuring shoulder function with the shoulder pain and disability index. The Journal of Rheumatology 22(4)727-732, 1995.

106 Slater, M.A., Hall, H.F., Atkinson, H. et al.: Pain and impairment beliefs in chronic low back pain: Validation of the Pain and Impairment Relationship Scale (PAIRS). Pain 44:51-56, 1991.

107 Millard, R.W.: The Functional Assessment Screening Questionnaire: Application for Evaluating Pain-Related Disability. Archives of Physical Medicine and Rehabilitation 70:303-307, 1989.

108 Miller, J., Neelon, V., Dalton, J. et al.: The assessment of discomfort in elderly confused patients: A preliminary study. Journal of Neuroscience Nursing 28(3):175-182, 1996.

109 Williams, R.C.: Toward a set of reliable and valid measures for chronic pain assessment and outcome research. Pain 35:239-251, 1988.

110 Callahan, L.F., Brooks, R.H., Summey, J.A., and Pincue, T.: Quantitative pain assessment for routine care of rheumatoid arthritis patients, using a pain scale based on activities of daily living and a visual analog pain scale. Arthritis and Rheumatism 30(6):630-636, 1987.

111 Varni, J.W., Thompson, K.L., and Hanson, V.: The Varni/Thompson Pediatric Pain Questionnaire. I. Chronic musculoskeletal pain in juvenile rheumatoid arthritis. Pain 28:27-38, 1987.

112 McGrath, P.A.: An assessment of children's pain: A review of behavioral, physiological and direct scaling techniques. Pain 31:147-176, 1987.

113 Thomas, R.J., McEwen, J., and Asbury, A.J.: The Glasgow Pain Questionnaire: A new generic measure of pain; development and testing. International Journal of Epidemiology 25(5): 1060-1067, 1996.

114 Farrar, J.T., Young, J.P., LaMoreaux, L. et al.: Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 94:149-158, 2001.

115 Bergh, I., Sjostrom, B., Oden, A., and Steen, B.: An application of pain rating scales in geriatric patients. Aging Clinical and Experimental Research 12(5):380-387, 2000.

116 LaChapelle, D.L., Hadjistavropoulos, T., and Craig, K.D.: Pain measurement in persons with intellectual disabilities. The Clinical Journal of Pain 15:13-23, 1999.

117 Langley, G.B., and Sheppeard, H.: The visual analogue scale: Its use in pain measurement. Rheumatology International 5:145-148, 1985.

118 Gracely, R.H., and Kwilosz, D.M.: The descriptor differential scale: Applying psychophysical principles to clinical pain assessment. Pain 35:279-288, 1988.

119 Cook, A.J., and Chastain, D.C.: The classification of patients with chronic pain: Age and sex differences. Pain & Aging 6(3):142-151, 2001.

120 Herr, K.A., Mobily, P.R., Kohout, F.J., and Wagenaar, D.: Evaluation of the Faces Pain Scale for use with the elderly. The Clinical Journal of Pain 14(1):29-38, 1998.

121 Stuppy, D.J.: The Faces Pain Scale: Reliability and validity with mature adults. Applied Nursing Research 11(2):84-89, 1998.

122 Keck, J.F., Gerkensmeyer, J.E., Joyce, B.A. et al.: Reliability and validity of faces and word descriptor scales to measure procedural pain. Journal of Pediatric Nursing 11(6):368-374, 1996.

123 Niezgoda, H.E., and Pater, J.L.: A validation study of the domains of the core EORTC Quality of Life Questionnaire. Quality of Life Research 2:319-325, 1993.

124 Kaasa, T., and Wessel, J.: The Edmonton Functional Assessment Tool: Further Development and Validation for Use in Palliative Care. Journal of Palliative Care 17(1):5-11, 2001.

125 Bombardier, C., Melfi, C.A., Paul, J. et al.: Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. Medical Care 33(4):AS131-AS144, 1995.

126 Gronblad, M., Hupli, M., Wennerstrand, P. et al.: Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. The Clinical Journal of Pain 9:189-195, 1993.

127 Keefe, F.J.: Behavioral assessment and treatment of chronic pain: Current status and future directions. Journal of Consulting and Clinical Psychology 50(6):296-911, 1982.

128 Guck, T.P., Fleischer, T.D., Willcockson, J.C. et al.: Predictive validity of the pain and impairment relationship scale in a chronic nonmalignant pain population. Archives of Physical Medicine and Rehabilitation 80:91-95, 1999.

129 Granger, C.V., Ottenbacher, K.J., Baker, J.G., and Sehgal, A.: Reliability of a brief outpatient functional outcome assessment measure. American Journal of Physical Medicine & Rehabilitation 74(6):469-475, 1995.

130 Davis, G.C.: Measurement of the chronic pain experience: Development of an instrument. Research in Nursing & Health 12:221-227, 1989.

131 Rucker, K.S., Metzler, H.M., and Kregel, J.: Standardization of chronic pain assessment: A multiperspective approach. The Clinical Journal of Pain 12(2):94-110, 1996.

132 Bernstein, I.H., Jaremko, M.E., and Hinkley, B.S.: On the Utility of the West Haven-Yale Multidimensional Pain Inventory. Spine 20(8):956-963, 1995.

133 Love, A., Leboeuf, C., and Crisp, T.: Chiropractic chronic low back pain sufferers and self-report assessment methods. Part I. A reliability study of the Visual Analogue Scale, the Pain Drawing and the McGill Pain Questionnaire. Journal of Manipulative and Physiological Characteristics 12(2):21-25, 1989.

134 Jensen, M., and Karoly, P.: Ch. 9 Self-report scales and procedures for assessing pain in adults. In Handbook of Pain Assessment (eds. Turk DC, Melzack R). Guilford Press. pp. 135-151, 1992.

135 McGrath, P.J.: Ch. 10 Behavioral measures of pain. In Measurement of Pain in Infants and Children, Progress in Pain Research and Management (eds. Finley GA, McGrath PJ). IASP Press. pp. 83-102, 1998.

136 Herr, K.A., and Mobily, P.R.: Comparison of selected assessment tools for use with the elderly. Applied Nursing Research 6(1):39-46, 1993.

137 AGS Panel on Chronic Pain in Older Persons: The management of chronic pain in older persons. Journal of the American Geriatric Society 46(5):635-651, 1998.

PHYSICAL FUNCTION

1 Beaton, D.E., Jogg-Johnson, Sheilah, and Bombardier: Evaluation changes in health status: Reliability and responsiveness in five generic health status measures in workers with musculokeletal disorders. Journal of Clinical Epidemiology 50(1):79-93, 1997.

2 Ware, J.E. and Sherborne, C.D.: The MOS 36-item Short Form Health Survey (SF-36): Conceptual Framework and item selection. Medical Care (30):473-483, 1992.

3 Nouri, F.M., and Lincoln, N.B.: An extended activities of daily living scale for stroke patients. Clinical Rehabilitation (1):301-305, 1987.

4 Seymour, D.G., Russell, E.M., and Garratt, A.M.: Problems in using health survey questionaires in older patients with physical disabilities: The reliability and validity of the SF-36 and the effect of cognitive impairment. Journal of Evaluation in Clinical Practice 7(4): 411-418, 2001.

5 Fisher, A.G.: The assessment of IADL motor skills: An application of many-faceted Rasch analysis. The American Journal of Occupational Therapy 47(4):319-329, April 1993.

6 Johansen, K. L., Painter, P., and Kent-Braun, J. A.: Validation of questionnaires to estimate physical activity and functioning in end-stage renal disease. International Society of Nephology 59:1121-1127, 2001.

7 Grey, N., and Kennedy, P.: The functional independence measure: A comparative study of clinician and self rating. Paraplegia 31:457–461, 1993.

8 Dodds, T.A., Martin, D.P., Stolov, W.C., and Deyo, R.: A validation of the functional independence measure. Archive of Physical and Medical Rehabilitation 74: 531–536, May 1993.

9 Bates, B.E., and Stineman, M.E.: Outcome indicators for stroke. Archive of Physical and Medical Rehabilitation 81:1468 1478, November 2000.

10 Wood-Dauphinee, S.L., Opzoomer, M.A., and Williams, J.: Assessment of global function. Archive of Physical and Medical Rehabilitation 69:583-590, August 1988.

11 Kelly, S., and Jessop, E.G.: A comparison of measures of disability and health. Journal of Public Health Medicine 18(2):169-174, 1996.

12 Wolinsky, F.D., Wan, G.J., and Tierney, W.M.: Changes in the SF-36 in 12 months in a clinical. Medical Care 36(11):1589-1598, 1998.

13 Stineman, M.G.: Defining the population, treatments, and outcomes of interest: Reconciling the rules of biology with meaningfulness. American Journal of Physical and Medical Rehabilitation 80:147–159, February 2001.

14 Kutner, N.G., and Jassal, S.V.: Quality of life and rehabilitation of elderly dialysis patients. Seminars in Dialysis 15(2):107–112, March-April 2002.

15 Kim, P., Warren, S., Madill, H., et al.: Quality of life of stroke survivors. Quality of Life Research 8:293-301, 1999.

16 Oleske, D.M., Andersson, G.B.J., Lavender, S.A.: Association between recovery outcomes for work-related low-back disorder and personal, family, and work factors. SPINE 25(10):1259–1265, 2000.

17 Kersten, P., Mullee, M.A., Smith, J.A.E., et al.: Generic health status measures are unsuitable. Clinical Rehabilitation 13:219–228, 1999.

PRE-MORBIDITY FUNCTION

1 Berkman, L.F., Leo-Summers, L., and Horwitz, R.I.: Emotional support and survival after myocardial infarction. Annals of Internal Medicine 117:1003-1009, 1992.

2 Blazer, D.G.: Social support and mortality in an elderly community population. American Journal of Epidemiology 115(5):684-694, 1982.

3 Byles, J., Byrne, C., Boyle, M. et al.: Ontario child health study: reliability and validity of the general functioning subscale of the McMaster family assessment device. Family Process 27:97-104, 1988.

4 Case, R.B., Moss, A.J., Case, N. et al.: Living alone after myocardial infarction: Impact on prognosis. JAMA 267(4):515-519, 1992.

5 Clark, M.S., and Smith, D.S.: Factors contributing to patient satisfaction with rehabilitation following stroke. International Journal of Rehabilitation Research 21:143-154, 1998.

6 Clark, M.S.: Abnormal illness behavior in rehabilitation from stroke. Clinical Rehabilitation 11:162-170, 1997.

7 Clark,M.S., and Smith, D.S.: Psychosocial correlates of outcome following rehabilitation from stroke. Clinical Rehabilitation 13:129-140, 1999.

8 Colantonio, A., Kasl, S., Ostfeld, A.M. et al.: Psychosocial predictors of stroke outcomes in an elderly population. Journal of Gerontology: Social Sciences 48(5):S261-S268, 1993.

9 Corrigan, J.D., Smith-Knapp, K., and Granger, C.V.: Outcomes in the first five years after traumatic brain injury. Archives of Physical Medicine and Rehabilitation 79:298-305, 1998.

10 Cummings, S.R., Phillips, S.L., Wheat, M.E. et al.: Recovery of function after hip fracture: the role of social supports. Journal of the American Geriatrics Society 36:801-806, 1988.

11 Epstein, N.B., Baldwin, L.M., and Bishop, D.S.: The McMaster Family Assessment Device. Journal of Marital and Family Therapy 9(2):171-180, 1983.

12 Evans, R.L., Bishop, D.S., and Haselkorn, J.K.: Factors predicting satisfactory home care after stroke. Archives of Physical Medicine and Rehabilitation 72:144-147, 1991.

13 Evans, R.L., Bishop, D.S., Matlock, A. et al.: Family interaction and treatment adherence after stroke. Archives of Physical Medicine and Rehabilitation 68:513-517, 1987.

14 Evans, R.L., Bishop, D.S., Matlock, A.L. et al.: Prestroke family interaction as a predictor of stroke outcome. Archives of Physical Medicine and Rehabilitation 68:508-512, 1987.

15 Evans, R.L., Pomeroy, S., van der Weele, T. et al.: Reliability of a stroke care information test for family caretakers. International Journal of Rehabilitation Research 8(2):199-201, 1985.

16 Felmingham, K.L., Baguley, I.J., and Crooks, J.: A comparison of acute and postdischarge predictors of employment 2 years after traumatic brain injury. Archives of Physical Medicine and Rehabilitation 82:435-439, 2001.

17 Fiscella, K., and Franks, P.: Individual income, income inequality, health, and mortality: what are the relationships? Health Services Research 35(1):307-318, 2000.

18 Freeman, J.A., Langdon, D.W., Hobart, J.C. et al.: Inpatient rehabilitation in multiple sclerosis: do the benefits carry over into the community? Neurology 52:50-56, 1999.

19 Friedland, J., and McColl, M.: Social support and psychosocial dysfunction after stroke: buffering effects in a community sample. Archives of Physical Medicine and Rehabilitation 68:475-480, 1987.

20 Grand, A., Grosclaude, P., Bocquet, H. et al.: Predictive value of life events, psychosocial factors and self-rated health on disability in an elderly rural French population. Social Science and Medicine 27(12):1337-1342, 1988.

21 Grissom, S.P., and Dunagan, L.: Improved satisfaction during inpatient rehabilitation after hip and knee arthroplasty. American Journal of Physical Medicine and Rehabilitation 80(11):798-803, 2001.

22 Harwood, R.H., Gompertz, P., and Ebrahim, S.: Handicap one year after stroke: Validity of a new scale. Journal of Neurology, Neurosurgery, and Psychiatry 57:825-829, 1994.

23 Harwood, R.H., Prince, M., Mann, A., et al.: Associations between diagnoses, impairments, disability, and handicap in a population of elderly people. International Journal of Epidemiology 27:261-268, 1998.

24 Harwood, R.H., Rogers, A., Dickinson, E. et al.: Measuring handicap: The London handicap scale, a new outcome measure for chronic disease. Quality Health Care 3(1):11-16, 1994.

25 Hoofien, D., Vakil, E., Gilboa, A. et al.: Comparison of the predictive power of socio-economic variables, severity of injury, and age on long-term outcome of traumatic brain injury: Sample specific variables versus factors as predictors. Brain Injury 16(1):9-27, 2002.

26 Johansson, S.E., Konlaan, B.B., and Bygren, L.O.: Sustaining habits of attending cultural events and maintenance of health: A longitudinal study. Health Promotion International 16(3):229-234, 2001.

27 Johnson, T.M., Kincade, J.E., Bernard, S.L. et al.: The association of urinary incontinence with poor self-rated health. Journal of the American Geriatric Society 46(6):693-699, 1998.

28 Kabacoff, R.I., Miller, I.W., Bishop, D.S. et al.: A psychometric study of the McMaster family assessment device in psychiatric, medical, and nonclinical samples. Journal of Family Psychology 3(4):431-439, 1990.

29 Kotila, M., Waltimo, O., Niemi, M. et al.: The profile of recovery from stroke and factors influencing outcome. Stroke 15(6):1039-1044, 1984.

30 Krause, J.S., Sternberg, M., Maides, J. et al.: Employment after spinal cord injury: Differences related to geographic region, gender, and race. Archives of Physical Medicine and Rehabilitation 79:615-624, 1998.

31 Kutner, N.G., Brogan, D., Fielding, B. et al.: Black/white differences in symptoms and health satisfaction reported by older hemodialysis patients. Ethnicity and Disease 10:328-334, 2000.

32 Kutner, N., Lin, L., Fielding, B. et al.: Continued survival of older hemodialysis patients: Investigation of psychosocial predictors. American Journal of Kidney Diseases 24(1):42-49, 1994.

33 Leach, L.R., Frank, R.G., Bouman, D.E. et al.: Family functioning, social support and depression after traumatic brain injury. Brain Injury 8(7):599-606, 1994.

34 Lee, A.J., Huber, J., and Stason, W.B.: Poststroke rehabilitation in older Americans: The Medicare experience. Medical Care 34(8):811-825, 1996.

35 Lindberg, M., Angquist, K., Fodstad, H. et al.: Self-reported prevalence of disability after subarachnoid haemorrhage, with special emphasis on return to leisure and work. British Journal of Neurosurgery 6:297-304, 1992.

36 Lubben, J.E.: Assessing social networks among elderly populations. Family Community Health 11(3):42-52, 1988.

37 Melamed, S., Heruti, I., Shiloh, S. et al.: Emotional reactivity and debilitating beliefs during hospitalization predict future adjustment to first myocardial infarction in men. Scandinavian Journal of Rehabiliation Medicine 31:23-30, 1999.

38 Melamed, S., Ring, H., and Najenson, T.: Prediction of functional outcome in hemiplegic patients. Scandinavian Journal of Rehabiliation Medicine Suppl12:129-133, 1985.

39 Mendes de Leon, C.F., Fillenbaum, G.G., Williams, C.S. et al.: Functional disability among elderly blacks and whites in two diverse areas: The New Haven and North Carolina EPESE. American Journal of Public Health 85(7):994-998, 1995.

40 Miller, I.W., Epstein, N.B., Bishop, D.S. et al.: The McMaster family assessment device: Reliability and validity. Journal of Marital and Family Therapy 11(4):345-356, 1985.

41 Pilowsky, I.: Dimensions of illness behavior as measured by the illness behavior questionnaire: a replication study. Journal of Psychosomatic Research 37(1):53-62, 1993.

42 Pilowsky, I., Bassett, D., Barrett, R. et al.: The Illness Behavior Assessment Schedule: Reliability and validity. International Journal of Psychiatry in Medicine 13(1):11-28, 1983-84.

43 Porell, W., and Miltiades, H.B.: Regional differences in functional status among the aged. Social Science and Medicine 54:1181-1198, 2002.

44 Porell, F.W., and Miltiades, H.B.: Access to care and functional status change among aged Medicare beneficiaries. Journal of Gerontology: Social Sciences 56B(2):S69-S83, 2001.

45 Powell, L.H., and Thoresen, C.E.: Behavioral and physiologic determinants of long-term prognosis after myocardial infarction. Journal of Chronic Disease 38(3):253-263, 1985.

46 Robinson, R.G., Murata, Y., and Shimoda, K.: Dimensions of social impairment and their effect on depression and recovery following stroke. International Psychogeriatrics 11(4):375-384, 1999.

47 Ruberman, W., Weinblatt, E., Goldberg, J. et al.: Psychosocial influences on mortality after myocardial infarction. The New England Journal of Medicine 311(9):552-559, 1984.

48 Sander, A.M., Carosellis, J.S., High, W.M. et al.: Relationship of family functioning to progress in post-acute rehabilitation programme following traumatic brain injury. Brain Injury 16(8):649-657, 2002.

49 Sarkisian, C.A., Liu, H., Ensrud, K.E. et al.: Correlates of attributing new disability to old age. Journal of the American Geriatrics Society 49(2):134-141, 2001.

50 Scherr, P.A., Albert, M.S., Funkenstein, H.H. et al.: Correlates of cognitive function in an elderly community population. American Journal of Epidemiology 128(5):1084-1101, 1988.

51 Seeman, T.E., and Berkman, L.F.: Structural characteristics of social networks and their relationship with social support in the elderly: Who provides support. Social Science and Medicine 26(7):737-749, 1988.

52 Seeman, T.E., Berkman, L.F., Charpentier, P.A. et al.: Behavioral and psychosocial predictors of pyhsical performance: MacArthur studies of successful aging. Journal of Gerontology: Medical Sciences 50A(4):M177-M183, 1995.

53 Seeman, T.E., Bruce, M.L., and McAvay, G.J.: Social network characteristics and onset of ADL disability: McArthur studies of successful aging. Journal of Gerontology: Social Sciences 51B(4):S191-S200, 1996.

54 Seeman, T.E., Charpentier, P.A., Berkman, L.F. et al.: Predicting changes in physical performance in a high-functioning elderly cohort: MacArthur studies on successful aging. Journal of Gerontology: Medical Sciences 49(3):M97-M108, 1994.

55 Stuck, A.E., Walthert, J.M., Nikolaus, T. et al.: Risk factors for functional status decline in community-living elderly people: A systematic literature review. Social Science & Medicine 48:455-469, 1999.

56 Tate, D.G., and Forchheimer, M.: Enhancing community reintegration after inpatient rehabilitation for persons with spinal cord injury. Top Spinal Cord Injury Rehabilitation 4(1):42-55, 1998.

57 Thomas, S.A., Browning, C.J., and Greenwood, K.M.: Rehabilitation of older injured workers. Disability and Rehabilitation 16(3):163-170, 1994.

58 Watts, R., and Perlesz, A.: Psychosocial outcome risk indicator: predicting psychosocial outcome following traumatic brain injury. Brain Injury 13(2):113-124, 1999.

59 Whiteneck, G.G., Charlifue, S.W., Gerhart, K.A. et al.: Quantifying handicap: A new measure of long-term rehabilitation outcomes. Archives of Physical Medicine and Rehabilitation 73:519-526, 1992.

60 Willer, B., Ottenbacher, K.J., and Coad M.L.: The community integration questionnaire. American Journal of Physical Medicine and Rehabilitation 73:103-111, 1994.

61 Williams, R.B., Barefoot, J.C., Califf, R.M. et al.: Prognostic importance of social and economic resources among medically treated patients with angiographically documented coronary artery disease. JAMA 267(4):520-524, 1992.