



Considerations and recommendations for selection and utilization of upper extremity clinical outcome assessments in human spinal cord injury trials

Linda A. T. Jones¹ · Anne Bryden² · Tracey L. Wheeler¹ · Keith E. Tansey^{3,4,5} · Kim D. Anderson⁶ · Michael S. Beattie⁷ · Andrew Blight⁸ · Armin Curt^{9,10} · Edelle Field-Fote^{11,12,13} · James D. Guest⁶ · Jane Hsieh^{14,15} · Lyn B. Jakeman¹⁶ · Sukhvinder Kalsi-Ryan^{17,18} · Laura Krisa¹⁹ · Daniel P. Lammertse^{20,21} · Benjamin Leiby¹⁹ · Ralph Marino¹⁹ · Jan M. Schwab²² · Giorgio Scivoletto²³ · David S. Tulskey²⁴ · Ed Wirth²⁵ · José Zariffa^{17,18} · Naomi Kleitman¹ · Mary Jane Mulcahey¹⁹ · John D. Steeves²⁶

Received: 31 May 2017 / Revised: 16 August 2017 / Accepted: 18 August 2017 / Published online: 28 December 2017
© The Author(s) 2018. This article is published with open access

Abstract

Study design This is a focused review article.

Objectives This review presents important features of clinical outcomes assessments (COAs) in human spinal cord injury research. Considerations for COAs by trial phase and International Classification of Functioning, Disability and Health are presented as well as strengths and recommendations for upper extremity COAs for research. Clinical trial tools and designs to address recruitment challenges are identified.

Methods The methods include a summary of topics discussed during a two-day workshop, conceptual discussion of upper extremity COAs and additional focused literature review.

Results COAs must be appropriate to trial phase and particularly in mid-late-phase trials, should reflect recovery vs. compensation, as well as being clinically meaningful. The impact and extent of upper vs. lower motoneuron disease should be considered, as this may affect how an individual may respond to a given therapeutic. For trials with broad inclusion criteria, the content of COAs should cover all severities and levels of SCI. Specific measures to assess upper extremity function as well as more comprehensive COAs are under development. In addition to appropriate use of COAs, methods to increase recruitment, such as adaptive trial designs and prognostic modeling to prospectively stratify heterogeneous populations into appropriate cohorts should be considered.

Conclusions With an increasing number of clinical trials focusing on improving upper extremity function, it is essential to consider a range of factors when choosing a COA.

Sponsors Craig H. Neilsen Foundation, Spinal Cord Outcomes Partnership Endeavor.

Introduction

As of 2016, more than half of all spinal cord injuries (SCI), ~58%, affect the cervical spine resulting in tetraplegia and less than 1% of people with SCI experience complete

recovery by hospital discharge [1]. In persons with tetraplegia, upper extremity (UE) function is affected. Initial human trials of high-risk drug or cell-based interventions are typically tested in individuals with thoracic SCI followed by cervical SCI trials. Trials in thoracic SCI are conducted to establish safety, as the potential loss of function in thoracic SCI has less impact than loss of cervical segments. However, evidence of efficacy is difficult to assess in the thoracic spine as the standard SCI motor exam (International Standards for Neurological Classification—ISNCSCI) defers to segmental sensory assessments, as accurate testing using the ISNCSCI motor scoring system is only available for muscles of the extremities [2]. In contrast, muscles in the UE innervated by cervical segments are

Naomi Kleitman, Mary Jane Mulcahey and John D. Steeves contributed equally to this work

✉ Linda A. T. Jones
linda@chnfoundation.org

Extended author information available on the last page of the article

readily testable, thus efficacy assessed by motor function is easier to detect.

Restoration of UE function is consistently identified as one of the top priorities for individuals living with tetraplegia [3]. Neurological improvement within cervical cord segments resulting in functional change has the potential to significantly impact quality of life, as regaining function of even one cervical level can make the difference between independence and dependence for many ADLs such as transfers and self-care [4]. Loss of UE function impacts other activities, which are also top priorities for individuals with SCI, such as ability to manage one's own bowel and bladder program. Due to the readily testable motor function in the UE and the importance of UE function, pivotal (confirmatory) SCI trials (a study intended to provide definitive evidence of efficacy prior to drug marketing approval) will likely be conducted in persons living with cervical SCI. In fact, based on data extracted from ClinicalTrials.gov between 2012 and 2016, there was an increase of 22% (from 13 to 35%) of drug, cell, or surgical interventions using UE clinical outcome assessments (COAs).

The types of tasks performed with the arm and hand are complex and diverse. Measures of sensation and strength (such as those assessed by ISNCSCI) are important, but do not capture this diversity or directly measure clinically meaningful function. UE tasks are often asymmetrical, temporally and spatially varied, with many functional goals and degrees of freedom. For example, multiple grasp patterns are used during ADLs, ranging from power grasp, where large forces are required to lift a glass or use a knife, to prehension tasks, which require precision to pick up a pin or hold a pencil. In order to capture this diversity, UE COAs may also be complex. Simple measures of time to complete a given task, such as those used for speed of gait in the lower extremity, may provide valuable information, but speed of performance is only one aspect of function. Thus, more comprehensive assessments are needed to characterize UE function.

As a direct result of the increase in trials that include UE function as an end point, and the lack of consensus on which UE COAs are best suited for clinical trials, a workshop was convened in January 2015 by the Craig H. Neilsen Foundation (www.chnfoundation.org) in association with the Spinal Cord Outcomes Partnership Endeavor (www.scope-sci.org). The goal of this workshop was to assemble leading experts (predominantly from North America—Appendix 1) in SCI, UE COAs, and SCI clinical trials, to identify important concepts in COA and provide guidance on how emerging and existing measures may be used to assess experimental therapies in SCI. An additional goal was to examine methods to facilitate recruitment strategies for future trials. The focus was on UE COAs assessing

impairment and/or function, in individuals with cervical complete and incomplete SCI in Phase II–III trials, where the therapeutic intervention was directed to alter a central nervous system target. This manuscript highlights key recommendations based on the expertise of the workshop participants, the concepts that led to those recommendations, and additional focused literature review of themes central to the workshop topics.


Key concepts

Four major areas are discussed herein: (1) Define important considerations of COAs; (2) Compare features of select UE COAs; (3) Discuss broad functional assessment scales that may allow for more inclusive trial designs and; (4) Define remaining gaps and options to address recruitment.

Important features of clinical outcome assessments

Beyond reliability, validity, and responsiveness of a COA tool, there are numerous factors that are important to consider in the choice of specific COAs for a given trial. Considerations include trial phase and measurement target, whether the COA is able to distinguish between recovery and compensation, the impact of lower motor neuron injury, clinical significance of the observed changes, gaps in content, and floor and ceiling effects. Although the topics discussed focus on different aspects of COAs for clinical trials, they are important to consider in aggregate. The choice of primary outcomes and end points should be driven by the therapeutic target of the intervention and clinical trial phase. Table 1 highlights clinical trial phases with associated relevant measurement domains. In addition to referencing traditional drug trial phases, we use the terms “early, mid, and late” phase to encompass trials that may not follow the drug and biologics path, but may focus on rehabilitative interventions or devices. Measurement domains are based on the International Classification of Functioning, Disability and Health (ICF) terminology [5]. The ICF provides a comprehensive framework for health-related function with interactions between the health condition (disorder or disease), body structures and function (impairment), activities (function) and participation. As COAs are considered along the ICF spectrum toward activities and participation, the number of factors that can influence outcomes increases. At the body function and structure level (impairment), anatomical and physiological variables are assessed and influenced by central nervous system integrity (sensory/motor neural circuits). In contrast, at the participation levels, more factors play a role including personal and social characteristics (e.g., choice, motivation, and social background), as

Table 1 Considerations for UE neurological and functional activity outcome measurement tools in biological and pharmacological clinical trials according to trial phase and ICF domains

	Clinical trial phase	Phase I (first in man to evaluate safety, tolerability and feasibility)	Phase II (safety and efficacy, administration, dose, timing – may evaluate evidence of biological activity and assess candidate measures for use in Phase III)	Phase III (pivotal efficacy, safety)
Increasing number of independent variables influencing outcomes in each domain 		Early phase Pharmacokinetics for safety ISNCSCI (UEMS) for safety (assess neurological deterioration)	Mid phase Electrophysiology, imaging to assess evidence of biological activity ISNCSCI (UEMS) to assess neurological improvement and deterioration	Late phase ISNCSCI (UEMS) primarily to assess efficacy
Examples of outcome measures corresponding to ICF domains (Independent of expected outcome levels)	Body function & structure <i>Change measured:</i> Improvement (or impairment) at the neurological and physiological level. <i>Factors influencing outcomes:</i> CNS integrity, neural circuits (sensory/motor)	NA	Activity measures may be secondary or exploratory to assess feasibility in later phase trials SCI specific UE measures: GRASSP ¹ CUE-T ²	Activity measures may be primary, secondary or exploratory SCI specific UE measures: GRASSP ¹ CUE-T ²
	Activities <i>Change measured:</i> Task execution/functional capacity (or limitation) <i>Factors influencing outcomes:</i> CNS integrity, neural circuits (sensory/motor), adaptive behaviors, rehabilitation, psychology/motivation	NA	Non-UE specific measures: SCIM – self-care SCI-FI ² SCAR ^{2,3}	Non-UE specific measures: SCIM – self-care SCI-FI ² SCAR ^{2,3}
	Participation <i>Change measured:</i> Involvement in life situations, quality of life and participation (or restriction) <i>Factors influencing outcomes:</i> CNS integrity, neural circuits (sensory/motor), adaptive behaviors, rehabilitation, psychology/motivation, community/family support, finances, work/school, barriers in the structural and societal environment	NA	NA	Optional

^aThese COAs measure components of body structure and function and activity

^bThese recently developed COAs may provide additional insight and have been attracting attention in the field of SCI but are not fully explored and require further application/validation in clinical trials

^cSCAR combines selected ordinal aspects of ISNCSCI and SCIM III and transforms those scores into a linear continuous scales, thus a separate assessment is not required

well as environmental factors (e.g., physical, social, and attitudinal conditions). Typically, early phase trial outcomes focus on safety at the base level of the ICF domains (neurological impairment), whereas later stage trials focus on changes in function needed to accomplish ADLs, and possibly participation within the community. However, it is important to establish the link between impairment and function in mid and late-phase trials. Assessing neurological impairment alone is a proxy measure and provides no direct information about the impact of change on functional ability and thus clinical meaningfulness. Conversely, assessments that focus solely on the ability to execute select tasks may capture behavioral compensation as opposed to actual recovery of functional neural circuitry.

Drug and biological interventions typically target neurological recovery vs. compensation, thus it is important to define these terms and choose COAs that reflect recovery. Levin et al. [6] considered recovery vs. compensation at different structural levels in relation to stroke. Although the terminology used by these authors was intended to follow the ICF, they used these terms differently than they are typically used in SCI, therefore we have simplified the terminology herein. At the physiological level, recovery may occur through neuroplasticity, or repair of injured neural circuits, whereas compensation may occur by activation of alternative pathways. At the level of task execution, recovery is attained by the completion of a movement in the same manner in which it was accomplished prior to injury, whereas compensation reflects performance of a previous task with a novel adaptive behavioral strategy. The movement may be slower, or coordinated by different patterns of muscle activation such as a tenodesis grasp pattern (passive hand grasp and release mechanism, occurring by active wrist extension or passive wrist flexion). At the level of task outcome, true functional recovery is demonstrated with concomitant improvement in neurological restitution and correlated with the normal activation of the same limb or muscles. Conversely, compensation is achieved using different joints or muscles; for example, opening a bag of chips with one hand and the mouth instead of two hands. For mid-late-phase drug or biologics studies COAs should, to some degree, reflect attainment of pre-injury motor behaviors at the level of task execution and outcome at the activity level in the ICF classification.

Although the impact of recovery and compensation can be discussed separately, they do not occur in isolation. They may occur together and influence one another, both at physiological and functional levels. However, the distinction between recovery and behavioral compensation is important in the context of clinical trials as behavioral compensation can result from rehabilitation or other strategies that do not reflect the therapeutic effect being tested. If a clinical trial is using functional electrical stimulation to

increase strength and coordination of movement to achieve a more normal grasp pattern, simply assessing time to accomplish the task does not necessarily reflect the target of the intervention. One may use a completely different strategy (function) resulting in increased speed, without commensurate changes at the underlying neurological level (impairment), in this case, strength and coordination.

An example of a COA that focuses on function, but does not necessarily distinguish between behavioral compensation and recovery, is the spinal cord independence measure (SCIM). SCIM III measures aspects of self-care, respiration, and mobility at the activity level, but does not measure changes in body structure (neurological impairment) resulting in improved activity. Curt et al. [7] examined the contributions of compensation, neural plasticity, and repair based on neurological assessment of motor function (ISNCSCI), functional assessments (SCIM and Walking Index for Spinal Cord Injury II), and electrophysiological recordings, respectively. In individuals with a motor complete injury, a change in motor score was not correlated with a change in SCIM. In individuals with motor incomplete injury, there was a moderate relationship. Thus, in this study, SCIM did not differentiate recovery from behavioral compensation. In contrast, a study by Rudhe et al. [8] found strong correlations between SCIM III self-care and UE motor scores. Other tools, such as the neuromuscular recovery scale (NRS) intend to assess an individual's capacity without compensation by grading movements in relation to pre-injury movement patterns [9]. The NRS is still in the development stage, with the recent addition of three UE tasks, thus further validation is required to confirm that it is capturing recovery in the absence of compensation [10]. Although it is not possible to clearly separate contributions of recovery and compensation with most COAs, it is important to correlate the contribution of neurological improvement to a change in functional recovery in the clinical trial setting.

When interpreting the outcomes of a given intervention, one needs to consider the contribution of upper motoneuron (UMN) and lower motoneuron (LMN) lesions. The extent of LMN damage may vary by level and severity of injury and across individuals with similar levels of injury. Mulcahey et al. [11] found that in children and youths with C1–C4 injuries, between 2% (triceps) and 91% (biceps complex) of tested muscles did not respond to stimulation using surface electrical stimulation, indicating significant LMN damage. Curt et al. [12] found a high rate of significant denervation in median and ulnar nerves in individuals with lower (C6–T1) cervical vs. mid (C4,5) injuries. These two studies illustrate the prevalence of LMN damage, both at the site of injury and one-to-two levels below the injury. LMN changes may occur from actual traumatic loss of motoneurons, severe proximal ventral root injury, or

changes in motor unit properties as a result of denervation (motoneuron sprouting with enlarged motor units). Lower motoneuron dysfunction is also associated with other segmental neuronal loss, including interneurons that coordinate activity across motoneuron pools and damage to dorsal roots and/or sensory processing in the dorsal horn. The functional impact of LMN loss varies based on the extent of loss at single and multiple segments. LMN injury is important to assess as this may have a significant impact on response rate for therapeutics targeting central nervous system (UMN) dysfunction. Clinically, muscular atrophy and associated hyporeflexia are signs of LMN injury, whereas spasticity and associated hyperreflexia are signs of UMN injury. However, they can often be seen in combination after cervical SCI. Surface electrical stimulation may be used as an initial assessment; peripheral nerve stimulation of muscles impacted by LMN damage commonly requires more current or may not be excitable. Additional electro-diagnostic testing (F waves, H reflexes) may follow as needed to provide additional information about LMN patency [13]. Results of these tests should be considered in the overall interpretation of results when looking at responders vs. non-responders to a given therapeutic approach.

The impact of an intervention in any late-stage clinical trial should be clinically meaningful (how a person feels, functions, or survives) to persons with the condition under study. In successful interventional SCI trials, clinical meaningfulness of an observed functional effect must be demonstrated. One way that clinical meaningfulness is determined is by establishing the minimal clinically important difference (MCID). While definitions may vary, MCID has traditionally been defined as “the smallest difference in score, within the domain of interest, which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive costs, a change in the patient’s management” [14]. As MCID may be expressed as either an improvement or a deterioration, Hart et al. [15] further separates MCID to reflect the “minimal clinically important improvement” (MCII). Both MCID and MCII may be impacted by baseline assessment scores, which vary widely by level and severity of SCI, as well as demographic characteristics such as age and possibly gender. For example, for individuals with orthopedic knee impairments, Wang et al. [16] assessed MCII for the lower extremity functional scale, a computer adaptive patient-reported outcome (PRO). They found that younger, male participants and those with more acute symptoms required a greater change in performance for a measure to be judged clinically meaningful when compared to older females or less severely injured subjects. Therefore, it is difficult to establish a universally accepted MCII or MCID threshold for any given COA. As discussed by Wu et al.

[17], there are numerous methods to establish MCID but ideally MCID is established from utilization of a COA in a clinical trial setting. As few late-phase SCI trials have been completed, there are very few established MCIDs for SCI measures. In the absence of an established MCID, the MCID for a given COA can also be assessed in mid-phase feasibility studies, prior to use in a pivotal trial.

Beyond measuring meaningful change, COAs should also be sensitive to change in all severities of injury, without gaps in content, and without floor or ceiling effects. Gaps in content occur when discrete but clinically important skills are not captured in a COA. Gaps may be present in the upper and/or lower end of a COA, resulting in floor and ceiling effects, respectively, but may occur anywhere along a scale, resulting in the inability to track important but discrete aspects of function. Floor effects are present when the lower levels of an outcome measure may be too difficult, thus scores are clustered at the lower range and do not differentiate lower levels of performance. Ceiling effects occur when the upper levels of a COA can no longer distinguish or track additional performance improvement. Many of the existing ordinal COAs such as ISNCSCI, and SCIM III may not adequately track the extent of recovery in less severe cases of SCI, for example, individuals with an American Spinal Injury Association (ASIA) impairment scale D or “AIS D” classification [2]. People with injuries classified as AIS D can perform many of the existing items within the available ordinal COAs over the first few months after injury while continued improvements may not be captured by these COAs. Because of their extensive functional recovery, and the safety concerns associated with novel drugs or cell transplants, AIS D participants are rarely recruited to drug and cell-based SCI studies. Until validation of existing tools for individuals with AIS D occurs and the tools are found to be sensitive to small change in function, it is recommended that recruitment of these individuals in clinical trials be undertaken cautiously. In order to expand inclusion criteria to participants with less severe SCI, assessment tools must be expanded to be sensitive to improvement at high levels of function.

In summary, when planning a clinical trial, investigators should ask the following questions when choosing a COA:

1. Is the COA appropriate for the trial phase? Mid- and late-phase trials should include measures of both impairment and function.
2. What is the impact of UMN vs. LMN dysfunction on the interpretation of the COA and response to a given therapeutic?
3. If the intended population encompasses individuals with injuries ranging from thoracolumbar to cervical injuries, are the COAs sensitive to change in all severities and levels of SCI?

4. Are the outcomes clinically meaningful? Statistically significant change that is not clinically meaningful will not support regulatory registration of an investigational product.

Features of select emerging UE clinical outcome assessments

A number of established instruments and others in development that are available for UE assessment, such as those recently included in the National Institutes of Neurological Disorders and Stroke Common Data Elements for SCI research, were discussed at the workshop [18]. These tools were developed for different uses, including classification, monitoring of clinical improvement, and utilization in clinical trials. Additionally, they span the range from assessment of impairments, such as strength and sensation, to capacity and performance. Measures of capacity are commonly observed by a health care team in the hospital or clinical setting. Measures of performance may also be observed by a member of the health care team in the home environment, but are commonly assessed by the patient, as a PRO. While performance is more important to an individual, measures of performance collected by PRO do not differentiate recovery from compensation and, therefore, are likely to be used as secondary or exploratory clinical trial end points (vs. primary). Measures of capacity are more likely to be chosen as primary outcome measures, as it is possible (with some COAs) to consider the contribution of compensatory techniques. Table 2 summarizes strengths and recommendations for use of selected COAs to assess UE impairment and/or function in clinical trials.

Two COAs that are specific to UE function after SCI are under development to provide improved sensitivity to change in the cervical SCI population, relative to more “global” activity measures such as SCIM III. The graded refined assessment of strength, sensibility and prehension (GRASSP) and capabilities of upper extremities test (CUE-T) both measure capacity of UE function, but use different approaches [19, 20]. The GRASSP is multidimensional, with separate scales for sensation, motor, and prehensile hand abilities. The CUE-T is unidimensional, with subscales for right/left arm and right/left hand. The CUE-T contains elements that assess proximal arm function in addition to hand function, while the GRASSP focuses only on hand function. Both instruments use standardized objects and methods to assess the typical grasp patterns (e.g., cylindrical grasp, lateral key grip, and tip-to-tip pinch). The GRASSP allows for the use of compensatory strategies for prehension but gives a lower score when compensation is identified. The CUE-T does not permit alternative grasp patterns for a particular task. While CUE-T is at an earlier

stage of development, establishment of psychometric properties for both measures is underway, with more psychometric data currently available on the GRASSP [19–25]. Both measures are being used as secondary COAs in clinical trials.

Emerging functional assessment scales to support inclusive trials

For clinical trials involving a broad range of subjects, e.g., those with different levels and severities of SCI, assessment of upper and lower extremity function may be indicated, thus a more “global” COA may be useful. Although there are several COAs that have been developed specifically for SCI, two new tools have broad application across different types of SCI: The spinal cord injury-functional index (SCI-FI) and spinal cord ability ruler (SCAR). Although neither tool is UE-specific, both include elements of UE function [26–28]. Both have a similar underlying approach to scaling, which is to assess specific neurological (SCAR) and functional activities (SCAR and SCI-FI) along a continuum of increasing difficulty.

The SCI-FI is a PRO-based tool that is available as a computer adaptive test (CAT) and short form to capture individual performance of physical function [26, 27, 29]. CAT first presents an item in the middle of the range of difficulty (not too hard, not too easy) and uses the response to estimate ability and determine the next appropriate item to present. This process continues until a predefined precision is met or a predefined stop-rule (e.g., stop at 5-items, or 10-items) is reached. The SCI-FI was developed as a PRO with the goal of including comprehensive content to cover all aspects of physical function in five-item banks (basic mobility, ambulation, wheelchair mobility, self-care, and fine motor function) while incorporating the perspective of individuals living with SCI across the full range of SCI severity and level of injury [26]. Self-care and fine motor item banks are the most relevant to the assessment of UE function. The tool was developed using item response theory, which is a conceptual model for design, analysis, and scoring of instruments that is intended to measure abilities or traits. The statistical procedure calibrates (or aligns) the items in the bank according to difficulty and discriminability parameters.

Precision is achieved by using item response theory calibrations, which, when administered by CAT, minimizes the likelihood of administering irrelevant items. For example, if an individual is not able to write with a pen or a pencil, items related to more difficult tasks (e.g., picking up a coin from the ground) would not be administered. CAT administration is tailored to subject ability and provides an estimation of function, achieving precision with reduced

Table 2 Strengths and recommendations for use of select UE clinical outcome assessments

	Primary construct	Strengths	Recommendations for use in clinical trials
Upper extremity motor scores (UEMS) derived from ISNCSCI	Upper extremity strength	<p>SCI specific</p> <p>Web based and in person training available through the American Spinal Injury Association</p> <p>Can be assessed early post injury in all severities of SCI</p> <p>Psychometrics established in pediatric and adult populations</p> <p>Well established, standardized test with normative data available</p>	<p>May be used for all trial phases (Phase I safety deterioration), Phase II, and III safety and efficacy)</p> <p>If a comprehensive assessment of UE muscle impairment is required (e.g., shoulder function, comprehensive hand function) non-key muscles must be added as the UEMS only assesses 5 (bilateral) key muscles</p>
Spinal Cord Independence Measure version III (SCIM III)—self-care and grooming sub-scales	Independence level in ADLs	<p>SCI specific</p> <p>Validated for observation, interview, or self-report</p> <p>Commonly used test with normative data available</p> <p>Psychometrics established in adults and youths (self-report)</p>	<p>Phase II (secondary or exploratory), III (primary or secondary) with impairment-based measure to differentiate recovery from compensation</p> <p>Development of a manual is needed to ensure systematic administration</p> <p>If assessing high cervical injury supplement with another measure</p> <p>Consider co-administration of a COA with a construct focusing on UE function</p>
Capabilities of Upper Extremities Questionnaire (CUE-Q)	Proximal and distal UE function	<p>SCI specific</p> <p>Validated in adolescent and adult populations</p> <p>Assesses proximal and distal function</p> <p>Items worded in a way to minimize compensation</p>	<p>Phase II, III (secondary or exploratory)</p> <p>PRO—should be used in conjunction with an impairment-based measure</p> <p>Consider language and cultural differences when used with non-English speakers</p>
Jebsen Taylor Hand Function Test	Hand function	<p>Standardized performance-based measure with normative references in healthy controls</p> <p>Requires mid-line crossing</p> <p>Kit available for purchase</p>	<p>Phase II, III (secondary or exploratory)</p> <p>Consider addition of SCI-specific measures due to limited psychometric testing in SCI</p> <p>Useful for individuals with incomplete mid-lower level cervical injuries</p> <p>Consider an additional measure if assessing opening and closing of the hand</p> <p>An impairment-based measure should be administered in conjunction to differentiate recovery from compensation</p>
Capabilities of UE function test (CUE-T)	Proximal and distal UE function	<p>SCI specific</p> <p>Items designed to reflect requirements of ADLs</p> <p>Compensation not allowed</p> <p>Components of proximal and distal UE function</p> <p>Assesses uni- and bi-manual function</p>	<p>Phase II, III (secondary or exploratory)</p> <p>Consider additional measures in conjunction as responsiveness has not been published</p> <p>Useful with incomplete SCI</p> <p>Requires tolerance to sitting</p>
Graded Redefined Assessment of Strength, Sensibility, and Prehension (GRASSP)	Multidimensional—hand impairment and function	<p>SCI specific</p> <p>Scores reduced with compensation</p> <p>Manual and kit available for purchase</p> <p>Normative data available</p>	<p>Strength and sensibility: Phase I (safety—deterioration), II and III (safety and efficacy—primary or secondary)</p> <p>Prehension: Phase II (secondary or exploratory), Phase III (primary or secondary)</p> <p>Useful to assess hand function</p> <p>Supplement with other measures if assessing entire upper limb</p> <p>Prehension component requires tolerance to sitting</p>

Table 2 (continued)

Primary construct	Strengths	Recommendations for use in clinical trials
Spinal Cord Injury-Functional Index (SCI-FI) Computer Adaptive Testing (CAT) or Short Form (self-care and fine motor)	SCI specific Large bank of UE items including fine motor Available as short form or computer adaptive test (two versions: one with use of technology, one does not allow use of technology) Links to pediatric SCI-FI	Phase II, III (secondary or exploratory) PRO—should be used in conjunction with an impairment-based measure Consider language and cultural differences when used with non-English speakers
Spinal Cord Ability Ruler (SCAR)	SCI specific Linear assessment Simulated validation (using Rasch analysis) from two databases Statistically combines two common, validated measures (ISNCSCI and SCIM III) so no additional data collection required	Phase II, III (secondary or exploratory) Consider other measures for central cord syndrome Consider the addition of a measure with upper extremity function construct

• Measures considered at the workshop and in this table are listed in the National Institutes of Neurological Disorders and Stroke (NINDS), SCI Common Data Elements (CDEs), or are under development. Recommendations are based on the views of the authors, considering the following: goals of each trial phase (see Table 1), validation (published psychometrics including responsiveness), type of assessment (observed vs. patient reported), and availability. For example, for a Phase III trial using a measure for which validation is not complete or the COA is a patient-reported outcome, the measure is recommended as secondary or exploratory. If validation is complete, and the measure is assessed by observation, the measure is recommended as a primary or secondary outcome measure

• Please note the following websites contain additional information on details, psychometrics, and recommendations for these and other UE COAs:

○ NINDS, CDEs (SCI specific): https://www.commondataelements.ninds.nih.gov/SCI.aspx#tab=Data_Standards

○ Spinal Cord Injury Research Evidence (SCI specific): <https://scireproject.com/outcome-measures/>

○ OSCE Evidence Database to Guide Effectiveness (SCI specific): <http://www.neuropt.org/professional-resources/neurology-section-outcome-measures-recommendations/spinal-cord-injury>

○ Rehab measures.org (non-SCI specific): <http://www.rehabmeasures.org/default.aspx>

subject burden. Capacity measures (e.g., functional independence measure and SCIM) with a limited number of observable items may not adequately identify discrete changes in function due to gaps in content.

Like all PROs, experience performing the activities and an individual's knowledge of their functional limitations are needed for accurate reporting. Thus, the SCI-FI is potentially a good measure of change in performance when used at relatively chronic time points after injury. Consideration should be given to a complementary capacity assessment (by observation in the clinic) using the concepts of CAT, namely that the activities assessed are tailored and appropriate to an individual's level of function.

Traditionally, many of the COA tools used in SCI are ordinal, or rank ordered, but the intervals between rankings are either unknown, unequal, or highly variable. For example, improvement from an ISNCSCI segmental motor score of 0–1 (0 = total paralysis, 1 = palpable or visible contraction) does not imply the same degree of improvement as a change from 3 to 4 (3 = active movement, full range of motion against gravity alone, 4 = active movement, full range of motion against gravity and moderate resistance). Moreover, the numerical changes in one muscle group are not necessarily functionally equivalent to the same changes in a different group, which are then combined to create a total score. The use of such ordinal measures limits the type, accuracy, and strength of the statistical analyses that can be applied and creates challenges in interpreting significant outcomes across a spectrum of impairment and functional deficits.

The SCAR is a new capacity-based assessment described as a measure of “volitional performance” developed to address challenges with ordinal measures [28]. Volitional performance is defined as voluntary task-specific physical actions contributing to independence in activities of daily living (ADLs), e.g., specific SCIM III items associated with voluntary movement. SCAR is designed to inclusively measure recovery across the entire range of severity and level of injury. It combines aspects of volitional performance (function) with strength (impairment), a key contributor to volitional performance. As SCAR contains elements of impairment and function, one may assess recovery vs. behavioral compensation over time. For example, an individual with improved mobility due to compensation alone will have a lower score than an individual with improved strength and mobility, reflecting recovery. The SCAR relies on select items from commonly used ordinal assessments (ISNCSCI and SCIM III), which are converted to a linear measure of volitional performance (Rasch analysis). Psychometric properties of the SCAR were tested and validated using Rasch analysis (retrospective modeling), and were found to be linear, interval

measures with a meaningful total score that reflects changes in functional recovery [28].

As mentioned above, one issue identified with current functional assessment tools (e.g., SCI-FI and SCIM III) is that they are difficult to complete at early time points after injury, whereas it is possible to collect neurological impairment measures (ISNCSCI motor scores) within hours following SCI. When SCAR is utilized in an acute SCI trial, the ISNCSCI component of the measure can be collected at the earliest time points and utilized to estimate recovery by Rasch modeling. As an interval measure, the SCAR provides meaningful scores across the continuum of severity and level of injury, however, further testing and validation of this measure is required.

For established and emerging measures used in clinical trials that assess a single underlying construct, Rasch analysis or similar methods should be considered. While not the focus of this article, a thorough discussion of the rationale for assessing and/or creating linear interval level COAs is addressed by Steeves et al. [28].

Remaining gaps and options to address recruitment

As with most neurological conditions, recruitment to SCI trials is challenging, particularly at acute and subacute time points. Although the incidence of SCI in the United States is ~17,000 per year, patients are dispersed among numerous level I and II trauma centers [1]. Once trial inclusion/exclusion criteria are applied, the number of individuals who are eligible for and consent to participate in clinical trials often becomes low. In addition, SCI is heterogeneous, with varying levels and severities of injury and a high rate of concomitant injuries. Due to these factors and others, there is a “funnel” effect in clinical trial recruitment leading to an extended time frame for the completion of a clinical trial (e.g., due to slow recruitment), resulting in increased costs and decreased enthusiasm on the part of trial sponsors. Traditionally, SCI clinical trials have been completed in a methodical, sequential fashion as they progressed from complete thoracic SCI participants to either thoracic incomplete or cervical complete to cervical incomplete. Although safety studies in a limited SCI thoracic population may be necessary during early phase safety trials of high-risk interventions, the ability to inclusively recruit and simultaneously enroll a broad range of cervical and thoracic complete and incomplete SCI participants in mid- and late-phase trials, would improve trial efficiency. Newer statistical designs and alternative clinical trial methodologies, such as adaptive trial designs, may facilitate more streamlined enrollment.

Prognostic modeling may be used to prospectively stratify initially heterogeneous populations into homogeneous

study cohorts based on identification of those baseline clinical characteristics that significantly predict a future clinical end point. Such modeling techniques may facilitate inclusive enrollment of parallel vs. sequential cohorts, thus decreasing the overall clinical trial duration. One method of prognostic modeling recently used in SCI is unbiased recursive partitioning, a type of regression analysis [30, 31]. The desired clinical end point is chosen by the investigator; unbiased recursive partitioning is then used to determine which early clinical factors (e.g., ISNCSCI motor scores, SCIM) significantly contribute to the end point of choice. Participants are then stratified (split) into homogeneous groups based on the combinations of the early (baseline) predictors. Unbiased recursive partitioning has been tested in both cervical complete and incomplete SCI models as a means to be “inclusive” in trial recruitment while identifying individuals who should be excluded due to ceiling or floor effects. Validation against multiple SCI databases and against established statistical approaches has been undertaken [31]. Once the predictive models have been generated, the investigator can determine which groups (based on the baseline assessment) should be enrolled in the study and determine appropriate target outcomes. Prospective modeling approaches hold promise to appropriately broaden trial participant inclusion, as well as identify applicable outcomes for comparable homogenous cohorts.

Adaptive trial design may be one way to address the challenges associated with recruitment, and should be considered in the context of enrollment and clinical trial progression in order to find the “most direct route to a statistically significant result using the smallest number of subjects” [32]. Adaptive trial design allows for modification of a trial based on incoming data without impacting the validity of the trial. Preset conditions are designated where modifications to dose, sample size, or eligibility criteria may occur based on the analysis of an early subset of the data. For example, in drug-dosing studies, five dosing arms may be initiated simultaneously. In a traditional dose finding study, subjects may be enrolled in doses sequentially or in parallel with a predetermined sample size per group. Adaptive trial design may allow an investigator to stop enrolling in non-efficacious doses, while continuing with more promising doses, thus, saving time and minimizing subject exposure to potential risks. The conditions and decision points are preset, which may lead to a longer period for trial design and setup, but a more efficient overall clinical trial timeline.

Results from prior adaptive or traditional trial designs may also be used to determine trial eligibility criteria. Based on data accumulated in early stage trials, an aggregate score for inclusion may be indicated. Those participants who score above a certain threshold may be recruited, whereas those with scores below would not be recruited as they are

least likely to demonstrate improvement. Not only does this approach streamline enrollment, it also minimizes exposure to unproven therapeutics in individuals who are unlikely to benefit. For example, older individuals with SCI may not recover as well as younger individuals, which may vary depending on the severity of SCI [33]. Additional surrogate markers such as severity of injury based on magnetic resonance imaging may also factor into an aggregate cutoff score. Combining information gained from prognostic modeling using large cohorts and early phase studies to develop “cutoff” scores for trial inclusion could be beneficial.

Conclusion

Optimizing UE function following SCI is essential to enhance independence in ADLs leading to improved quality of life. As clinical trials related to drugs and biologic intervention emerge, there is great need for valid and sensitive COAs that capture improvement in function. In recent years, the field of SCI has matured with the development of SCI-specific COAs, some of which are tailored to assess UE function, others that are more inclusive in their coverage, allowing assessment of upper and lower extremity function. In selecting COAs for a clinical trial, many factors, such as showing evidence of recovery and the ability to demonstrate clinically meaningful change, need to be considered at the trial development stage. Emerging methods for inclusive COAs measures, prognostic modeling, patient stratification, and adaptive trial designs are likely to improve SCI study efficiency and should be further validated in the clinical trial setting. The continued understanding and development of existing measures ensures the field of SCI research will be well positioned to assess the efficacy of emerging interventions and the impact on UE function.

Acknowledgments Organizational support for this work was provided by the Craig H. Neilsen Foundation and Spinal Cord Outcomes Partnership Endeavor. Financial support was provided by the Craig H. Neilsen Foundation. All authors reviewed and contributed to this manuscript. All authors except J.D.G. and G.S. participated in the workshop.

Compliance with ethical standards

Conflict of interest L.A.T.J., T.L.W., and N.K. are employed by the Craig H. Neilsen Foundation. The remaining authors declare no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits use, duplication, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s)

and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

Appendix 1: Workshop Participants

Kim Anderson, PhD, Miami, Florida, USA
 Michael Beattie, PhD, San Francisco, California, USA
 Andrew Blight, PhD, Ardsley, New York, USA
 Anne Bryden, OT, Cleveland, Ohio, USA
 Armin Curt, MD, Zurich, Switzerland
 Edelle Field-Fote, PT, PhD, Atlanta, Georgia, USA
 Jane Hsieh, MSc, Toronto, Ontario, Canada
 Lyn Jakeman, PhD, Bethesda, Maryland, USA
 Linda Jones, PT, MS, Encino, California, USA
 Sukhvinder Kalsi-Ryan, PT, PhD, Toronto, Ontario, Canada
 Naomi Kleitman, PhD, Encino, California, USA
 Laura Krisa, PhD, Philadelphia, Pennsylvania, USA
 Daniel Lammertse, MD, Denver, Colorado, USA
 Benjamin Leiby, PhD, Philadelphia, Pennsylvania, USA
 Ralph Marino, MD, Philadelphia, Pennsylvania, USA
 Mary Jane Mulcahey, OT, PhD, Philadelphia, Pennsylvania, USA
 Jan Schwab, MD, PhD, Columbus, Ohio, USA
 John Steeves, PhD, Vancouver, British Columbia, Canada
 Keith Tansey, MD, PhD, Jackson, Mississippi, USA
 David Tulskey, PhD, Newark, Delaware, USA
 Ed Wirth, MD, PhD, Fremont, California, USA
 José Zariffa, PhD, Toronto, Ontario, Canada
 Tracey Wheeler, PhD, Encino, California, USA

References

- National Spinal Cord Injury Statistical Center. Facts and figures at a glance. Birmingham, AL: University of Alabama at Birmingham; 2016.
- American Spinal Injury Association. International standards for neurological classification of spinal cord injury. Atlanta, GA: ASIA (revised 2011, updated 2015).
- Simpson LA, Eng JJ, Hsieh JT, Wolfe DL. The health and life priorities of individuals with spinal cord injury: a systematic review. *J Neurotrauma*. 2012;29:1548–55.
- Consortium for Spinal Cord Medicine. Outcomes following traumatic spinal cord injury: clinical practice guidelines for health-care professionals. *J Spinal Cord Med*. 2000;23:289–316.
- WHO. International classification of functioning, disability, and health: ICF. Geneva: World Health Organization; 2001.
- Levin MF, Kleim JA, Wolf SL. What do motor “recovery” and “compensation” mean in patients following stroke? *Neurorehabil Neural Repair*. 2009;23:313–9.
- Curt A, Van Hedel HJ, Klaus D, Dietz V. Recovery from a spinal cord injury: significance of compensation, neural plasticity, and repair. *J Neurotrauma*. 2008;25:677–85.
- Rudhe C, van Hedel HJ. Upper extremity function in persons with tetraplegia: relationships between strength, capacity, and the spinal cord independence measure. *Neurorehabil Neural Repair*. 2009;23:413–21.
- Behrman AL, Ardolino E, Vanhiel LR, Kern M, Atkinson D, Lorenz DJ, et al. Assessment of functional improvement without compensation reduces variability of outcome measures after human spinal cord injury. *Arch Phys Med Rehabil*. 2012;93:1518–29.
- Harkema SJ, Shogren C, Ardolino E, Lorenz DJ. Assessment of functional improvement without compensation for human spinal cord injury: extending the neuromuscular recovery scale to the upper extremities. *J Neurotrauma*. 2016;33:2181–90.
- Mulcahey MJ, Smith BT, Betz RR. Evaluation of the lower motor neuron integrity of upper extremity muscles in high level spinal cord injury. *Spinal Cord*. 1999;37:585–91.
- Curt A, Dietz V. Neurographic assessment of intramedullary motoneurone lesions in cervical spinal cord injury: consequences for hand function. *Spinal Cord*. 1996;34:326–32.
- Bryden AM, Hoyen HA, Keith MW, Mejia M, Kilgore KL, Nemunaitis GA. Upper extremity assessment in tetraplegia: the importance of differentiating between upper and lower motor neuron paralysis. *Arch Phys Med Rehabil*. 2016;97:S97–104.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407–15.
- Hart DL, Wang YC, Stratford PW, Mioduski JE. A computerized adaptive test for patients with hip impairments produced valid and responsive measures of function. *Arch Phys Med Rehabil*. 2008;89:2129–39.
- Wang YC, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. *Phys Ther*. 2011;91:675–88.
- Wu X, Liu J, Tanadini LG, Lammertse DP, Blight AR, Kramer JL, et al. Challenges for defining minimal clinically important difference (MCID) after spinal cord injury. *Spinal Cord*. 2015;53:84–91.
- Biering-Sorensen F, Alai S, Anderson K, Charlifue S, Chen Y, DeVivo M, et al. Common data elements for spinal cord injury clinical research: a National Institute for Neurological Disorders and Stroke project. *Spinal Cord*. 2015;53:265–77.
- Kalsi-Ryan S, Beaton D, Curt A, Duff S, Popovic MR, Rudhe C, et al. The graded redefined assessment of strength sensibility and prehension: reliability and validity. *J Neurotrauma*. 2012;29:905–14.
- Marino RJ, Patrick M, Albright W, Leiby BE, Mulcahey M, Schmidt-Read M, et al. Development of an objective test of upper-limb function in tetraplegia: the capabilities of upper extremity test. *Am J Phys Med Rehabil*. 2012;91:478–86.
- Kalsi-Ryan S, Beaton D, Ahn H, Askes H, Drew B, Curt A, et al. Responsiveness, sensitivity, and minimally detectable difference of the graded and redefined assessment of strength, sensibility, and prehension, version 1.0. *J Neurotrauma*. 2016;33:307–14.
- Kalsi-Ryan S, Beaton D, Curt A, Duff S, Jiang D, Popovic MR, et al. Defining the role of sensation, strength, and prehension for upper limb function in cervical spinal cord injury. *Neurorehabil Neural Repair*. 2014;28:66–74.
- Kalsi-Ryan S, Curt A, Verrier MC, Fehlings MG. Development of the graded redefined assessment of strength, sensibility and prehension (GRASSP): reviewing measurement specific to the upper limb in tetraplegia. *J Neurosurg Spine*. 2012;17:65–76.
- Velstra IM, Bolliger M, Tanadini LG, Baumberger M, Abel R, Rietman JS, et al. Prediction and stratification of upper limb function and self-care in acute cervical spinal cord injury with the

- graded redefined assessment of strength, sensibility, and prehension (GRASSP). *Neurorehabil Neural Repair*. 2014;28:632–42.
25. Velstra IM, Curt A, Frotzler A, Abel R, Kalsi-Ryan S, Rietman JS, et al. Changes in strength, sensation, and prehension in acute cervical spinal cord injury: European multicenter responsiveness study of the GRASSP. *Neurorehabil Neural Repair*. 2015;29:755–66.
 26. Tulsy DS, Jette AM, Kisala PA, Kalpakjian C, Dijkers MP, Whiteneck G, et al. Spinal cord injury-functional index: item banks to measure physical functioning in individuals with spinal cord injury. *Arch Phys Med Rehabil*. 2012;93:1722–32.
 27. Jette AM, Tulsy DS, Ni P, Kisala PA, Slavin MD, Dijkers MP, et al. Development and initial evaluation of the spinal cord injury-functional index. *Arch Phys Med Rehabil*. 2012;93:1733–50.
 28. Reed R, Mehra M, Kirshblum S, Maier D, Lammertse D, Blight A, et al. Spinal cord ability ruler: an interval scale to measure volitional performance after spinal cord injury. *Spinal Cord*. 2017;55:730–38.
 29. Heinemann AW, Dijkers MP, Ni P, Tulsy DS, Jette A. Measurement properties of the spinal cord injury-functional index (SCI-FI) short forms. *Arch Phys Med Rehabil*. 2014;95:1289–97.e5
 30. Tanadini LG, Hothorn T, Jones LA, Lammertse DP, Abel R, Maier D, et al. Toward Inclusive trial protocols in heterogeneous neurological disorders: prediction-based stratification of participants with incomplete cervical spinal cord injury. *Neurorehabil Neural Repair*. 2015;29:867–77.
 31. Tanadini LG, Steeves JD, Hothorn T, Abel R, Maier D, Schubert M, et al. Identifying homogeneous subgroups in neurological disorders: unbiased recursive partitioning in cervical complete spinal cord Injury. *Neurorehabil Neural Repair*. 2014;28:507–15.
 32. Babbs CF. Choosing inclusion criteria that minimize the time and cost of clinical trials. *World J Methodol*. 2014;4:109–22.
 33. Wilson JR, Davis AM, Kulkarni AV, Kiss A, Frankowski RF, Grossman RG, et al. Defining age-related differences in outcome after traumatic spinal cord injury: analysis of a combined, multi-center dataset. *Spine J*. 2014;14:1192–8.

Affiliations

Linda A. T. Jones¹ · Anne Bryden² · Tracey L. Wheeler¹ · Keith E. Tansey^{3,4,5} · Kim D. Anderson⁶ · Michael S. Beattie⁷ · Andrew Blight⁸ · Armin Curt^{9,10} · Edelle Field-Fote^{11,12,13} · James D. Guest⁶ · Jane Hsieh^{14,15} · Lyn B. Jakeman¹⁶ · Sukhvinder Kalsi-Ryan^{17,18} · Laura Krisa¹⁹ · Daniel P. Lammertse^{20,21} · Benjamin Leiby¹⁹ · Ralph Marino¹⁹ · Jan M. Schwab²² · Giorgio Scivoletto²³ · David S. Tulsy²⁴ · Ed Wirth²⁵ · José Zariffa^{17,18} · Naomi Kleitman¹ · Mary Jane Mulcahey¹⁹ · John D. Steeves²⁶

¹ Craig H. Neilsen Foundation, Encino, CA, USA

² Case Western Reserve University, Cleveland, OH, USA

³ University of Mississippi Medical Center, Jackson, MS, USA

⁴ Methodist Rehabilitation Center, Jackson, MS, USA

⁵ Veterans Administration Medical Center, Jackson, MS, USA

⁶ University of Miami, Miami, FL, USA

⁷ University of California, San Francisco, CA, USA

⁸ Acorda Therapeutics, Ardsley, NY, USA

⁹ University Hospital Balgrist, Zurich, Switzerland

¹⁰ University of Zurich, Zurich, Switzerland

¹¹ Shepherd Center, Atlanta, GA, USA

¹² Emory University, Atlanta, GA, USA

¹³ Georgia Institute of Technology, Atlanta, GA, USA

¹⁴ Wings for Life, Salzburg, Austria

¹⁵ Parkwood Institute, London, ON, Canada

¹⁶ National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA

¹⁷ Toronto Rehabilitation Institute, Toronto, ON, Canada

¹⁸ University of Toronto, Toronto, ON, Canada

¹⁹ Thomas Jefferson University, Philadelphia, PA, USA

²⁰ Craig Hospital, Englewood, CO, USA

²¹ University of Colorado, Aurora, CO, USA

²² The Ohio State University, Columbus, OH, USA

²³ Foundation Saint Lucia, Rome, Italy

²⁴ University of Delaware, Newark, DE, USA

²⁵ Asterias Biotherapeutics, Fremont, CA, USA

²⁶ University of British Columbia, Vancouver, Canada