

Review

Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: clinical trial outcome measures

JD Steeves^{*1}, D Lammertse², A Curt¹, JW Fawcett³, MH Tuszynski⁴, JF Ditunno⁵, PH Ellaway⁶, MG Fehlings⁷, JD Guest⁸, N Kleitman⁹, PF Bartlett¹⁰, AR Blight¹¹, V Dietz¹², BH Dobkin¹³, R Grossman¹⁴, D Short¹⁵, M Nakamura¹⁶, WP Coleman¹⁷, M Gaviria¹⁸ and A Privat¹⁸

¹ICORD, University of British Columbia (UBC) and Vancouver Coastal Health (VCH) Research Institute, Vancouver, BC, Canada; ²Craig Hospital, Englewood, CO, USA; ³Cambridge University Centre for Brain Repair, Robinson Way, Cambridge, UK; ⁴Center for Neural Repair, University of California at San Diego, La Jolla, CA, USA; ⁵Jefferson Medical College, Thomas Jefferson University, Philadelphia, PA, USA; ⁶Department of Movement and Balance, Division of Neuroscience and Mental Health, Imperial College London, Charing Cross Campus, London, UK; ⁷University of Toronto, Krembil Neuroscience Center, Head Spine and Spinal Cord Injury Program, Toronto Western Hospital, Toronto, Ontario, Canada; ⁸Department of Neurological Surgery and the Miami Project to Cure Paralysis, Lois Pope LIFE Center, Miami, FL, USA; ⁹National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, USA; ¹⁰Queensland Brain Institute, University of Queensland, St Lucia, Queensland, Australia; ¹¹Acorda Therapeutics, Hawthorne, NY, USA; ¹²Spinal Cord Injury Center, Balgrist University Hospital, Zurich, Switzerland; ¹³Department of Neurology, University of California Los Angeles, Geffen School of Medicine, Neurologic Rehabilitation and Research Program, Los Angeles, CA, USA; ¹⁴Baylor College of Medicine, Department of Neurosurgery, One Baylor Plaza, Houston, TX, USA; ¹⁵Midlands Centre for Spinal Injuries, Robert Jones and Agnes Hunt Orthopaedic and District Hospital NHS Trust, Oswestry, Shropshire, UK; ¹⁶Department of Orthopaedic Surgery, Keio University, School of Medicine, Tokyo, Japan; ¹⁷WPCMath, Buffalo, NY, USA; ¹⁸Institut des Neurosciences – CHU St Eloi, INSERM U-583, Montpellier, France

An international panel reviewed the methodology for clinical trials of spinal cord injury (SCI), and provided recommendations for the valid conduct of future trials. This is the second of four papers. It examines clinical trial end points that have been used previously, reviews alternative outcome tools and identifies unmet needs for demonstrating the efficacy of an experimental intervention after SCI. The panel focused on outcome measures that are relevant to clinical trials of experimental cell-based and pharmaceutical drug treatments. Outcome measures are of three main classes: (1) those that provide an anatomical or neurological assessment for the connectivity of the spinal cord, (2) those that categorize a subject's functional ability to engage in activities of daily living, and (3) those that measure an individual's quality of life (QoL). The American Spinal Injury Association impairment scale forms the standard basis for measuring neurologic outcomes. Various electrophysiological measures and imaging tools are in development, which may provide more precise information on functional changes following treatment and/or the therapeutic action of experimental agents. When compared to appropriate controls, an improved functional outcome, in response to an experimental treatment, is the necessary goal of a clinical trial program. Several new functional outcome tools are being developed for measuring an individual's ability to engage in activities of daily living. Such clinical end points will need to be incorporated into Phase 2 and Phase 3 trials. QoL measures often do not correlate tightly with the above outcome tools, but may need to form part of Phase 3 trial measures.

Spinal Cord (2007) 45, 206–221. doi:10.1038/sj.sc.3102008; published online 19 December 2006

Keywords: spinal cord injury; clinical trial; neurologic assessment; outcome measures; functional recovery

*Correspondence: J Steeves, ICORD, University of British Columbia (UBC) and Vancouver Coastal Health (VCH) Research Institute, c/o 2469-6270 University Boulevard, Vancouver, BC, Canada V6T 1Z4

Introduction

The second International Campaign for Cures of spinal cord injury Paralysis (ICCP) Clinical Guidelines Panel

meeting focused on the outcome measures to be used during spinal cord injury (SCI) clinical trials for the evaluation of a therapeutic intervention. Given the small number of clinical trials that have been undertaken for SCI, it is not surprising that until now there has been little opportunity to develop agreement as to the most appropriate and accurate clinical end points (ie outcome measures) for demonstrating the efficacy of an experimental therapeutic intervention.¹ The various possible outcome measures with their advantages and disadvantages are reviewed in this article.

Challenges for assessing SCI outcomes or benefits of therapeutic interventions

In terms of designing a specific SCI clinical trial with the most accurate assessment of neurological or functional outcome, a consideration of the following issues is suggested:

- Phase of clinical trial, as primary and secondary outcome measures and thresholds are likely to differ or evolve from Phase 1 (safety) to Phase 3 (therapeutic confirmatory trials).
- Level of spinal injury, including the extent of the zone of partial preservation (ZPP).
- Severity of spinal injury (varying degrees of incomplete to complete sensorimotor loss).
- Time since injury (early acute to late chronic; ie from unstable to more stable functional capacities after SCI)
- Appropriate nature of outcome measure to the capacity or capability being evaluated (eg sensorimotor impairment, autonomic function, personal functional capacity, performance, or community participation). Different clinical targets normally require distinct outcome assessment tools.
- Sensitivity of outcome measure (ie detection threshold).
- Accuracy and validation of outcome assessment tool.
- Reliability of measurements between assessments by a single investigator and between investigators (ie intra- and inter-rater reliability).
- Feasibility for using selected outcome measurement tools in a particular center or across multiple centers.
- Adoption of standardized outcome assessment procedures and data sets across multiple trial centers.

We will discuss these and other influences as they impact the selection of outcome measures for SCI clinical trials.

Categories of outcome assessments

Assessment methodologies for evaluating a clinical end point for an SCI trial fall into three main categories:

- (a) Assessments aimed at describing the neurological connectivity of the spinal cord, irrespective of the ability of the patient to functionally use those connections in everyday activity. The American

Spinal Injury Association (ASIA) scale would be an example of such an assessment. This would also include assessments of neurological capacity that are independent of the environment (eg electrophysiological recordings or imaging assessments). If these outcome tools can be shown to accurately predict the long-term functional benefits (clinical endpoints) resulting from a therapeutic intervention, they can also be thought of as surrogate end points.

- (b) Assessments of the abilities of a patient with SCI to perform activities associated with everyday life. Examples would be the Functional Independence Measure (FIM) and the Spinal Cord Independence Measure (SCIM). Functional evaluations may be a more direct measurement of a clinically meaningful change in the functional capacity of a study subject, but the changes in functional outcomes may not always be the result of a demonstrated change in spinal-neurological activity or connectivity. In short, any change in a person's functional capacity after SCI may be due to adaptive changes (or plasticity) within and/or without the central nervous system (CNS), including environmental accommodations and/or alternative compensatory strategies.
- (c) Assessments of an individual's level of participation in societal activities. Quality of life (QoL) can be defined as a person's perception of his position in life, within the context of both his personal and society's values and culture, and relate to his personal concerns, standards and goals. The short form 12- or 36-item medical outcomes health survey (SF-12 and SF-36) are examples of a QoL survey.

Improvement of functional abilities, reflected in activities of daily living (see above) will be the most meaningful and valued outcomes. However, the early phase clinical trials (Phase 1 and 2) that have been completed to date (using pharmaceutical therapeutics), have focused on assessment of neurological connectivity to provide 'proof of principle' measures. It is likely that neurological assessments will continue to be used as primary outcome measures, indicating the likelihood that a treatment will improve the functional capacities and performance of a subject in later phases of clinical studies. However, no experimental intervention will be considered effective for the treatment of people living with SCI unless it improves their ability to function and engage in everyday life within their society. Outcome assessment tools that accurately and sensitively demonstrate such benefits will need to be incorporated into the more definitive and confirmatory Phase 3 clinical trials.

Clinical trial phases and corresponding categories of outcome measures

Phase 1

The objectives of Phase 1 trials can be quite varied, from the initial exploration of tolerability, through study of human pharmacokinetics and metabolism, to identification of the maximum safely tolerated dose of a

candidate therapeutic (see also Lammertse *et al*²). A Phase 1 trial is specifically designed to evaluate the safety of the intervention and expose any adverse or toxic side effects, usually in small numbers of subjects with a simple open label design. Participant's who choose to take part in a Phase I trial may experience significant risks with a limited probability of receiving individual benefit. Preliminary Phase 2 (proof of concept or evidence of activity) data are sometimes collected during a Phase 1 trial, but only to develop a preliminary sense of potential efficacy and to assist in the identification of appropriate outcome measures to be used in subsequent properly powered Phase 2 or 3 trials. Many of the currently conceived therapeutics for the possible treatment of SCI involve an invasive intervention, such as the direct infusion of a drug or cellular transplant into, or around the injured spinal cord. As a consequence, healthy volunteers (without SCI) are unlikely to be recruited for a Phase 1 SCI clinical trial of this type.

SCI is a heterogeneous disorder in terms of level of spinal injury, severity of injury and timing of treatments after injury. Some types of SCI (eg central cord syndrome and cauda equina injuries) have higher spontaneous rates of overall sensory and motor recovery. Thus, they may not be the best subjects to be included with other types of traumatic SCI during a Phase 1 or Phase 2 trial, as they could increase the variability of the outcome data. They may also be inappropriate, based on the proposed mechanism of action for the experimental intervention.

Patients with complete ASIA A thoracic injuries are frequently suggested as being the 'preferred' group of SCI participants for early phase SCI clinical trials. By confining the administration of the experimental therapeutic to the thoracic cord, it is probable that any adverse effects on spinal function would not seriously alter a person's functional capabilities (ie not spread to more rostral cervical levels and compromise arm, hand or respiratory function). Complete ASIA A, thoracic-injured patients are a small proportion of total SCI cases, and there are, as yet, no validated outcome measures for changes in thoracic cord motor function (although some are under development, see below). Sensory function can be evaluated using the ASIA examination or other measures.

General Phase 1 trial safety outcome measures include: ongoing assessment of standard vital signs, physical examination data (eg temperature, respiration, heart rate, and blood pressure), clinical laboratory tests (eg hematology and urine analysis), as well as the appearance of any systemic adverse event (observed or reported by a trial subject). Depending on the therapeutic drug or cell line being evaluated and the route of administration, other Phase 1 safety outcome measures may include the evaluation of unintended effects on the CNS or other body tissues, including infection, inflammation, or immune reactions.

A more specific measure of neurological state is the ASIA assessment³ to determine whether there is any change in neurological level or any sensorimotor

deterioration, as well as to subsequently track any changes in the ASIA score. An improvement in ASIA scores is a possibility during a Phase 1 trial, indicating possible efficacy of the treatment, but this is not the primary reason for including an ASIA assessment at this stage of clinical study. An ASIA assessment, just before randomization of a subject to a clinical trial study arm, can be most useful to assure that the candidate meets all inclusion criteria and whether the participating subjects should be stratified (into a sub-category) on the basis of their ASIA score, so only appropriately matched experimental and control subjects are compared thereafter.

Inclusion of ongoing standardized ASIA assessments is warranted on the grounds that this examination: (1) has been widely adopted throughout the world, enabling the comparison of data between centers, (2) can be readily undertaken with a minimum of equipment, and (3) can provide important reference data between different phases of a clinical trial or with previous trial (historical) data. In several previous randomized control trials (RCT),⁴⁻⁸ motor and sensory assessments, comparable with the current ASIA standards have been used as an overall indicator of the general severity of neurological impairment after SCI (especially in terms of segmental motor function, see below).

Later in this document, we will discuss when and how often an ASIA assessment should be undertaken, the strengths and limitations of the ASIA examination, the separation of upper and lower limb ratings, as well as the intra- and inter-rater reliability of the ASIA assessment (see below).

Phase 2

During a Phase 2 study (sometimes referred to as the Proof of Concept level), an exploratory evaluation of efficacy becomes more prominent, with the objective of determining potential effect size and variability of an experimental therapy in comparison to a useful control group. Information is gained regarding choice of optimal end points for a larger Phase 3 confirmatory trial of efficacy. During a Phase 2 trial, additional information is also obtained regarding safety. Combined Phase 1/2 trials, where safety and bioactivity of the therapeutic are evaluated together can often occur when the Phase 1 trial does not involve healthy subjects and is restricted to people having the clinical disorder. It is possible for SCI clinical trials to be designed in this manner. Nevertheless, the data from such a combined Phase 1/2 trial must be able to satisfy the essential outcome end points for each respective trial phase.

The preferred Phase 2 design would be a RCT where each participant is recruited prospectively and randomly assigned to either the experimental or control arm of the study and where the investigators and, if at all possible the participants, are blinded to which study arm they have been assigned. If available, Phase 2 trials could employ surrogate end points, which are expected to be predictors of functional improvement, to estimate

presumed effective doses, and to allow trials of shorter duration and smaller size to be conducted.

Phase 3

Phase 3 (therapeutic confirmatory) trials are generally the definitive clinical trial phase and typically undertaken as a RCT. The object is to confirm the preliminary evidence obtained at the Phase 2 stage with a statistically significant clinical benefit of the therapeutic in a wider group of subjects across multiple study centers. For a more detailed discussion of Phase 3 and Phase 4 trial stages, see accompanying article – SCI Guidelines 4 (Lammertse *et al*²).

SCI therapies conceived as early interventions or acute stage treatments are likely to be administered within days of spinal injury and it is important that the outcome tools have the ability to accurately and sensitively track meaningful changes across a broad chronological timeframe. Several assessment tools are available or are being developed, each with their individual strengths and limitations. We will discuss each separately.

ASIA impairment scale

ASIA assessments

As mentioned above, the ASIA Impairment Scale has become a standardized and routinely adopted classification for most patients suspected of suffering a SCI.³ It is especially useful for classification of motor-complete and sensory-complete SCI (ASIA A) as well as motor-complete, sensory-incomplete SCI (ASIA B). During the acute stages of SCI, there have been concerns about how soon after injury the ASIA examination can provide useful prognostic information about the eventual degree of impairment. It has been argued that an ASIA assessment within the first 24 h may not provide an accurate prognosis and that a later 72 h examination is a more reliable indicator, as the patient is medically more stable.^{9–11} At chronic time points (greater than 12 months after SCI), the ASIA assessment may not capture the most important aspects of functional changes after SCI. Nevertheless, it is still valuable for classifying and stratifying participants for a clinical trial. Functional tests (see below) are perhaps more useful primary outcome tools for chronic studies.

Regardless of these concerns, it is essential that steps should be taken to standardize and optimize the accuracy of the ASIA assessment. For all patients being considered for entry into a trial, the clinical trial center(s) must conduct an independent and blind ASIA assessment, just before randomization to the therapeutic intervention or relevant control treatment. Subsequent follow-up ASIA assessments should also be undertaken at relevant time points over the course of recovery, as defined for that trial (eg first few weeks, first couple of months, and then at fixed intervals, every few months, throughout the duration of the study) in the same blinded fashion, and preferably by the same examiner.

In the absence of a more sensitive and accurate outcome tool, such ASIA assessments enable any initial detriments or benefits to be identified and followed.

The Panel strongly recommends that ASIA assessors undergo standardized training with an intra- and inter-rater reliability test being completed at the end of the training session. Follow-up training of the same examiners should be undertaken at reasonable intervals (eg every 6–12 months) by the same qualified trainers. This is especially important when it is necessary to undertake the clinical trial at more than one site. Although the ASIA assessment paradigm seems simple in its description, experience has indicated that rigorous adherence to the definitions, based on training, is necessary to obtain consistent data that can be meaningfully compared both within and across clinical studies or centers.

Previous SCI clinical trial experience^{4–8} suggests that requiring the improvement of one or two ASIA grades over and above spontaneous recovery (eg ASIA B to ASIA C or ASIA D), as a primary outcome end point (to document the benefit of a therapeutic intervention), may be too demanding a threshold (ie is a relatively insensitive measure for a therapeutic effect). A candidate therapeutic with a very large effect size could be addressed with such a challenging clinical point. However, an intervention with a potentially smaller effect size might require a more sensitive outcome measure, such as a statistically significant change in ASIA motor score.

ASIA motor score

In many respects, the ASIA motor score is considered more reliable than the ASIA sensory score in predicting functional outcome after SCI.¹² The Panel recommends that upper and lower limb motor scores should be compiled separately as the upper-extremity motor score (UEMS) and lower-extremity motor score (LEMS). This enables a change in motor function to be more clearly tracked and recorded as specific to either the cervical or lumbar levels (Table 1). Separation of the motor scores into UEMS and LEMS also reduces the influence that a large change in the functional strength in one or a few muscles might have on the interpretation of therapeutic benefit.

In general, establishing a functionally meaningful ASIA motor score threshold to document the benefit of a therapeutic intervention is dependent both on the level and severity of the SCI,¹³ as well as the degree of spontaneous recovery after SCI with conventional treatment (Table 2 and Fawcett *et al*¹¹). As shown in Table 2, previous studies have indicated that a low-cervical, ASIA A-injured patient is likely to spontaneously improve about 10 ASIA motor points during the first year after SCI.^{7,8,14,15} Thus, to demonstrate the efficacy of a therapeutic intervention, a response to treatment of an additional 10-point improvement in the ASIA motor score (efficacy threshold now being 20 point) might be considered a valid primary outcome end point (cf Fawcett *et al*¹¹).

Table 1 Key muscles used for ASIA motor score assessment, with muscle grades categorizing functional assessment of each muscle's contraction

<i>Left side (max. grade)</i>	<i>Key muscles for ASIA motor score assessment and primary level of spinal innervation</i>	<i>Right side (max. grade)</i>
5	Elbow flexors (biceps brachialis) – C5	5
5	Wrist extensors (extensor carpi radialis longus and brevis) – C6	5
5	Elbow extensor (triceps) – C7	5
5	Finger flexors (flexor digitorum profundus, middle finger) – C8	5
5	Finger abductors (abductor digiti minimi, little finger) – T1	5
25	Upper Extremity Motor Score (UEMS)	25
5	Hip flexors (iliopsoas) – L2	5
5	Knee extensors (quadriceps) – L3	5
5	Ankle dorsiflexors (tibialis anterior) – L4	5
5	Long toe extensors (extensor hallucis longus) – L5	5
5	Ankle plantar flexors (gastrocnemius, soleus) – S1	5
25	Lower Extremity Motor Score (LEMS)	25
50	Total ASIA motor score (= 100 for both sides)	50

ASIA muscle grades: 0 = total paralysis; 1 = palpable or visible contraction; 2 = active movement, gravity eliminated; 3 = active movement, against gravity; 4 = active movement, against some resistance; 5 = active movement

Table 2 Spontaneous' improvement in ASIA motor scores for complete and incomplete cervical SCI at 1 year

<i>Initial ASIA classification of cervical-level SCI</i>	<i>Geisler et al. (2001)</i>	<i>Marino et al. (1999)</i>	<i>Waters et al. (1993)</i>	<i>EMSCI</i>
ASIA A	12.3 ± 13.7 (n = 264)	9.6 ± 12.7 (n = 808)	10 (n = 61)	14 + 18.8 (n = 61)
ASIA B	37.1 ± 27.8 (n = 88)	28.2 ± 25.6 (n = 242)		32.0 + 22.3 (n = 12)
ASIA C		43.0 ± 20.4 (n = 295)		43 ± 17 (n = 31)
ASIA C and D	51.9 ± 18.1 (n = 105)			33 ± 17 (n = 66)
ASIA D		25.7 + 20.1 (n = 215)		24 + 12 (n = 35)

Different efficacy thresholds would need to be specified for a response at each level and severity of SCI. For example, the spontaneous recovery of ASIA B cervical patients, 1 year after a cervical SCI, has been reported to be about 30 motor points (Table 2), and thus might require an additional 20 point improvement to indicate a clinically meaningful benefit for an intervention. Such a threshold would allow demonstration of benefit with a reasonable number of trial subjects. However, these requirements could be complicated by a 'ceiling' in ASIA motor scores. As no ASIA motor score is collected between T2 and L1, only a physiological assessment of motor connectivity could be reliably undertaken with the thoracic region (see below). It should be noted that the absolute difference in the number of ASIA motor points between an experimental and appropriately matched control group is not as important as whether a statistically valid difference exists and whether that magnitude of difference confers a clinical benefit (ie an improved functional outcome) to the person with SCI.

Finally, several studies have reported a substantial (25–50) motor point improvement over the first year

after SCI for people with ASIA C and D classifications (Table 2), which is on top of their initial ASIA motor score. Thus, an ASIA motor score 'ceiling effect' may make it difficult to discriminate a statistical difference between the ASIA motor scores of SCI participants in the experimental and control arms of a study. In short, the spontaneous ASIA motor score may become so high within the recovery period that a treatment effect will not be detectable. Therefore, a functional test (see below) may be a more appropriate primary outcome tool for ASIA C and ASIA D trial participants.

Statistically speaking, the use of ASIA motor scores as a primary outcome end point is perhaps most useful for SCI subjects initially enrolled in a clinical trial as either ASIA A or ASIA B. The obvious drawback for ASIA A and ASIA B subjects is that they initially have motor-complete spinal injuries and it may be difficult to produce or discern a clinically meaningful improvement in their ASIA motor score.

For reasons arising from the underlying physiology and the natural history of spontaneous recovery, the ASIA motor scores may not always represent a normal, bell-shaped curve and this may make normal-theory

statistical procedures like the *t*-test and analysis of variance incorrect in small samples. As different inclusion and exclusion criteria can affect the representation of these subgroups in the total composition of the study sample (cf Tuszynski *et al*¹⁶), estimates of the standard deviation based on one trial may be inaccurate in predicting the standard deviation in a new trial. In a large sample, the number of patients with low or zero change in the ASIA motor scores, can skew the distribution to the left and leave a large peak. In any case, the changes can still show 'ceiling' effects in people with mild SCI.

These technical statistical problems suggest why it may sometimes be attractive to use a binary (success/failure) criterion as a trial's primary outcome measure, rather than an ordinal variable like the ASIA motor score. Although binary variables always have a completely known, parametric probability distribution that can be used by statisticians confidently, they are likely to mask underlying clinical complexities and/or variability.

Some reports have expressed the ASIA motor score as the 'percent deficit recovered.' Although this strategy has an appealing rationale, it also has a potential danger. It may be that a mild SCI injury, with only a few points in ASIA motor deficit, has a larger chance for spontaneous recovery. Thus, this method would allow mildly injured patients to have disproportionate weight in one direction, whereas patients with severe motor deficit would count heavily in the other direction as they are least likely to improve. The method of presenting the ASIA score as the number of motor points changed from baseline can give more potential weight to the severely spinal injured group (whether you use an individual baseline or the mean of the subgroup), as they have numerically more room to improve. Perhaps the best solution might be to use the number of motor points changed, but to compensate by stratifying the subject population into cohorts or subgroups on the basis of the initial classification of ASIA impairment scale (AIS) severity.

Adjusting for baseline differences has been used, as in the NASCIS III study.⁶ Simply introducing a baseline term in an analysis of covariance may not be sufficient, as the amount of correction required may be different for patients with mild, moderate, and severe SCI in a manner that is not linearly proportional across the range of SCI severity. Also, this introduces a mathematical relation between the outcome variable (the change in ASIA motor score) and the predictor (the initial baseline score) that could make the envelope of data points depart from the commonly assumed model, where the scatter of the data above and below the regression line has a normal distribution with uniform variance.

The outcome of a trial can depend strongly on its mixture of population subgroups and clinical covariates (also see Lammertse *et al*²). In order to design a clinical trial properly, it is important to recognize and distinguish the different questions and problems: (1) as the natural history (ie spontaneous recovery) is different for different SCI severities. For example, if more patients

with AIS grade A and fewer patients with grade C are disproportionately assigned to the test treatment, then that treatment will appear artificially of less benefit as ASIA grade A subjects will probably always exhibit the smallest treatment effect, (2) even if the outcome of the trial is positive, any randomization imbalance will provide ammunition for skeptics to find *post hoc* rationalizations for disbelieving otherwise sound results, (3) even if there is no randomization imbalance at all, there is still the possibility that the test treatment will be less effective in certain groups. For example, it is likely that the target and functional recovery mechanisms available in a subject with an ASIA C injury will differ from those in a patient with an ASIA A injury, (4) even aside from the question of power it may be scientifically and clinically important at the end of the trial to know if there are effect differences among identifiable cohorts or subgroups, and (5) any result is more scientifically credible if hypothesized in advance than if found *ad hoc* or *post hoc*. Therefore, the most important covariates should be identified during trial design and included in the primary analysis. Indeed, a major purpose of the current series of papers is to provide designers with historical data that can be used in calculations, sensitivity analyses, and simulations that can help a designer to determine whether a planned trial is likely to succeed (see Lammertse *et al*²).

There are three means available to deal with covariates and subgroups: (1) to include them as strata in a block randomization, (2) to model them as explicit terms in the trial's single, prospectively specified 'primary efficacy analysis', and (3) to include them in prospectively specified secondary analyses. None of these approaches is unrestrictedly useful and trial designers should probably employ all three.

Stratified randomization only protects against randomization imbalance, not differences in effect size. Also, it would be a bad idea to include too many factors as strata, as, if the block size becomes large compared to the recruitment at the individual centers, too many incomplete blocks will be left at the end of the trial and this would precisely defeat the purpose.

Identifying and restricting the number of study covariates to a small number normally has the effect of increasing power (in the overall test, rather than in the individual subgroup tests) and therefore decreasing the necessary number of study subjects (ie sample size). In general terms, unexplained variability is reduced when individuals are considered within their own more homogeneous subgroup, and this increases statistical power. However, if relatively unimportant covariates are included 'for completeness,' then statistical tests will exact a penalty and the power will actually be less and not more. Also, as the number of factors rises, it may require very considerable skill in analysis and interpretation to tease out any treatment effect.

Given the small number of SCI clinical trials completed to date, identifying important covariates is not yet an exact science (cf Tuszynski *et al*¹⁶). We have reanalyzed some of the GM-1 trial data and found that

baseline AIS is a very strong covariate as is the level of injury (eg cervical or thoraco-lumbar). Certain types of spinal injuries (eg a suspected central cord or conus injury or one not involving a fracture dislocation) have a prognostic value (usually for a significant spontaneous functional recovery). Younger patients with incomplete injuries recover better than older ones; but younger patients tend to be more severely injured so that, on the whole, their recovery is no better. Other possibilities (use of spinal surgery or direct admission to tertiary care) did not have a readily detectable effect in the GM-1 study.^{7,8}

The ICCP Clinical Guidelines Panel is continuing to examine the raw data from previous SCI trials to determine if a valid therapeutic threshold for ASIA motor scores can be established for different levels and severities of SCI.

Zone of partial preservation

Below the most caudal 'functional' ASIA motor level (ASIA motor grade of 3, 4, or 5) the ZPP consists of those myotomes and dermatomes that remain partially innervated (Table 1), but at a level that may not be functionally meaningful (eg ASIA motor grade of 1 or 2). The exact numbers of segments, so affected, make up the ZPP. The term is used only when there is a motor-complete spinal injury. As outlined in the preceding article,¹¹ it is often difficult to discern the mechanism underlying any neurological or functional improvement when it occurs within the ZPP; it could be due to central repair (plasticity, sprouting, or regeneration) and/or due to similar peripheral modifications, such as peripheral sprouting, as some muscles are innervated from multiple spinal segments.¹⁷

There is little doubt that improved recovery of function within the ZPP can provide new and meaningful capabilities for a person with SCI, especially those individuals with a cervical level injury. All the same, the ZPP can also complicate the accurate interpretation of therapeutic action because the extent of recovery within the ZPP can be variable. Spontaneous changes within the ZPP introduce 'background noise' into the determination of therapeutic efficacy. There was general agreement that functional changes within the ZPP need to be interpreted with caution.¹⁸ Any improvement in function ascribed to an experimental intervention that is confined to the first two segments caudal to the last functional ASIA motor level may be due to plastic changes within the ZPP rather than to the formation of new spinal connections across the level of injury. Furthermore, there was recognition that in many previous therapeutic studies clear chronological description of ZPP function has been lacking; future trials should make provision to clearly describe changes in segments adjacent to the level of spinal injury.

ASIA sensory score

The lack of sophistication of the ASIA sensory score for accurately describing preserved sensory levels after SCI

or as a valid outcome measure has long been recognized. The ordinal 3-point scale for light touch (normal, abnormal, or absent) is highly variable at different assessment times and between ASIA assessors. The ASIA pin-prick score appears to be the more useful clinical measure of preserved spinal sensory function (eg sacral sparing in people with an ASIA B classification), as well as a predictor for future recovery.^{19,20} The ASIA light touch score does not necessarily correlate with subsequent sensory functions accurately and does not seem to be particularly useful as an SCI clinical trial outcome measure.

Quantitative sensory testing

Quantitative sensory testing (QST) is emerging as a potential adjunct to the neurological exam in the evaluation of sensory dysfunction after SCI.^{21–23} Commonly, QST has used quantitatively controlled thermal (warm and cool), mechanical (monofilaments/von Frey hairs) and vibratory stimuli (eg 100 Hz) with psychophysical scaling against established normative values, to differentiate the contributions from small and large diameter peripheral sensory afferent projections or distinguish the contributions of ascending spinal sensory pathways (spinothalamic and dorsal columns, respectively). QST measures appear to correlate with somatosensory-evoked potential (SSEP) recordings and with ASIA sensory scores.

Although further validations of QST techniques are required, QST appears to be a more sensitive technique than the ASIA sensory score, but it is a time-consuming evaluation. With repeated measures, QST might be considered as a secondary outcome measure of spinal cord function. Nevertheless, the Panel currently has more confidence in the sensitivity, accuracy, reliability, and reproducibility of motor function tests than in QST, primarily because QST can be a lengthy procedure with a number of highly variable stimulation parameters. A recent simple adjunct for the sensory evaluation of SCI, which overcomes some of the complexities of the QST, is the electrical perceptual threshold (EPT) test.²⁴ EPT supplies a measure of sensory threshold for each dermatome and provides a more quantitative map of the level and completeness of SCI, including the ZPP.^{22,25,26}

Electrophysiological assessments

Electrophysiological measurements such as SSEP, electromyographic (EMG), and motor-evoked potential (MEP) recordings provide objective data (latencies and amplitudes) for assessing spinal conductivity that can be analyzed by a blinded investigator in the form of truly quantitative values, in contrast to measures such as the ASIA scores that are a nonlinear ordinal scale.^{27–31} Furthermore, electrophysiological recordings have the advantage that they can be performed on comatose or otherwise unresponsive subjects. EMG recordings are useful in the assessment of function, both in response

to voluntary effort or when combined with electrical or magnetic stimulation of peripheral nerves (reflexes) or motor cortex (ie MEP).

Complementary to the neurological assessment, a combination of SSEP, MEP and/or EMG measurements provides information about spinal cord function that is not retrievable by other clinical means and may have additional value in predicting functional outcomes.^{32,33} Changes in conduction velocity and the magnitude of the compound action potentials, as an outcome measure, must be interpreted with caution. An increased conduction velocity may accurately reflect a remyelination of fiber tracts, which could be the targeted aim of a SCI trial, but in itself, may not herald the recovery of function or improvement in neurological condition.^{34,35} Strong correlations between AIS scores and electrophysiological measurements are not always evident.³⁶ In general, the Panel felt that electrophysiological measures were most useful when combined with other outcome tools and could be useful in determining the mechanism of therapeutic action.^{37,38}

Assessment of thoracic cord function

Currently, there are no agreed methods for assessing motor levels in the thoracic cord, although sensory levels are assessed during the standard ASIA examination. This is a significant problem for determining the potential efficacy of an intervention, given the expectation that it is safer to perform initial human studies in patients with a thoracic level injury. The electrophysiological studies described in recent papers^{22,39} provide methods aimed at detecting changes in motor and autonomic function, as well as providing information on the level and completeness of injury to the thoracic cord. Motor assessments have been developed using transcranial magnetic stimulation to elicit MEPs in paraspinal, intercostal, and abdominal muscles. Quantitative measures that appear to be promising include: thresholds, latencies, and recruitment (input/output curves) of MEPs from trunk muscles innervated at different thoracic levels.⁴⁰ Mechanically evoked reflexes, recorded as EMGs in paraspinal muscles, also show abnormalities directly related to the level of spinal injury.⁴⁰

In summary, these tests may be used to indicate functional improvement or deterioration following treatment. However, the innervation of trunk muscles by multiple thoracic spinal levels means that the resolution of these motor techniques is not as precise as might be achieved in the cervical cord. The tests may be used to indicate motor level within two or three levels (plus or minus).

Autonomic function testing

The accurate evaluation of impaired autonomic nervous system (ANS) function after SCI is currently limited. In addition to the motor and sensory deficits associated with SCI, coincident ANS impairments are common (cf Claydon *et al*⁴¹). Individuals with SCI often exhibit

autonomic dysreflexia, which results in episodes of uncontrolled hypertension. The recognition and management of cardiovascular dysfunctions following SCI represent challenging clinical issues, as well as important therapeutic targets since cardiovascular disorders in the acute and chronic stages of SCI are the cause of death in individuals with SCI.^{42,43}

As sympathetic vasomotor control is disrupted below the level of a complete sensorimotor SCI lesion, reflex vasodilatation owing to local heating of the skin in people with chronic SCI is diminished.⁴⁴ Thus, it has been suggested that assessment of reflex vasodilatation may be a useful noninvasive outcome measure to detect the preservation of any central autonomic pathways after SCI and possibly to document any change in spinal autonomic functions after a therapeutic intervention.^{23,45}

Tracking standard vital signs is imperative throughout the entire phase of any clinical trial, especially as the influence of the ANS on any of these measured functions is well established. Interestingly, measurement of the sympathetic skin response (SSR) has been suggested to delineate the level and extent of spinal sympathetic function, as a measure of autonomic dysfunction.^{22,32,41,46} It may reveal an incomplete lesion in terms of autonomic function in cases of complete motor and sensory injury.⁴⁷ However, SSR remains a controversial measure²² of overall spinal function and, if adopted as an outcome measure, should be limited to testing the efficacy of an intervention on ANS function and used in conjunction with a number of other outcome measures. Further development of valid outcome tools for the assessment of ANS function after SCI is imperative.

Imaging assessments

Magnetic resonance imaging (MRI) has become a cornerstone of radiologic technique to detect the location (and to some degree the severity) of an acute SCI, as well as to detect possible complications arising during chronic SCI, such as syringomyelia. At present, MRI along with computerized axial tomography and X-ray images are useful diagnostic tools and potentially helpful for screening participants to be included or excluded from a clinical trial.

MRI has been useful in determining the extent of cord compression,^{48–50} outlining hemorrhages and edema after human spinal injury and in the near future, might be useful in monitoring progressive changes in spinal cord tracts, such as demyelination after spinal injury. Indeed, recent data from the Spine Trauma Study Group, indicates that the extent of cord compression and the presence of hemorrhage and cord swelling are highly predictive of ASIA motor score outcomes at one, 1 year post-SCI.⁵⁰

MRI has also been proposed as a potential SCI assessment tool after a therapeutic intervention, and as a means of tracking implanted cells. In experimental models of SCI, diffusion tensor imaging (DTI) can

delineate both disrupted and intact axonal fiber tracts within the spinal cord, as well as the orientation of glial scarring surrounding a spinal lesion.⁵¹ With further development, MR technologies may develop a useful early 'surrogate' end point measure that would accurately predict the long-term functional benefits of an experimental intervention after SCI (cf Schwartz *et al*⁵¹).

Nevertheless, MRI is still largely a qualitative measure and quantitative standards, in relation to functionally measured SCI outcomes, will need to be developed and validated before MRI can be used as an outcome tool (cf Miller⁵²). It is hoped that MRI and Magnetic Resonance Spectroscopy technologies will rapidly mature, with more sophisticated algorithms (including DTI and functional MRI), such that imaging will become a valuable non-invasive assessment tool.

Functional tests

General considerations

For chronic SCI studies (greater than 12 months after initial SCI), ASIA assessments may not be a sufficient tool as an outcome measure, especially for studies on incomplete SCI where the ASIA motor score is likely to be substantial and highly variable between individuals. Nevertheless, an ASIA assessment, before randomization, is valuable for classifying and stratifying participants in a clinical trial. At acute and sub-acute stages after SCI, the value of functional outcome tools is less clear, especially for motor-complete SCI (ASIA A and ASIA B), which are likely to be the initial subjects in early Phase trials. If the expected therapeutic benefit is modest, a dramatic improvement in functional performance may not be readily evident. Nevertheless, functional outcome assessments should be undertaken as a secondary outcome measure.

There was agreement from the ICCP Clinical Guidelines Panel that an improvement in the measurable performance of meaningful function is necessary for any therapeutic intervention to be universally accepted as beneficial (for a review, see Ditunno *et al*⁵³). The World Health Organisation (WHO), specifically the International Classification of Functioning, Disability and Health (or ICF), has rigorously defined function and impairment, as well as activities of life and disability (see below). ICF-1 is a health sphere of influence classification system that describes, among other things, body functions and structures, activities and participation. In short, reduced function in a body structure can result in difficulty executing an activity of daily living.

ICF complements the WHO's International Classification of Diseases (ICD; eg latest version is ICD-10) and is currently being reviewed for the next iteration, ICF-2. ICF is useful to understand and measure functional outcomes after SCI and all clinical researchers are encouraged to become familiar with these classifications and definitions (<http://www3.who.int/icf/>).

Lower limb function

For clinical trials involving people with motor-incomplete SCI (ASIA C and ASIA D), at acute, subacute, and chronic SCI stages, several validated tests of ambulatory performance have been developed, including the Walking Index for Spinal Cord Injury (WISCI) and a number of timed walking tests.^{54,55} WISCI is a 21-level hierarchical scale of walking based on physical assistance, need of braces and devices, with an ordinal range from 0 (unable to walk) to 20 (walking without assistance for at least 10 m). It is an example of a more sensitive and precise scale for rating a specific functional activity in people with incomplete SCI. WISCI is currently a valid outcome measure for strategies directed to improve ambulation by subjects with incomplete SCI.⁵⁴

Although the WISCI has been validated as a qualitative outcome measure for the assessment of standing and walking after incomplete SCI, the opinion of the ICCP Clinical Guidelines Panel is that a more accurate assessment may be provided by a combination of WISCI and some of the more quantitative timed walking tests. Such quantitative walking tests include the timed up and go, time taken for a 10-min walk test (10 MWT) or one of the many similar variants (25, 30 ft, 8 m) and the distance traversed during a 6-min walk test.⁵⁵ There may be some redundancy between tests like the 10 MWT and the 6-min walk and it may be pragmatically easier to undertake a short timed walk test as the more routine walking assessment, especially in trials that involve centers, which may not have adequate facilities for measurement of longer duration walks.

Upper limb function

The number of people surviving with a cervical level spinal injury has risen dramatically over the past few decades and cervical SCI now accounts for approximately 50% of all people living with a SCI. Thus, validating a functional outcome tool to assess arm and hand capacity after a cervical spinal injury was identified as a top priority by the Panel.

At the present time, there is a lack of agreement on what might be the most useful test of arm and hand function after SCI (for a review, see van Tuijl *et al*⁵⁶). Many of the scales developed have been deemed too insensitive to track small, but potentially meaningful functional gains. The majority of tests have been developed within the domains of stroke or hand surgery, but less often to describe the impairment and course of hand function recovery after SCI, particularly for acute tetraplegic patients. Many previous studies examined tetraplegics after functional reconstructive surgery of the upper limb or application of a hand neuroprosthesis and did not provide randomized control data.

It is generally accepted that the assessment of hand function has to include several components including: (1) proximal arm and trunk stabilization (reaching out), as well as placement of the arm and hand, (2) sensory

testing of at least two sensory qualities (touch sensation, vibration, temperature, two-point discrimination, proprioception), (3) manual muscle testing of intrinsic (small hand muscles) and extrinsic muscles (forearm) involved in hand control, (4) description of different grasp forms (like pulp and lateral pinch), and (5) the effect of tenodesis on hand function, specifically for opening and closing of fingers and the fist.

The Quadriplegia Index of Function (QIF) was developed in the 1980s⁵⁷ as a scale for evaluating 10 areas of self-care and mobility for people living with tetraplegia. The QIF has been noted to be a better indicator of motor recovery than the FIM (when compared with ASIA motor scores) and a more sensitive measure of small gains in arm function.^{58,59}

One of the more established hand function assessment tools is the Sollerman test⁶⁰ although the test was not developed for SCI. The Sollerman test has limited resolution for hand function in tetraplegics, requires specialized equipment, and is a long duration examination (60–90 min). Another common test is the Manual Muscle Test, which has been used to evaluate handgrip strength, although it has been criticized as not sensitive enough to distinguish small or moderate changes in human subjects.⁶¹

The Action Research Arm Test looks at different types of pinches and provides a qualitative scoring, but has been mainly applied in stroke patients. The Jebsen (Taylor test) is most frequently used in stroke and includes writing, lifting cans, simulated feeding, stacking checkers, and picking up paper clips and coins. However, it does not detect changes of intrinsic muscles and allows compensatory trunk and shoulder movements to accomplish any tasks.

Other upper limb outcome assessment tools have recently been introduced. As an example, there is the motor capacities scale (MCS).⁶² This scale was developed and tested in France with the participation of 52 motor-complete C5–C7 tetraplegics, although some had received restorative upper limb surgery. The MCS initially involved 36 items associated with activities of daily living (ADL), including: transfers, repositioning in a prone and seated position, use and control of either a manual or powered wheelchair, bilateral reaching to a predetermined target, and bilateral hand grasping. High inter-rater reliability (correlation coefficient of 0.99) was noted for the MCS, as was a high correlation with the Sollerman test (correlation coefficient of 0.96). Initial correlation with ASIA motor scores was lower (0.74). Because of redundancies, this list has now been reduced to 31 items associated with ADL. The MCS is undergoing further testing and validation.

A Toronto group recently developed the Tetraplegia Hand Measure, which combines a modified Sollerman test with quantitative assessments of sensory function. A Zurich group developed a hand function test, which also uses certain key elements of the Sollerman test. An initiative is now underway across Canada, the United States, and Europe to develop an integrated hand function test as a valid assessment tool for SCI clinical trials.

Comprehensive functional outcome tools

The FIM was first developed in the 1980s (cf Stineman *et al*⁶³). The FIM is a proprietary global disability outcome assessment tool, which has been used for rating the functional performance of individuals, with a variety of different disorders and disabilities, on a series of ADL. It has been used as the functional outcome measure in many trials, such as the NASCIS III clinical trial.⁶ Because of its application to a broad range of disabilities, it has become a standard tool for decisions on support and reimbursement as a person re-integrates back into their home community (ie it has been called a ‘burden of care’ tool). For the purposes of SCI clinical trials, some of the FIM subsections are not directly relevant to people living with SCI (eg communication and social cognition) and FIM scores, and ASIA motor scores are not tightly correlated.^{58,59}

A more recently developed functional measure is the Spinal Cord Independence Measure (SCIM) and it appears to be a more sensitive and accurate functional assessment for ADL after SCI. SCIM has now gone through a few iterations^{64–66} and is undergoing further refinement in multinational studies. The SCIM is a 100-point disability scale developed specifically for SCI with emphasis on 18 activities associated with:

1. self-care (feeding, bathing, dressing, grooming), max. = 20 points
2. respiration and sphincter management (ventilation, bladder, bowel, use of toilet), max. = 40 points (clinically weighted)
3. mobility (in bed, transfers, indoors and outdoors, wheelchair, walking), max. = 40 points.

Preliminary findings suggest that the SCIM may be a more relevant and a useful outcome tool for SCI clinical trials than the FIM. However, the well-established nature of FIM may slow the adoption of SCIM. It may be too much to expect that one comprehensive functional outcome tool will accurately and sensitively track all SCI clinically meaningful benefits after a therapeutic intervention; a number of functional outcome measures may be required initially.

QoL surveys

QoL assessments for people with SCI have been intensely debated as clinical trial endpoint tools (cf Dijkers *et al*⁶⁷). The inclusion of a QoL assessment is often recommended as one outcome measure to be included in any clinical trial assessment, though often as a secondary outcome. WHO defines QoL as a person’s perception of his position in life within the context of the culture and value systems in which he lives and in relation to his goals, standards, and concerns. As outlined above, WHO published the ICF in 2002 with three distinctive dimensions:

1. body structure and function/impairment at organ level

2. activity/activity limitation at personal level
3. participation/restriction at societal level.

Several QoL surveys have been developed, along two paths, and are illustrated by the two following examples:

1. SF-36 (Medical Outcomes Study 36-item Short Form health survey) is a profile where the investigator determines the domains of life that are pertinent and the assumption is that the same domains are important to all people in that group. SF-36 reflects the perspective and choices made by the 'outsider' (investigator) rather than the subjective point of view of the 'insider' (subject).
2. SWLS (satisfaction with life survey) is an example of an alternate self-reported appraisal, where statements (eg I am satisfied with my life) are rated on a 7-point Likert-type scale (ranging from 'strongly disagree' to 'strongly agree'). SWLS is an example of a more global QoL where the individual (insider) is allowed to either adjust the weighting of a domain or in some cases self-nominate a domain as to its relative importance on their QoL. This can make comparisons between subjects or between study arms difficult.

Thus, QoL tools are either investigator-determined (eg SF-36), enabling statistical comparisons between an experimental and control group or they are more individualized (eg SWLS), allowing the participating subject to weigh the value (importance) of any individual field in the self-assessment of their own QoL.

In terms of SCI clinical trials of pharmaceutical drugs or cell-based transplants, especially during Phase 1 and 2, the former type of QoL survey (eg SF-36) is not suitable as a primary outcome measure, and should only be used in combination with other types of outcome data (eg ASIA motor scores or a functional outcome measure). Which precise QoL survey is best suited to a specific SCI trial has not been determined. It may be advisable to use more than one type of assessment.

The concern of the Panel was that any choice made by a subject during a QoL survey might accurately relate to a change in QoL, but be unrelated to an observable change in neurological impairment or functional capacity. Likewise, a small but significant improvement in neurological function might not influence the responses on a QoL survey. The consensus of the Panel was that changes in neurological function or functional outcomes should be used as the primary measure for Phase 1 or 2 SCI clinical trials that evaluate the activity of a pharmaceutical or cell-based transplant intervention.

Spasticity

A velocity-dependent, abnormal increase in muscle tone with exaggerated tendon jerks is one definition of spasticity,⁶⁸ which is a common complication of SCI and a variety of other CNS disorders.^{69,70} Spasticity can lead to incoordination of muscle action, reduced

functional limb movement and in its more severe forms may result in chronic pain, muscle contracture, and permanent muscle shortening. Several treatments have been developed to minimize spastic symptoms, including systemic or intrathecal Lioresal (Baclofen) and (more recently) the direct intramuscular injections of Botulinum toxin (Botox) into specific affected muscles.

The level of spasticity is known to vary over time, thus a single clinical assessment will not necessarily reflect accurately an individual's overall level of spasticity. The principal clinical outcome measure for spasticity has been the long-established Ashworth Scale or the modified Ashworth Scale, even though both scales have less than ideal inter-rater reliability⁷¹ and have a poor correlation with self-rated assessments of spasticity.⁷⁰ The scale determines the amount of resistance felt during the passive displacement of a limb, but it does not accurately account for the dependence of the resistance to the velocity of the stretch, which can be highly variable from examiner to examiner.

Pain

It has been suggested that over 50% of people living with SCI reported experiences of chronic neuropathic pain. Agreement on classifying pain (as musculoskeletal, neuropathic, or visceral forms) after SCI has been elusive, but the classification of Siddall *et al*⁷² has been widely quoted. Sharp, stabbing, or burning pain within the dermatomes at or just above the level of SCI is often termed at-level neuropathic pain, whereas similar types of pain below the level of the lesion have been called below-level neuropathic pain.

There are a few RCTs that have evaluated the benefits of gabapentin⁷³ and lidocaine⁷⁴ for the treatment of neuropathic pain after SCI (for a review, see Finnerup and Jensen⁷⁵). Nevertheless, causing pain as a result of an experimental treatment is also a major concern, especially as some of the emerging therapeutics have the potential to stimulate axonal fiber outgrowth or functional plasticity within central pain pathways. Thus, the Panel felt that inclusion of specific pain measures would be an important component of SCI therapeutics' outcome testing. The most straightforward assessment would rely on patient's self-reports of any increased pain during treatment. Several tools have been developed, including the visual analogue scale⁷⁶ and the neuropathic pain scale.⁷⁷ Nevertheless, these may not always provide an accurate reflection of neuropathic pain, especially as an individual's emotional health and/or social interactions can modify pain perception.

In an acute or subacute situation, the source of an individual's pain may be difficult to locate or originate outside the CNS pain sphere (eg result from concomitant injury to another body tissue or due to a preceding condition). Clinical trials may want to consider a more direct measure for a change in central pain threshold. For example, components of the QST and/or EPT may be useful evaluations (cf Savic *et al*²⁵, Savic *et al*²⁶).

There are many pain perception surveys available, including the well-known McGill pain questionnaire.⁷⁸ However, which pain assessment is the most accurate and easiest to use is a matter of debate. In more chronic SCI situations, pain management is an important clinical goal. One approach to mapping whether a pain management strategy is having a meaningful benefit is to assess how pain intensity interferes with ADL. Two common measurement scales of pain interference, the graded chronic pain (GCP) disability scale and three versions of the brief pain inventory (BPI), have recently been examined for their reliability and validity as pain assessment tools in a survey of 127 people living with chronic pain after SCI.⁷⁹ The self-report data asked questions on how pain interfered with ADL. Needless to say, increasing pain intensity caused increased interference with ADL. Both GCP and the three different length versions of the BPI were found to be internally consistent and related to the reported level of pain experienced.

Another issue is to carefully distinguish between neuropathic and normal musculoskeletal pain. A therapy that restores some normal pain sensation may make a patient aware of conditions that were previously unfamiliar to the spinal injured individual, such as lower-back pain or other forms of normal, internally referenced visceral pain.

Summary and recommendations for the future

Objective outcome measures are critical in designing useful SCI therapeutic clinical trials. Different clinical targets (eg sensorimotor tasks, autonomic function, personal functional capacity, performance, or community participation) normally require distinct and appropriate outcome assessment tools, which have been validated as both sensitive and accurate.

The most common outcome assessment tools currently being employed are the ASIA impairment grades and ASIA motor scores. The accuracy of initial and subsequent ASIA examinations is essential to ascribing a therapeutic benefit in neurological recovery. For example, a candidate drug or cell transplant with a very large effect size might rely on statistically significant differences in ASIA grades between the experimental and control arms of an SCI study. However, an intervention with a potentially smaller effect size might target a more specific and sensitive neurological outcome measure, such as a statistically significant difference between experimental and control groups for the ASIA motor score.

Establishing valid treatment effect thresholds for ASIA motor scores requires calculation of the spontaneous improvement of ASIA motor scores for each severity and level of SCI within 'untreated' control populations. Such an initial evaluation is now being undertaken by the ICCP Clinical Guidelines Panel. Nevertheless, any first table of ASIA motor score thresholds will require ongoing monitoring and updating to maintain relevance.

Valid and clinically meaningful sensory assessment tools for SCI remain a challenge where current assessment tools are either inadequate or insufficiently validated. Electrophysiological assessment tools exist and would benefit from broader application and standardization. Such evaluations are currently underway. Likewise, there is a need to develop a number of clinically valid autonomic function tests.

An improvement in the measurable performance of a meaningful function or behavior is necessary for any therapeutic intervention to be universally accepted as clinically beneficial. Thus, accurate and sensitive functional outcome measures are critical to SCI clinical trials and this will be especially true for any Phase 3 studies. The FIM scale is not specific to SCI and not suitable, although the recently developed SCIM assessment may be a more specific and accurate outcome tool for detecting clinical end points in SCI. The continued development and validation of tests that quantify highly relevant behaviors such as walking or hand function are most important; such tools may have greater utility for documenting the subtle benefit of a therapeutic than a more global scale of disability.

The inclusion of QoL measures in SCI trials is important, but which precise QoL survey is best suited to a specific SCI trial and their importance in the overall assessment of an intervention has not been determined. It may be best to use more than one type of assessment. The concern of the Panel was that any choice made by a subject during a QoL survey might be unrelated to an observable change in neurological or functional outcome. Likewise, a small, but significant, improvement in neurological function might not influence the responses on a QoL survey, which are often governed by attitude and social integration and not by physical disability.

Given the paucity of Phase 3 SCI clinical trial experiences and thus the emerging nature of SCI clinical studies, the current opinion of the Panel was that changes in neurological function or functional outcomes should be used as the primary measure for Phase 1 or 2 SCI clinical trials designed to evaluate the safety and/or provide evidence of activity of a pharmaceutical or cell-based transplant intervention. Neurological function tests should remain an element of the outcome assessment in Phase 3 trials.

Glossary of definitions

(Additional glossaries are included in the three accompanying papers)

Neurological level of spinal injury is generally the lowest segment of the spinal cord with normal sensory and motor function on both sides of the body. However, the spinal level at which normal function is found often differs on each side of the body, as well as in terms of preserved sensory and motor function. Thus, up to four different segments may be identified in determining the neurological level and each of these segments is recorded separately and a single-level descriptor is not used. Note

that the level of spinal column injury may not correlate with the neurological level of SCI.

ASIA (American Spinal Injury Association) Impairment Scale (or AIS) describes the completeness of a spinal injury (see Marino *et al*³). An individual with an ASIA A grade has no motor or sensory function at the level of S4–S5 sacral segments. ASIA B has some sensory function below the neurological level, including S4–S5, but not motor function. ASIA C has some motor function below the neurological level, but more than half of the key muscles involved have a muscle strength score that is less than 3 (Table 1). ASIA D has motor function below the neurological level but more than half of the key muscles have a muscle grade of 3 or more. ASIA E indicates normal motor and sensory function.

Tetraplegia (quadriplegia) is the term used to refer to loss of motor and/or sensory function owing to damage to the spinal cord, with impairment of the upper extremities as well as trunk, legs, and pelvic organs. This implies damage to the spinal cord at or above the C8 level.

Paraplegia is the equivalent term used to refer to functional loss below the level of the upper extremities, which may involve loss of motor and/or sensory function within the trunk, and/or the lower extremities. This implies damage to the spinal cord below the level of C8 and may include damage to conus medullaris or cauda equine (ie neural tissue within the spinal canal).

Complete and incomplete SCI are other terms used to describe the overall severity of SCI. Technically, SCI is classified as complete if there is no motor or sensory function preservation in the sacral (most caudal) spinal segments. Thus, incomplete SCI is when there is some preserved motor or sensory function at the lowest sacral spinal level (S4–5). There can be extensive variability in the degree of preserved function after incomplete SCI.

ASIA Sensory and Motor Assessments form the basis for the International Standards for Neurological and Functional Classification of Spinal Cord Injury (the ASIA International Standards) and are conducted in the supine position and involve a qualitative grading of sensory responses to touch and pin-prick at each of 28 dermatomes along each side of the body and a qualitative grading of the strength of contraction within 10 representative (key) muscles, primarily identified with a specific spinal level, 5 for the upper extremity (C5–T1) and 5 for the lower extremity (L2–S1) on each side of the body (Table 1)

ASIA Motor Score is calculated by assigning to one muscle group, innervated and primarily identified with a specific spinal level, a score between 0 (no detectable contraction) and 5 (active movement and a full range of movement against maximum resistance). C5–T1 and L2–S1 are tested, giving 10 levels on each side of the body for a possible maximum score of 100.

LEMS is the lower extremity motor score which is a maximal 50-point subset of the ASIA motor score for the representative leg and foot muscles.

UEMS is the upper extremity motor score which is a maximal 50-point subset of the ASIA motor score for the representative arm and hand muscles.

Motor level is defined as the most caudal spinal level as indexed by the key muscle group for that level having a muscle strength of 3 or above while the key muscle for the spinal segment above is normal (= 5).

ASIA sensory score is calculated by testing a point on the dermatome for each spinal level from C2 to S4–5 for both light touch and pin-prick sensation. Each point is assigned a score from 0 (absent sensation) through 1 (abnormal sensation) to 2 (normal sensation). This gives a possible maximum score of 56 on each side for a maximum total of 112 each for light touch and pin-prick.

Sensory level is defined as the spinal segment corresponding with the most caudal dermatome having a normal score of 2/2 for both pin-prick and light touch.

Zone of partial preservation (ZPP) is only used when SCI is complete and refers to those segments below the neurological level of injury where there is some preservation of impaired motor or sensory function (usually, but not always, within a few segments of the neurological level).

Acknowledgements

We are grateful for the support of The International Campaign for Cures of spinal cord injury Paralysis (ICCP), which provided the funding for the authors' travel and accommodation expenses. The ICCP represents the following member organizations: Christopher Reeve Foundation (USA), Institut pour la Recherche sur la Moëlle Epinière (FRA), International Spinal Research Trust (UK), Japan Spinal Cord Foundation, Miami Project to Cure Paralysis (USA), Paralyzed Veterans of America (USA), Rick Hansen Man In Motion Foundation (CAN), SpinalCure Australia, and Spinal Research Fund of Australia. We thank the European Multicenter study in Spinal Cord Injury (EM-SCI) for sharing their data on spontaneous recovery after spinal cord injury. ICORD (International Collaboration on Repair Discoveries) in Vancouver provided all logistical coordination and support. All panel members (authors) volunteered their time and effort. Finally, we are most grateful for the input and constructive comments from a countless number of SCI investigators over the past 2.5 years.

References

- 1 Steeves J, Fawcett J, Tuszynski M. Report of International Clinical Trials Workshop on spinal cord injury February 20–21, 2004, Vancouver, Canada. *Spinal Cord* 2004; **42**: 591–597.
- 2 Lammertse D *et al*. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the International Campaign for Cures of spinal cord Paralysis (ICCP) Panel: Clinical trial design. *Spinal Cord* 2006 [E-pub ahead of print: 19 December 2006; doi:10.1038/sj.sc.3102010].
- 3 Marino R *et al*. International standards for neurological classification of spinal cord injury (6th edn). *J Spinal Cord Med* 2003; **26**(Suppl 1): S49–S56.

- 4 Bracken MB *et al*. A randomized, controlled trial of methylprednisolone or naloxone in the treatment of acute spinal cord injury. Results of the second National Acute Spinal Cord Injury Study. *N Engl J Med* 1990; **322**: 1405–1411.
- 5 Bracken MB *et al*. Administration of methylprednisolone for 24 or 48 hours or tirilazad mesylate for 48 hours in the treatment of acute spinal cord injury. Results of the Third National Acute Spinal Cord Injury Randomized Controlled Trial. National Acute Spinal Cord Injury Study. *JAMA* 1997; **277**: 1597–1604.
- 6 Bracken MB *et al*. Methylprednisolone or tirilazad mesylate administration after acute spinal cord injury: 1-year follow-up. Results of the third national acute spinal cord injury randomized controlled trial. *J Neurosurg* 1998; **89**: 699–706.
- 7 Geisler FH, Coleman WP, Grieco G, Poonian D, the Sygen[®] Study Group. Measurements and recovery patterns in a multicenter study of acute spinal cord injury. *Spine* 2001a; **26**: S68–S86.
- 8 Geisler FH, Coleman WP, Grieco G, Poonian D, the Sygen[®] Study Group. The Sygen[®] multicenter acute spinal cord injury study. *Spine* 2001b; **26**: S87–S98.
- 9 Blaustein DM, Zafonte RD, Thomas D, Herbison GJ, Ditunno Jr JF. Predicting recovery of motor complete quadriplegic patients: twenty-four-hour *versus* 72-h motor index scores. *Arch Phys Med Rehabil* 1991; **72**: 786.
- 10 Burns AS, Lee BS, Ditunno Jr JF, Tessler A. Patient selection for clinical trials: the reliability of the early spinal cord injury examination. *J Neurotrauma* 2003; **20**: 477–482.
- 11 Fawcett JW, Curt A, Steeves JD, Coleman WP, Tuszynski MH. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP Panel: Spontaneous recovery after spinal cord injury and statistical power needed for therapeutic clinical trials. *Spinal Cord* 2006 [E-pub ahead of print: 19 December 2006; doi:10.1038/sj.sc.3102007].
- 12 Marino RJ, Graves DE. Metric properties of the ASIA motor score: subscales improve correlation with functional activities. *Arch Phys Med Rehabil* 2004; **85**: 1804–1810.
- 13 Coleman WP, Geisler FH. Injury severity as a primary predictor of outcome in acute spinal cord injury: retrospective results from a large multicenter clinical trial. *Spine J* 2004; **4**: 373–378.
- 14 Waters RL, Adkins RH, Yakura JS, Sie I. Motor and sensory recovery following complete tetraplegia. *Arch Phys Med Rehabil* 1993; **74**: 242–247.
- 15 Marino RJ, Ditunno JF, Donovan WH, Maynard F. Neurologic recovery after traumatic spinal cord injury: data from the Model Spinal Cord Injury Systems. *Arch Phys Med Rehabil* 1999; **80**: 1391–1396.
- 16 Tuszynski MH, Steeves JD, Fawcett JW, Lammertse D, Kalichman M. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP Panel: Clinical trial inclusion/exclusion criteria and ethics. *Spinal Cord* 2006 [E-pub ahead of print: 19 December 2006; doi:10.1038/sj.sc.3102009].
- 17 Marino RJ, Herbison GF, Ditunno JF. Peripheral sprouting as a mechanism for recovery in the zone of injury in acute quadriplegia: a single-fiber EMG study. *Muscle Nerve* 1994; **17**: 1466–1468.
- 18 Dietz V, Curt A. Neurological aspects of spinal cord repair: promises and challenges. *Lancet Neurol* 2006; **5**: 688–694.
- 19 Crozier KS, Graziani V, Ditunno JF, Herbison GJ. Spinal cord injury: prognosis for ambulation based on sensory examination in patients who are initially motor complete. *Arch Phys Med Rehabil* 1991; **72**: 119–121.
- 20 Katoh S, el Masry WS. Motor recovery of patients with motor paralysis and sensory sparing following cervical spinal injuries. *Paraplegia* 1995; **30**: 506–509.
- 21 Hayes KC *et al*. Clinical and electrophysiological correlates of quantitative sensory testing in patients with incomplete spinal cord injury. *Arch Phys Med Rehabil* 2002; **83**: 1612.
- 22 Ellaway PH *et al*. Towards improved clinical and physiological assessments of recovery in spinal cord injury: a clinical initiative. *Spinal Cord* 2004; **42**: 325–337.
- 23 Nicotra A, Ellaway PH. Thermal perception thresholds: assessing the level of human spinal cord injury. *Spinal Cord* 2006; **44**: 617–624.
- 24 Davey NJ, Nowicky AV, Zaman R. Somatopy of perceptual threshold to cutaneous electrical stimulation in man. *Exp Physiol* 2001; **86**: 127–130.
- 25 Savic G *et al*. Quantitative sensory tests (perceptual thresholds) in patients with spinal cord injury. *J Rehab Res Dev* 2006 (in press).
- 26 Savic G, Bergstrom EMK, Frankel HL, Jamous MA, Ellaway PH, Davey NJ. Perceptual thresholds to cutaneous electrical stimulation in patients with spinal cord injury. *Spinal Cord* 2006; **44**: 560–566.
- 27 Curt A, Dietz V. Ambulatory capacity in spinal cord injury: Significance of somatosensory-evoked potentials and ASIA protocols in predicting outcome. *Arch Phys Med Rehabil* 1997; **78**: 39–43.
- 28 Curt A, Keck ME, Dietz V. Functional outcome following spinal cord injury: significance of motor-evoked potentials. *Arch Phys Med Rehabil* 1998; **79**: 81–86.
- 29 Davey NJ, Smith HC, Wells E, Maskill DW, Savic G, Ellaway P. Frankel HL responses of the thenar muscles to transcranial magnetic stimulation of the motor cortex in incomplete spinal cord injury patients. *J Neurol Neurosurg Psychiatry* 1998; **65**: 80–87.
- 30 Davey NJ, Smith HC, Savic G, Maskill DW, Ellaway PH, Frankel HL. Comparison of input-output patterns in the corticospinal system of normal subjects and incomplete spinal cord injured patients. *Exp Brain Res* 1999; **127**: 382–390.
- 31 Kirshblum S, Lim S, Garstang S, Millis S. Electrodiagnostic changes of the lower limbs in subjects with chronic complete cervical spinal cord injury. *Arch Phys Med Rehabil* 2001; **82**: 604–607.
- 32 Curt A, Dietz V. Electrophysiological recordings in patients with spinal cord injury: significance for predicting outcome. *Spinal Cord* 1999; **37**: 157–165.
- 33 Metz GA, Curt A, van de Meent H, Klusman I, Schwab ME, Dietz V. Validation of the weight-drop contusion model in rats: a comparative study of human spinal cord injury. *J Neurotrauma* 2000; **17**: 1–17.
- 34 Diehl P, Kliesch U, Dietz V, Curt A. Impaired facilitation of motor-evoked potentials in incomplete spinal cord injury. *J Neurol* 2006; **253**: 51–57.
- 35 Wolfe DL, Hayes KC, Hsieh JT, Potter PJ. Effects of 4-aminopyridine on motor-evoked potentials in patients with spinal cord injury: a double-blinded, placebo-controlled crossover trial. *J Neurotrauma* 2001; **18**: 757–771.
- 36 Smith HC *et al*. Corticospinal function studied over time following incomplete spinal cord injury. *Spinal Cord* 2000; **38**: 292–300.
- 37 Laubis-Herrmann U, Dichgans J, Bilow H, Topka H. Motor reorganization after spinal cord injury: evidence of

- adaptive changes in remote muscles. *Restor Neurol Neurosci* 2000; **17**: 175–181.
- 38 Thomas SL, Gorassini MA. Increases in corticospinal tract function by treadmill training after incomplete spinal cord injury. *J Neurophysiol* 2005; **94**: 2844–2855.
- 39 Curt A, Schwab ME, Dietz V. Providing the clinical basis for new interventional therapies: refined diagnosis and assessment of recovery after spinal cord injury. *Spinal Cord* 2004; **42**: 1–6.
- 40 Kuppuswamy A *et al*. Motoneurone excitability in back muscles assessed using mechanically evoked reflexes in spinal cord injured patients. *J Neurol Neurosurg Psychiatry* 2005; **76**: 1259–1263.
- 41 Claydon VE, Steeves JD, Krassioukov A. Orthostatic hypotension following spinal cord injury: understanding clinical pathophysiology. *Spinal Cord* 2006; **44**: 341–351.
- 42 Devivo MJ, Krause JS, Lammertse DP. Recent trends in mortality and causes of death among persons with spinal cord injury. *Arch Phys Med Rehabil* 1999; **80**: 1411–1419.
- 43 Garshick E *et al*. A prospective assessment of mortality in chronic spinal cord injury. *Spinal Cord* 2005; **43**: 408–416.
- 44 Nicotra A, Asahina M, Mathias CJ. Skin vasodilator response to local heating in human chronic spinal cord injury. *Eur J Neurol* 2004; **11**: 835–837.
- 45 Nicotra A, Young TM, ASahina M, Mathias CJ. The effect of different physiological stimuli on skin vasomotor reflexes above and below the lesion in human chronic spinal cord injury. *Neurorehabil Neural Repair* 2005b; **19**: 325–331.
- 46 Cariga P, Catley M, Mathias CJ, Savic G, Frankel HL, Ellaway PH. Organisation of the sympathetic skin response in spinal cord injury. *J Neurol Neurosurg Psychiatry* 2002; **72**: 356–360.
- 47 Nicotra A, Catley M, Ellaway PH, Mathias CJ. The ability of physiological stimuli to generate the sympathetic skin response in human chronic spinal cord injury. *Restor Neurol Neurosci* 2005a; **23**: 331–339.
- 48 Fehlings MG *et al*. The optimal radiologic method for assessing spinal canal compromise and cord compression in patients with cervical spinal cord injury Part II: results of a multicenter study. *Spine* 1999; **24**: 605–613.
- 49 Bono CM *et al*. Measurement techniques for lower cervical spine injuries: consensus statement of the Spine Trauma Study Group. *Spine* 2006; **31**: 603–609.
- 50 Miyajiri F, Furlan JC, Aarabi B, Arnold PM, Fehlings MG. Correlation of MRI findings with neurological outcome in patients with acute cervical traumatic spinal cord injury: a prospective study in 100 consecutive patients. *Radiology* 2006 (in press).
- 51 Schwartz ED, Duda J, Shumsky JS, Cooper ET, Gee J. Spinal cord diffusion tensor imaging and fiber tracking can identify white matter tract disruption and glial scar orientation following lateral funiculotomy. *J Neurotrauma* 2005; **22**: 1388–1398.
- 52 Miller DH. Biomarkers and surrogate outcomes in neurodegenerative disease: lessons from multiple sclerosis. *J Am Soc Exp NeuroTherapeutics* 2004; **1**: 284–294.
- 53 Ditunno JF, Burns AS, Marino RJ. 2005. Neurological and functional capacity outcome measures: essential to spinal cord injury clinical trials. *J Rehab Res Dev* 2005; **42**(Suppl 1): 35–41.
- 54 Morganti B, Scivoletto G, Ditunno P, Ditunno JF, Molinari M. Walking index for spinal cord injury (WISCI): criterion validation. *Spinal Cord* 2005; **43**: 43–71.
- 55 van Hedel HJ, Wirz M, Dietz V. Assessing walking ability in subjects with spinal cord injury: validity and reliability of 3 walking tests. *Arch Phys Med Rehabil* 2005; **86**: 190–196.
- 56 van Tuijl JH, Janssen-Potten YJ, Seele HA. Evaluation of upper extremity motor function tests in tetraplegics. *Spinal Cord* 2002; **40**: 51–64.
- 57 Gresham GE, Labi ML, Dittmar SS, Hicks JT, Joyce SZ, Stehlik MA. The Quadriplegia Index of Function (QIF): sensitivity and reliability demonstrated in a study of thirty quadriplegic patients. *Paraplegia* 1986; **24**: 3–44.
- 58 Marino RJ *et al*. 1993 Assessing self-care status in quadriplegia: comparison of the quadriplegia index of function (QIF) and the functional independence measure (FIM). *Paraplegia* 1991; **31**: 225–233.
- 59 Yavuz N, Tezyurek M, Akyuz M. A comparison of two functional tests in quadriplegia: The quadriplegia index of function and the functional independence measure. *Spinal Cord* 1998; **36**: 832–837.
- 60 Sollerman C, Ejeskar A. Sollerman hand function test. A standardized method and its use in tetraplegic patients. *Scan. J Plast Reconstr Surg Hand Surg* 1995; **29**: 167–176.
- 61 Noreau L, Vachon J. Comparison of three methods to assess muscular strength in individuals with spinal cord injury. *Spinal Cord* 1998; **36**: 716–723.
- 62 Fattal C. Motor capacities of upper limbs in tetraplegics: a new scale for the assessment of the results of functional surgery on upper limbs. *Spinal Cord* 2004; **42**: 80–90.
- 63 Stineman MG *et al*. *A Prototype Classification System for Medical Rehabilitation*. American Rehabilitation Association: Washington DC 1994.
- 64 Catz A, Itzkovich M, Agranov E, Ring H, Tamir A. SCIM–spinal cord independence measure: a new disability scale for patients with spinal cord lesions. *Spinal Cord* 1997; **35**: 850–856.
- 65 Itzkovich M *et al*. Reliability of the Catz-Itzkovich Spinal Cord Independence Measure assessment by interview and comparison with observation. *Am J Phys Med Rehabil* 2003; **82**: 267–272.
- 66 Catz A *et al*. A multi-center international study on the spinal cord independence measure, version III: Rasch psychometric validation. *Spinal Cord* 2006 [E-pub ahead of print: 15 August 2006; doi:10.1038/sj.sc.3101960].
- 67 Dijkers MP. Individualization in quality of life measurement: instruments and approaches. *Arch Phys Med Rehabil* 2003; **84**(Suppl 1): S3–S14.
- 68 Young RR. Spasticity: a review. *Neurology* 1994; **44**: 12–20.
- 69 Hobart JC *et al*. Getting the measure of spasticity in multiple sclerosis: the multiple sclerosis spasticity scale (MSSS-88). *Brain* 2006; **129**: 224–234.
- 70 Lechner HE, Frotzler A, Eser P. Relationship between self- and clinically rated spasticity in spinal cord injury. *Arch Phys Med Rehabil* 2006; **87**: 15–19.
- 71 Pandyan AD, Johnson GR, Price CI, Curless RH, Barnes MP, Rodgers H. A review of the properties and limitations of the Ashworth and Modified Ashworth Scales as measures of spasticity. *Clin Rehabil* 1999; **13**: 373–383.
- 72 Siddall PJ, Taylor DA, McClelland JM, Rutkowski SB, Cousins MJ. Pain report and the relationship of pain to physical factors in the first six months following spinal cord injury. *Pain* 1999; **81**: 187–197.
- 73 Levendoglu F, Ogun CO, Ozerbil O, Ogun TC, Ugurlu H. Gabapentin is the first line drug for the treatment of neuropathic pain in spinal cord injury. *Spine* 2004; **29**: 743–751.

- 74 Finnerup NB *et al*. Intravenous lidocaine relieves spinal cord injury pain. *Anesthesiol* 2005; **102**: 1023–1030.
- 75 Finnerup NB, Jensen TS. Spinal cord injury pain – mechanisms and treatment. *Eur J Neurol* 2004; **11**: 73–82.
- 76 Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults.. In: Turk DC, Melzack R (eds). *Handbook of Pain Assessment*. Guilford Press: New York, NY 1992, pp 152–168.
- 77 Bradley S, Galer BS, Jensen MP. Development and preliminary validation of a pain measure specific to neuropathic pain: the neuropathic pain scale. *Rehabil Med* 1997; **48**: 332–337.
- 78 Melzack R. The McGill pain questionnaire: from description to measurement. *Anesthesiol* 2005; **103**: 199–202.
- 79 Raichle KA, Osborne TL, Jensen MP, Cardenas D. The reliability and validity of pain interference measures in persons with spinal cord injury. *J Pain* 2006; **7**: 179–186.