J Neurosurg Spine (Suppl) 17:141–150, 2012

Quantitative testing in spinal cord injury: overview of reliability and predictive validity

Maxwell Boakye, M.D., M.P.H.,[1] Susan Harkema, Ph.D.,[2] Peter H. Ellaway, Ph.D.,[3] and Andrea C. Skelly, Ph.D., M.P.H.[4]

[1]Outcomes and Translational Research Lab, Center for Advanced Neurosurgery, Department of Neurosurgery, University of Louisville, and Roblex Rex Veteran's Administration Medical Center, Louisville; [2]Department of Neurological Surgery, University of Louisville, and Frazier Rehab Institute, Louisville, Kentucky; [3]Division of Experimental Medicine, Imperial College, Charing Cross Campus, London, United Kingdom; and [4]Spectrum Research Inc., Tacoma, Washington

Object. The objective of this study was to identify commonly used physiological outcome measures and summarize evidence on the reliability and predictive validity of quantitative measures used in monitoring persons with spinal cord injury (SCI).

Methods. A systematic search of PubMed through January 5, 2012, was conducted to identify publications using common outcome measures in persons with SCI and for studies that were specifically designed to evaluate the reliability and predictive validity of selected quantitative measures. Quantitative measures were defined as tests that quantify sensory and motor function, such as amount of force or torque, as well as thresholds, amplitudes, and latencies of evoked potentials that might be useful in studies and monitoring of patients with SCI. Reliability studies reporting interclass correlation coefficients (ICCs) or weighted κ coefficients were considered for inclusion. Studies explicitly evaluating correlation between measures and specific functional outcomes were considered for predictive validity.

Results. From a total of 121 potentially relevant citations, 6 studies of reliability and 4 studies of predictive validity for quantitative tests met the inclusion criteria. In persons with incomplete SCI, ICCs for both interrater and intrarater reliability of electrical perceptual threshold (EPT) were ≥ 0.7 above the sensory level of SCI but were less reliable below the sensory level. Interclass correlation coefficients for interrater and intrarater reliability of the Graded Redefined Assessment of Strength, Sensibility, and Prehension (GRASSP) components ranged from 0.84 to 0.98. For electromyography, the ICC was consistently high for within-day tests. The overall quality of reliability of the majority of studies was poor, due to the potential for selection bias and small sample sizes. No classic validation studies were found for the selected measures, and evidence regarding the predictive validity of the measures was limited. Somatosensory evoked potentials (SSEPs) may be correlated with ambulatory capacity, as well as the Barthel Index and motor index scores, but this correlation was limited for evaluation of bladder function recovery in 3 studies that assessed the correlation between baseline or initial SSEPs and a specific clinical outcome at a later follow-up time. All studies used convenience samples and the overall sample quality was low.

Conclusions. Evidence on the reliability and validity of the quantitative measures selected for this review is limited, and the overall quality of existing studies is poor. There is some evidence for the reliability of the EPT, dermatomal SSEPs, and the GRASSP to suggest that they may be useful in longitudinal studies of patients with SCI. There is a need for high quality studies of reliability, responsiveness, and validity for quantitative measures to monitor the level and degree of SCI.

(http://thejns.org/doi/abs/10.3171/2012.5.AOSPINE1296)

Key Words • **quantitative sensory testing** • **spinal cord injury** • **reliability** • **validity** • **outcome**

The current gold standard for clinical assessment of SCI is the ASIA standard neurological classification, which includes tests of motor and cutaneous sensory function.[2] However, there are limitations to the AIS.[19–21] For sensory function, there is a limiting component of subjectivity, with sensory cutaneous evaluation of each dermatome scored simply as either normal, absent, or abnormal sensation. Abnormal sensation currently includes both heightened and lowered sensitivity, as well as allodynia. For motor function, only the upper and lower limbs are assessed, which only includes 5 muscle groups for each limb. The trunk is not evaluated, making assessment of neurological-level SCI in the thoracic region dependent solely on the sensory evaluation. Notably, the supraspinal pathways that remain intact or recover are not

Abbreviations used in this paper: AIS = ASIA impairment scale; ASIA = American Spinal Injury Association; BMCA = Brain Motor Control Assessment; EMG = electromyography; EPT = electrical perceptual threshold; GRASSP = Graded Redefined Assessment of Strength, Sensibility, and Prehension; ICC = interclass correlation coefficient; MEP = motor evoked potential; MMV = maximal movement velocity; SCI = spinal cord injury; SSEP = somatosensory evoked potential; TPT = thermal perceptual threshold; VPT = vibrational perceptual threshold.

identified using this assessment. In addition, significant concerns exist regarding sensitivity of the AIS to subclinical improvements through clinical trials and regenerative and therapeutic strategies.[30] It is therefore important to be able to develop quantitative tests that can assess neurological function longitudinally. An ideal quantitative test would be reliable, valid, and consistent across raters, and likely be more sensitive and responsive to neurological and subclinical improvement, and recovery of a few spinal segments.[9]

The primary question to be addressed in this paper is as follows: What sensory, motor, and autonomic physiological tests have been assessed for validity and reliability? The purpose of this paper is to provide an overview of studies describing the reliability and predictive validity of selected quantitative measures that may be useful for monitoring regeneration and progress in clinical trials and patient recovery.

Methods

A systematic search of PubMed through January 5, 2012, was conducted to identify publications using common outcomes measures in persons with SCI and for studies that were specifically designed to evaluate the reliability and validity of quantitative measures. There was no restriction on publication date or study type. Searches were limited to studies conducted in humans and published in English. Terms related to traumatic SCI [(((spinal cord injuries[MAJR]) OR (spinal cord injury[TI]) OR (spinal cord injured[TI]) OR (spinal cord injuries[TI]) OR (spinal cord lesions[TI]) OR (spinal cord lesion[TI]) OR (parapleg*[TI]) OR (quadripleg*[TI]) OR (tetrapleg*[TI])) NOT (neoplasms[MAJR] OR cancer*[TIAB] OR cancer*[TW] OR malignan*[TIAB] OR malignan*[TW] OR neoplas*[TIAB] OR neoplas*[TW] OR metasta*[TIAB] OR metasta*[TW]))] were added to specific search strategies for measures as described below.

Identification of Outcome Measures Used in Clinical Studies

Measures specific to patients with SCI in the following areas were identified by the clinical authors as most suitable for inclusion: pain, functional/potential measures, upper extremity potential measures, and motor/sensory measures. PubMed was searched to identify studies using such measures and obtain an estimate of how commonly they have been used in clinical studies. For each of the measures, a search strategy that included relevant key words, acronyms, and MeSH terms was added to the terms related to traumatic SCI described above. Titles and abstracts of studies identified were searched to determine whether the outcomes measure of interest was used in persons with SCI and estimate the number of studies using the measure.

Reliability, Responsiveness, and Predictive Validity of Quantitative Measures

Quantitative measures were defined as tests that quantify sensory, motor, or autonomic function (amount of force or torque, and thresholds, amplitudes, and latencies of evoked potentials) that might be useful for monitoring regeneration and progress in clinical trials and patient recovery in patients with SCI. The following measures were identified by the clinical authors as most suitable for inclusion as quantitative measures: SSEPs, MEPs, dermatomal SSEPs, contact heat-evoked potentials, quantitative sensory testing (EPT, TPT, and VPT), and autonomic measures (sudomotor/quantitative sudomotor axon reflex test, sympathetic skin response, postural challenge/"tilt test," Valsalva maneuver, sweating/thermoregulation sweating, and cardiovagal heart rate).

The systematic search of PubMed combined terms related to traumatic SCI described above with those related to studies of reliability or validity [(Reproducibility of Results[MeSH] OR reliab*[TI] OR valid* OR intertest* OR interobserv* OR intratest* OR intraobserv* OR interrat* OR intrarat* OR validation studies [Publication Type])] For each of the measures listed above, a search strategy that included relevant key words, acronyms, and MeSH terms was added to identify reliability and validity studies for the specific measure. Reference lists of seminal articles were also systematically checked for relevant studies.

Reliability evaluates the extent to which repeated measurements in stable patients (test-retest) yield similar responses.[26] Reproducibility measures whether patients can be differentiated from each other despite measurement error (relative measurement error).[26,31] Reliability studies reporting ICCs or weighted κ coefficients were considered for inclusion. The Pearson correlation coefficient is not considered an adequate measurement of reliability, because it does not account for systematic differences.[31]

The critical appraisal of the quality of reliability studies was based on the following factors: inclusion of a broad spectrum of persons with the expected condition; adequate description of methods for replication; blinded performance of tests, measurements, or interpretation; timing of second test appropriate for stage/period of disease, timing to avoid influence of interpretation from first test; and demonstration of an ICC or weighted κ coefficient ≥ 0.70 when measured in at least 50 patients. A good quality study (Level of Evidence I) meets all 5 of these criteria; a moderate quality study (Level II) meets 4, poor quality (Level III) meets 3, and a very poor quality study meets fewer than 3 of the criteria (Level IV).

Predictive validity refers to the extent to which the measure predicts a specific outcome (patient function based on a validated measure or mortality) in the patient population of interest and is closely related to outcomes. The question is whether a specified change in a measure correlates with a clinically meaningful change in a physical or functional outcome. A measure may be good at predicting one outcome but not another; thus, the outcome needs to be specified and well measured. For a measure to have predictive validity, it should predict outcome in a second population (a population independent from the population used to develop the measure). Studies explicitly evaluating correlation between baseline quantitative measures and specific functional outcomes measured at some later follow-up time were considered for predictive validity. Studies reporting formal statistical analysis using appropriate correlation or regression methods were sought. Studies that failed to report explicit evaluation of such a correlation with a specific outcome were excluded.

Responsiveness assesses whether a measure is able to detect clinically important changes over time (the score changes with the status of the patient).[26,31] Studies that formally evaluated the smallest detectable change, minimally important change, responsiveness ratio, or area under curve for a receiver-operator characteristic curve were sought.

Studies in adults with acute or chronic, complete or incomplete SCI were considered for inclusion if the study was designed to evaluate reliability, responsiveness, or predictive validity as described above. Studies in patients with peripheral nerve injury, spinal root injury, cancer, deformity (including scoliosis), or other neurological conditions were excluded. Studies of fewer than 10 patients with SCI were excluded, as were studies of animals, those with less than 50% of the population comprised of patients with SCI, and those exploring mechanisms or the basic feasibility of measures by comparing them to healthy control patients. The focus of this review is to provide information on the highest quality studies available to answer the clinical question.

Each retrieved citation was reviewed by 2 reviewers working independently. Most articles were excluded on the basis of information provided by the title or abstract. Citations that appeared to be relevant or that could not be unequivocally excluded from the title and abstract were identified, and the corresponding full text reports were evaluated by at least 2 reviewers. Disagreement with respect to inclusion or exclusion of these citations was resolved by consensus. Figure 1 summarizes the results from the literature search and exclusion of studies at various stages.

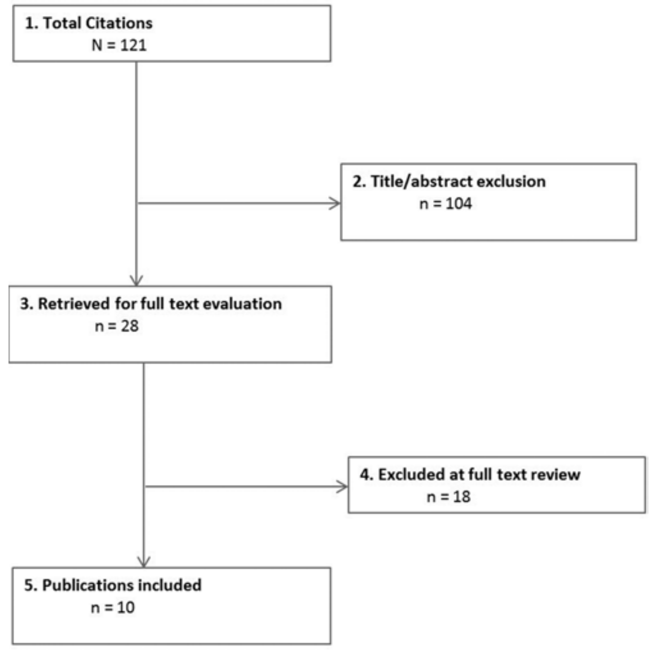

Fig. 1. Selection of studies of reliability and predictive validity of quantitative SCI outcome measures.

Results

Study Selection for Reliability and Predictive Validity Studies

The PubMed search for studies on reliability and predictive validity yielded 121 unique citations after initial exclusions by title. Regarding reliability, 5 studies were excluded at full text review: 3 did not include at least 50% of patients with SCI, 1 did not report reliability, and 1 involved blind persons without SCI. For predictive validity, 13 studies were excluded at full text review for 1 or more of the following reasons: timing of quantitative measure relative to outcome assessment not clear or not reported; no formal statistical evaluation of association between measure and outcome; or no reported effect size (Fig. 1).

Identification of Outcome Measures Used in Clinical Studies

Somatosensory and motor evoked potential measures appear to be the most commonly used among those selected, followed by manual muscle testing and the ASIA motor score (Fig. 2).

Reliability of Quantitative Measures

Studies designed to evaluate reliability in patients with SCI that met the inclusion criteria were found for the following measures: dermatomal SSEPs,[14] EPT,[13,14] TPT,[10,15] VPT,[10,15] EMG,[18] and the GRASSP.[12] No reliability studies in the SCI population were found for MEPs, autonomic, or contact heat-evoked potential tests. The overall quality of the studies was considered poor or very poor, and most studies were retrospective (Table 1). All populations appear to be convenience samples, and with the exception of 1 study,[12] failed to include a broad spectrum of persons with SCI to whom the test might apply, leading to possible selection bias and limiting the generalizability of the results. For some studies, the combination of test data for SCI patients with data from healthy controls precludes making conclusions about how the test will perform in the patients with SCI.

Table 2 summarizes basic characteristics and data from the included studies by quantitative measure. Table 1 summarizes critical appraisal elements for these studies.

Three studies included persons with incomplete SCI.[13,15,18] For EPT, ICCs for both interrater and intrarater reliability were ≥ 0.7 above the sensory level of SCI but were less reliable below the sensory level.[13] For TPT, the ICC for most dermatomes was < 0.7 for warm, cold, and cold pain thresholds.[15] For EMG, the ICC was consistently high for within-day tests.[18]

In persons with traumatic tetraplegia, intrarater reliability ICCs were high for unaffected dermatomal SSEP N1 latencies (0.97), but low for EPT testing (0.24) in 1 study.[14] Elapsed time between examinations ranged from 5 days to 1.3 years.

In a study of patients with SCI who had neuropathic pain,[10] an ICC of 0.9 was reported for VPT and 0.5 for TPT measures, but only 10 of the 22 patients with SCI were retested at 1–4 weeks.

The GRASSP reliability study[12] included a broader

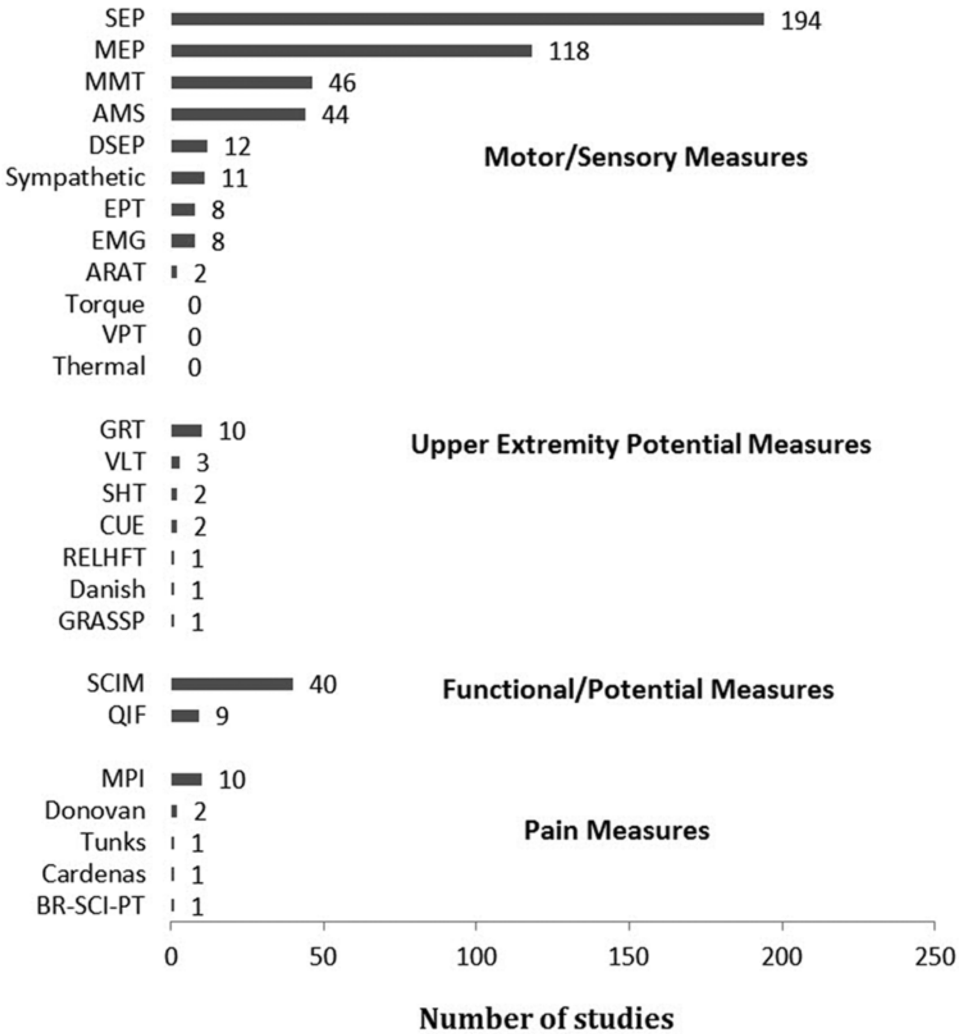

Fig. 2. Summary of selected common outcomes measures in patients with SCI. The number of studies found is given for each measure. Citations were retrieved from PubMed, and the outcome measures were identified in the titles and abstracts of articles. AMS = ASIA motor score; ARAT = action research arm testing; BR-SCI-PT = Bryce-Ragnarsson Pain Taxonomies; Cardenas = Cardenas Pain Classification; CUE = Capabilities of Upper Extremity instrument; Danish = Danish Tetraplegia Hand Measure; Donovan = Donovan Pain Classification Scheme; DSEP = dermatomal SSEP; GRT = Grasp and Release Test; MMT = Manual Muscle Test; MPI = Multidimensional Pain Inventory; QIF = Quadriplegia Index of Function; RELHFT = Rehabilitation Engineering Laboratory Hand Function Test; SCIM = Spinal Cord Independence Measure; SEP = somatosensory evoked potential; SHT = Sollerman Hand Function Test; Sympathetic = sympathetic reflex; Thermal = TPT; Torque = torque testing; VLT = Van Lieshout Test; Tunks = Tunks SCI Pain Classification.

range of persons with SCI, reporting high ICC values for all parameters evaluated. Interclass correlation coefficients for interrater and intrarater reliability of the GRASSP components ranged from 0.84 to 0.98. Construct and concurrent validity was established for the GRASSP measure in this study by comparisons to The International Standards for the Neurological Classification of Spinal Cord Injury scores and to Spinal Cord Independence Measures and Capabilities of Upper Extremity scores. No studies reported on the responsiveness of measures as defined for this review.

Predictive Validity

Limited evidence regarding the predictive validity of quantitative measures was found (Table 3). To assess predictive validity, a study needed to explicitly evaluate correlation (or similar measure of association providing an effect size) between baseline quantitative measures and specific functional or clinical outcomes measured at some later follow-up time. All studies appear to have used convenience samples; details of subject selection, number of individuals eligible but not enrolled, and description of enrollment procedures were not provided in any of the studies.

Three studies that evaluated a correlation between baseline or initial SSEPs and a specific clinical outcome at a later follow-up time were included.[4,7,17] Data from these studies suggest that SSEPs may be correlated with ambulatory capacity,[4] as well as the Barthel Index and motor index;[17] however, the correlation was limited for evaluation of bladder function recovery.[7]

TABLE 1: Critical appraisal of included reliability studies of SCI*

Methodological Principle	Kramer et al., 2010	King et al., 2009	Felix et al., 2009	Krassioukov et al., 1999	Lim et al., 2005	Kalsi-Ryan et al., 2012
broad spectrum of patients w/ expected condition	NA	NA	NA	NA	NA	yes
adequate description of methods for replication	yes	yes	yes	yes	yes	yes
blinded comparison of tests/interpretations	NA	yes	NA	NA	NA	NA
second test performed independently/timing appropriate	NA	NA	NA	NA	NA	NA
ICC or weighted κ ≥0.70 in ≥50 patients	NA	NA	NA	NA	yes	yes
level of evidence	IV	IV	IV	IV	IV	III

* NA = criterion not met or not reported by authors.

In 1 study of recovery of ankle dorsiflexion, correlation between MEP and the following outcomes was evaluated: maximal voluntary contraction, MMV, dexterity, gait, and ASIA scores. Authors accounted for change over time by calculating data as percentages (quotient 1-month result/6-month result) for MEP amplitude, MMV at 2.4 Hz, and gait speed, and as differences (difference of 1-month result subtracted from 6-month result) for ASIA motor score and Walking Index for SCI II score.[32]

Discussion

Evidence on the reliability of the quantitative measures selected for this review is limited, and the overall quality of existing studies is poor. From the included studies, it is not clear how the various tests may perform across a broad spectrum of persons with SCI and during the course of follow-up to include more acute and chronic phases and complete and incomplete SCI. All studies used convenience samples, leaving open the possibility of selection bias.

Even though a measure may be reproducible, it may not be valid. Classically, validity evaluates the extent to which an instrument measures what it is intended to measure and involves comparison with an appropriate "gold" standard that measures the "truth." No such studies were found based on the search conducted.

It is difficult to draw conclusions about the predictive validity as defined for this review for a number of reasons. To assess predictive validity, a study needs to explicitly evaluate correlation between baseline quantitative measures and specific functional outcomes measured at some later follow-up time. Few studies met these criteria. Included studies may have been subject to selection bias. Details regarding subject recruitment, inclusion/exclusion criteria, number of eligible patients who were not enrolled, and number lost to follow-up were not provided in the studies.

A number of studies examined the usefulness of electrophysiological measures and quantitative sensory testing. Many of these studies failed to meet our inclusion criteria because they were cross-sectional in nature; such studies cannot evaluate predictive associations. For example, the study by Curt and Dietz[5] examined the correlation between SSEP and hand function and concluded that median and ulnar SSEPs were predictive of hand function. However, this study was excluded because it did not describe the timing of measurements or report strength of correlation or effect size. It is possible that low SSEPs are evidence of poor function but do not predict changes in function. Another study of SSEPs and quantitative sensory testing[11] was excluded because it did not provide information on timing of measures or data on correlation between SSEPs and quantitative sensory testing.

The GRASSP is an SCI-specific quantitative measure of upper limb impairment. It was shown to have high intra- and interrater reliability. As discussed above, the GRASSP was shown to have construct and concurrent validity, but responsiveness and sensitivity to change have not been established as of this writing, although there are ongoing studies.

Significant challenges remain regarding use of routine electrophysiological tests such as MEPs and SSEPs in tracking recovery. Our review shows that dermatomal SSEPs are more reliable (ICC = 0.97) and are responsive to sensory recovery and possibly more useful than routine SSEPs. Dermatomal SSEPs allow for monitoring of neurophysiological changes in spinal segments.[14] Motor evoked potentials are the most commonly used quantitative tests of corticospinal tract function, but the correlation between amplitudes, latencies, and neurological recovery is poor.[6,8] More sophisticated electrophysiological tests such as short intracortical inhibition,[29] afferent regulation of evoked potentials,[28] and H-reflex modulation[3] have not been studied longitudinally. The BMCA was originally designed to identify and characterize residual supraspinal CNS influence on motor output following a severe SCI.[16,22–25] In the BMCA, composite motor unit activity recorded from multiple muscles is used to indicate the state of spinal motor excitability relative to a motor task requested or in response to maneuvers such as deep inspiratory breathing and Valsava maneuvers.[16,22–25] The BMCA measures the amplitude, duration, and time to peak of EMG activity of multiple muscles during standardized voluntary, passive, and reflexive maneuvers; however, it has not been validated in longitudinal studies. The BMCA has recently been modified into a new protocol, the functional neurophysiological assessment, which assesses neurological recovery in thoracic and cervical segments after SCI (see paper by Harkema in this issue).

Although a variety of tests are available for assessment of autonomic function in SCI (see review by Prévinaire et al.[27]), there were no studies that met our criteria for reliability and validity studies, or predictive validity

TABLE 2: Summary of studies evaluating reliability of selected quantitative motor, sensory, and autonomic measures of SCI*

Authors & Year†	Study Design	Pt Population	Uninjured Controls	Interobserver Reliability	Intraobserver Reliability or Test/Retest
dermatomal SSEP					
Kramer et al., 2010	retrospective cohort; electrode placements & recording configuration identical & performed by trained, qualified technician	n = 18 (traumatic tetraplegia); mean age 13.1 ± 16.4 yrs; 83% male; AIS Grade A (n = 6), Grade B (n = 4), Grade C (n = 3), Grade D (n = 5)	n = 5; mean age 35.7 ± 14.0 yrs; 80% male; exclusions: other neurological conditions	NR	ICC (tetraplegia): unaffected dermatomal SSEP N1 latencies 0.97 (p <0.001); ICC (controls): 0.9 (p <0.01)
EPT					
Kramer et al., 2010	retrospective cohort	n = 18 (traumatic tetraplegia); mean age 13.1 ± 16.4 yrs; 83% male; AIS Grade A (n = 6), Grade B (n = 4), Grade C (n = 3), Grade D (n = 5)	n = 5; mean age 35.7 ± 14.0 yrs; 80% male; exclusions: other neurological conditions	NR	ICC (tetraplegia): EPT 0.24 (p <0.05); ICC (controls): 0.5 (p <0.01)
King et al., 2009	prospective cohort; sensory levels determined for each side by same experienced examiner, raters blinded to results of each other	n = 12 (incomplete SCI); mean age 48 yrs; 83% male	n = 12; median age 33 yrs (range 22–65 yrs); 58% male; exclusions: clinical history of diabetes or any peripheral neuropathies	ICC Trial 1: sensory level above (0.77), at (0.74), below (0.56); Trial 2: sensory level above (0.91), at (0.79), below (0.52)	ICC sensory level above (0.80), at (0.56), below (0.65)
TPT					
Felix & Widerström-Noga, 2009	prospective cohort	n = 22 (SCI w/ chronic neuropathic pain); mean age 41.7 ± 15.5 yrs; 86% male	n = 10; mean age 30.4 ± 4.3 yrs; 60% male	NR	ICC (SCI): cold pain threshold 0.50, hot pain threshold 0.50; ICC (uninjured): cold pain threshold 0.49, hot pain threshold 0.68
Krassioukov et al., 1999	prospective cohort	n = 21 (incomplete SCI); mean age 38.9 ± 14.4 yrs; 71% male; AIS Grade B (n = 6), Grade C (n = 4), Grade D (n = 11)	n = 14; mean age 33.9 ± 9.6 yrs; 43% male; exclusions: peripheral nerve dysfunction, DM, or history of seizures or medical complications likely to impair ability to safely complete trial	NR	ICC in order of dermatome (rt S-1, L-4, L-5, lt S-1, L-4, L-5) for SCI: cold 0.55, 0.62, 0.81, 0.45, 0.68, & 0.79, warm 0.69, 0.25, 0.46, 0.56, 0.36, & 0.23, & cold pain 0.67, 0.65, 0.89, 0.75, 0.72, & 0.72; ICC (uninjured): cold 0.76, 0.79, 0.90, 0.75, 0.78, & 0.77, warm 0.73, 0.83, 0.84, 0.71, 0.75, & 0.36, & cold pain 0.95, 0.93, 0.91, 0.94, 0.91, & 0.93
VPT					
Felix & Widerström-Noga, 2009‡	prospective cohort	n = 22 (SCI w/ chronic neuropathic pain); mean age 41.7 ± 15.5 yrs; 86% male	n = 10; mean age 30.4 ± 4.3 yrs; 60% male	NR	ICC (SCI) 0.90, ICC (uninjured) 0.86

(continued)

TABLE 2: Summary of studies evaluating reliability of selected quantitative motor, sensory, and autonomic measures of SCI* ***(continued)***

Authors & Year†	Study Design	Pt Population	Uninjured Controls	Interobserver Reliability	Intraobserver Reliability or Test/Retest
VPT *(continued)*					
Krassioukov et al., 1999	prospective cohort	n = 21 (incomplete SCI); mean age 38.9 ± 14.4 yrs; 71% male; AIS Grade B (n = 6), Grade C (n = 4), Grade D (n = 11)	n = 14; mean age 33.9 ± 9.6 yrs; 43% male; exclusions: peripheral nerve dysfunction, DM, or history of seizures or medical complications likely to impair ability to safely complete trial	NR	ICC in order of dermatome (rt S-1, L-4, L-5, lt S-1, L-4, L-5) for SCI: NR, 0.76, 0.90, NR, 0.88, & 0.82; ICC for uninjured: NR, 0.96, 0.90, NR, 0.33, & 0.63
EMG					
Lim & Sherwood, 2005	prospective cohort	n = 69 (incomplete SCI); mean age 48.1 ± 4.6 yrs; 94% male; AIS Grade C (n = 34), Grade D (n = 35)	n = 15; mean age 36 ± 10 yrs; 73% male	NR	ICC (average of within-day tests for 10 motor tasks): AIS Grade C magnitude 0.92, similarity index 0.76; AIS Grade D magnitude 0.88, similarity index 0.87; total magnitude 0.93, similarity index 0.83; p <0.01 (magnitude versus similarity index)
GRASSP					
Kalsi-Ryan et al., 2011	prospective cohort	n = 72 (tetraplegia); mean age 39.7 ± 10.7 yrs; AIS Grade A (n = 28), Grade B (n = 18), Grade C (n = 12), Grade D (n = 14); exclusions: mod brain injury pts w/ neurological instability, or individuals w/ any pathology other than tetraplegia causing upper limb impairment		ICC: SWM rt 0.84, SWM lt 0.91, strength rt 0.95, strength lt 0.95, prehension ability rt 0.95, prehension ability lt 0.95, prehension performance rt 0.95, prehension performance lt 0.96; p <0.0001	ICC: SWM rt 0.95, SWM lt 0.86, strength rt 0.98, strength lt 0.98, prehension ability rt 0.98, prehension ability lt 0.98, prehension performance rt 0.93, prehension performance lt 0.96; p <0.0001

* DM = diabetes mellitus; mod = moderate; NR = not reported; Pt = patient; SWM = Semmes-Weinstein monofilaments.
† Responsiveness was not reported in any study.
‡ For sites with little to no vibratory sensation, rate was increased to avoid lengthy trials.

TABLE 3: Summary of included studies on predictive validity of quantitative measures of SCI*

Authors & Year	Study Design	Population	Measure & Outcome	Results (correlation, effect size)	Authors' Conclusions
SSEP					
Curt & Dietz, 1997	prospective cohort; 6-mo FU; convenience sample	n = 70 (acute SCI); mean age 42.6 ± 22.3 yrs (tetraplegia), 40.3 ± 17 yrs (paraplegia); 87% male (tetraplegia), 77% male (paraplegia); 100% FU (70/70)	measure: tibial SSEP, pudendal SSEP; outcome: ambulatory capacity; measurement timing: initial SSEP correlated w/ ambulatory capacity at ≥6 mos after injury	acute SCI (tetraplegic): tibial SSEP (r = 0.81, p <0.0001), pudendal SSEP (r = 0.92, p <0.0001); acute SCI (paraplegic): tibial SSEP (r = 0.72, p <0.0001), pudendal SSEP (r = 0.80, p <0.0001)	SSEP is related to ambulatory capacity in pts w/ acute SCI
Curt et al., 1997	prospective cohort; 6-mo FU; convenience sample	n = 70 (acute SCI); mean age 43 ± 22 yrs (tetraplegia), 40 ± 16 yrs (paraplegia); 87% male (tetraplegia), 77% (paraplegia); 100% FU (70/70)	measure: tibial SSEP, pudendal SSEP; outcome: autonomic nerve function, ambulatory capacity, bladder function, EUS function; measurement timing: initial SSEP correlated w/ bladder function at 6 mos	bladder function (tetraplegic): tibial SSEP (r = 0.50, p <0.003), pudendal SSEP (r = NR, p = 0.1); bladder function (paraplegic): tibial SSEP (r = NR, p <0.007), pudendal SSEP (r = NR, p <0.007)	limited relationship btwn SSEP & recovery of urodynamic bladder function; assessment of impairment of bladder function due to voluntary EUS function in pts w/ acute SCI & can indicate likelihood of EUS function recovery early after trauma
Li et al., 1990	prospective cohort; 6-mo FU; convenience sample	n = 36 (cervical SCI); mean age NR; % male NR; 100% FU (36/36)	measure: ulnar SSEP, pst tibial SSEP; outcome: Barthel Index, motor index; measurement timing: 1st SSEP w/in 2 wks of injury, Barthel Index & motor index measured 6 mos after injury	Barthel Index (pst tibial): multivariate regression (−R2 = 0.75, p <0.0001); motor index (pst tibial): multivariate regression (−R2 = NR); less mean improvement if SSEP of 1 vs >1; p = 0.0009	SSEP is a good prognostic indicator of functional recovery
MEP					
Wirth et al., 2008	prospective cohort; 6-mo FU; convenience sample	n = 12 (incomplete ASIA Grade C & D SCI) mean age 53.7 ± 18.5 yrs; 50% male; % FU NR, no. of eligible but not enrolled NR	measure: MEP (baseline data only available for 8/12 pts; outcome: MVC, MMV, dexterity, gait, ASIA scale; measurement timing: pts assessed at 1, 3, & 6 mos after injury†	MMV: r = 0.67 (1–3 mos), p = 0.03, r = 0.69 (1–6 mos), p = 0.02; MVC: r = 0.50 (1–3 mos), p = 0.14, r = 0.58 (1–6 mos), p = 0.06; ASIA: r = 0.08 (1–3 mos), p = 0.83, r = 0.29 (1–6 mos), p = 0.39; gait speed: r = 0.24 (1–3 mos), p = 0.48, r = 0.20 (1–6 mos), p = 0.53; walking aids: r = 0.17 (1–3 mos), p = 0.65, r = 0.26 (1–6 mos), p = 0.44	purpose: study motor recovery of ankle dorsiflexion; no obvious relationship btwn corticospinal tract function (torque-controlled MEP) & gross muscle strength & gait capacity

* EUS = external urethral sphincter; FU = follow-up; MVC = maximal voluntary contraction; pst = posterior.

† Data calculated and displayed as percentages (quotient 1-month result/6-month result) for MEP amplitude, MMV at 2.4 Hz, and gait speed, and as differences (difference of 1-month result subtracted from 6-month result) for ASIA motor score and Walking Index for SCI II score.

correlating autonomic testing results with clinical function, such as development of autonomic dysreflexia. Development of validated autonomic tests is urgently needed because these are not currently evaluated as part of the standard AIS clinical evaluation.[1]

Our results reveal a need for more studies of reliability and validity of non–SCI-specific quantitative sensory measures such as the THT and VBT. For EPT, ICCs for both interrater and intrarater reliability were ≥ 0.7 above the sensory level of SCI, but were less reliable below the sensory level. One recent study reported an ICC of 0.24 on normal dermatomes from patients with SCI, which is much lower than what has been reported in previous studies.[13,14] With the exception of this study, EPT appears to be a reliable quantitative measure of sensory function.[9]

Conclusions

Evidence on the reliability and validity of the quantitative measures selected for this review is limited, and the overall quality of existing studies is poor, with small sample sizes and potential for selection bias. In summary, there is some evidence that the EPT, dermatomal SSEPs, and the GRASSP may be reliable for use in longitudinal studies, but further studies in larger samples are needed. These measures are recommended as adjuncts to the AIS.[9] Future studies must also address the concept of minimally clinically important difference of these recommended measures, and unestablished quantitative measures need to be further evaluated in prospective longitudinal studies.[1] It is also important to establish the predictive value of these tests by comparing them to specific future AIS and functional outcomes.

Disclosure

Analytic support for this work was provided by Spectrum Research, Inc., with funding from AOSpine North America.

Author contributions to the study and manuscript preparation include the following. Conception and design: Harkama, Ellaway. Analysis and interpretation of data: Boakye, Skelly. Drafting the article: Harkama, Ellaway, Skelly. Critically revising the article: Harkama, Ellaway, Skelly. Reviewed submitted version of manuscript: Harkama, Ellaway, Skelly.

Acknowledgment

The authors wish to thank Daniel Hadidi for his assistance with the literature search and data abstraction.

References

1. Alexander MS, Anderson KD, Biering-Sorensen F, Blight AR, Brannon R, Bryce TN, et al: Outcome measures in spinal cord injury: recent assessments and recommendations for future directions. **Spinal Cord 47:**582–591, 2009
2. American Spinal Injury Association, International Spinal Cord Society: **International Standards for Neurological Classification of Spinal Cord Injury, ed 6.** Chicago: American Spinal Injury Association, 2006
3. Benito Penalva J, Opisso E, Medina J, Corrons M, Kumru H, Vidal J, et al: H reflex modulation by transcranial magnetic stimulation in spinal cord injury subjects after gait training with electromechanical systems. **Spinal Cord 48:**400–406, 2010
4. Curt A, Dietz V: Ambulatory capacity in spinal cord injury: significance of somatosensory evoked potentials and ASIA protocol in predicting outcome. **Arch Phys Med Rehabil 78:** 39–43, 1997
5. Curt A, Dietz V: Traumatic cervical spinal cord injury: relation between somatosensory evoked potentials, neurological deficit, and hand function. **Arch Phys Med Rehabil 77:**48–53, 1996
6. Curt A, Keck ME, Dietz V: Functional outcome following spinal cord injury: significance of motor-evoked potentials and ASIA scores. **Arch Phys Med Rehabil 79:**81–86, 1998
7. Curt A, Rodic B, Schurch B, Dietz V: Recovery of bladder function in patients with acute spinal cord injury: significance of ASIA scores and somatosensory evoked potentials. **Spinal Cord 35:**368–373, 1997
8. Curt A, Van Hedel HJ, Klaus D, Dietz V: Recovery from a spinal cord injury: significance of compensation, neural plasticity, and repair. **J Neurotrauma 25:**677–685, 2008
9. Ellaway PH, Kuppuswamy A, Balasubramaniam AV, Maksimovic R, Gall A, Craggs MD, et al: Development of quantitative and sensitive assessments of physiological and functional outcome during recovery from spinal cord injury: a clinical initiative. **Brain Res Bull 84:**343–357, 2011
10. Felix ER, Widerström-Noga EG: Reliability and validity of quantitative sensory testing in persons with spinal cord injury and neuropathic pain. **J Rehabil Res Dev 46:**69–83, 2009
11. Hayes KC, Wolfe DL, Hsieh JT, Potter PJ, Krassioukov A, Durham CE: Clinical and electrophysiologic correlates of quantitative sensory testing in patients with incomplete spinal cord injury. **Arch Phys Med Rehabil 83:**1612–1619, 2002
12. Kalsi-Ryan S, Beaton D, Curt A, Duff S, Popovic MR, Rudhe C, et al: The Graded Redefined Assessment of Strength Sensibility and Prehension (GRASSP): reliability and validity. **J Neurotrauma 29:**905–914, 2012
13. King NK, Savic G, Frankel H, Jamous A, Ellaway PH: Reliability of cutaneous electrical perceptual threshold in the assessment of sensory perception in patients with spinal cord injury. **J Neurotrauma 26:**1061–1068, 2009
14. Kramer JK, Taylor P, Steeves JD, Curt A: Dermatomal somatosensory evoked potentials and electrical perception thresholds during recovery from cervical spinal cord injury. **Neurorehabil Neural Repair 24:**309–317, 2010
15. Krassioukov A, Wolfe DL, Hsieh JT, Hayes KC, Durham CE: Quantitative sensory testing in patients with incomplete spinal cord injury. **Arch Phys Med Rehabil 80:**1258–1263, 1999
16. Lee DC, Lim HK, McKay WB, Priebe MM, Holmes SA, Sherwood AM: Toward an objective interpretation of surface EMG patterns: a voluntary response index (VRI). **J Electromyogr Kinesiol 14:**379–388, 2004
17. Li C, Houlden DA, Rowed DW: Somatosensory evoked potentials and neurological grades as predictors of outcome in acute spinal cord injury. **J Neurosurg 72:**600–609, 1990
18. Lim HK, Sherwood AM: Reliability of surface electromyographic measurements from subjects with spinal cord injury during voluntary motor tasks. **J Rehabil Res Dev 42:**413–422, 2005
19. Marino RJ, Barros T, Biering-Sorensen F, Burns SP, Donovan WH, Graves DE, et al: International standards for neurological classification of spinal cord injury. **J Spinal Cord Med 26 (Suppl 1):**S50–S56, 2003
20. Marino RJ, Graves DE: Metric properties of the ASIA motor score: subscales improve correlation with functional activities. **Arch Phys Med Rehabil 85:**1804–1810, 2004
21. Marino RJ, Jones L, Kirshblum S, Tal J, Dasgupta A: Reliability and repeatability of the motor and sensory examination of the international standards for neurological classification of spinal cord injury. **J Spinal Cord Med 31:**166–170, 2008
22. McKay WB, Lee DC, Lim HK, Holmes SA, Sherwood AM: Neurophysiological examination of the corticospinal system

and voluntary motor control in motor-incomplete human spinal cord injury. **Exp Brain Res 163:**379–387, 2005
23. McKay WB, Lim HK, Priebe MM, Stokic DS, Sherwood AM: Clinical neurophysiological assessment of residual motor control in post-spinal cord injury paralysis. **Neurorehabil Neural Repair 18:**144–153, 2004
24. McKay WB, Ovechkin AV, Vitaz TW, Terson de Paleville DG, Harkema SJ: Neurophysiological characterization of motor recovery in acute spinal cord injury. **Spinal Cord 49:**421–429, 2011
25. McKay WB, Stokic DS, Dimitrijevic MR: Assessment of corticospinal function in spinal cord injury using transcranial motor cortex stimulation: a review. **J Neurotrauma 14:**539–548, 1997
26. Norvell D, Dettori J, Suk M: What makes a quality outcomes instrument?, in Suk M, Hanson BP, Norvell DC, et al (eds): **AO Handbook of Musculoskeletal Outcomes Measures and Instruments.** Stuttgart: Thieme, 2005, pp 7–24
27. Prévinaire JG, Mathias CJ, El Masri W, Soler JM, Leclercq V, Denys P: The isolated sympathetic spinal cord: Cardiovascular and sudomotor assessment in spinal cord injury patients: a literature survey. **Ann Phys Rehabil Med 53:**520–532, 2010
28. Roy FD, Yang JF, Gorassini MA: Afferent regulation of leg motor cortex excitability after incomplete spinal cord injury. **J Neurophysiol 103:**2222–2233, 2010
29. Roy FD, Zewdie ET, Gorassini MA: Short-interval intracortical inhibition with incomplete spinal cord injury. **Clin Neurophysiol 122:**1387–1395, 2011
30. Steeves JD, Lammertse D, Curt A, Fawcett JW, Tuszynski MH, Ditunno JF, et al: Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: clinical trial outcome measures. **Spinal Cord 45:**206–221, 2007
31. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al: Quality criteria were proposed for measurement properties of health status questionnaires. **J Clin Epidemiol 60:**34–42, 2007
32. Wirth B, Van Hedel HJ, Curt A: Changes in corticospinal function and ankle motor control during recovery from incomplete spinal cord injury. **J Neurotrauma 25:**467–478, 2008

Manuscript submitted January 26, 2012.
Accepted May 17, 2012.
Please include this information when citing this paper: DOI: 10.3171/2012.5.AOSPINE1296.
Address correspondence to: Maxwell Boakye, M.D., M.P.H., Outcomes and Translational Research Center for Advanced Neurosurgery, University of Louisville, 220 Abraham Flexner Way, 1105, Louisville, Kentucky 40202. email: max.boakye@louisville.edu.