

REVIEW

Challenges for defining minimal clinically important difference (MCID) after spinal cord injury

X Wu^{1,2}, J Liu¹, LG Tanadini^{3,4}, DP Lammertse⁵, AR Blight⁶, John LK Kramer^{1,7}, G Scivoletto⁸, L Jones⁹, S Kirshblum¹⁰, R Abel¹¹, J Fawcett¹², E Field-Fote¹³, J Guest¹³, B Levinson¹⁴, D Maier¹⁵, K Tansey¹⁶, N Weidner¹⁷, WG Tetzlaff¹, T Hothorn⁴, A Curt³ and JD Steeves¹

Study design: This is a review article.

Objectives: This study discusses the following: (1) concepts and constraints for the determination of minimal clinically important difference (MCID), (2) the contrasts between MCID and minimal detectable difference (MDD), (3) MCID within the different domains of International Classification of Functioning, disability and health, (4) the roles of clinical investigators and clinical participants in defining MCID and (5) the implementation of MCID in acute versus chronic spinal cord injury (SCI) studies.

Methods: The methods include narrative reviews of SCI outcomes, a 2-day meeting of the authors and statistical methods of analysis representing MDD.

Results: The data from SCI study outcomes are dependent on many elements, including the following: the level and severity of SCI, the heterogeneity within each study cohort, the therapeutic target, the nature of the therapy, any confounding influences or comorbidities, the assessment times relative to the date of injury, the outcome measurement instrument and the clinical end-point threshold used to determine a treatment effect. Even if statistically significant differences can be established, this finding does not guarantee that the experimental therapeutic provides a person living with SCI an improved capacity for functional independence and/or an increased quality of life. The MDD statistical concept describes the smallest real change in the specified outcome, beyond measurement error, and it should not be confused with the minimum threshold for demonstrating a clinical benefit or MCID. Unfortunately, MCID and MDD are not uncomplicated estimations; nevertheless, any MCID should exceed the expected MDD plus any probable spontaneous recovery.

Conclusion: Estimation of an MCID for SCI remains elusive. In the interim, if the target of a therapeutic is the injured spinal cord, it is most desirable that any improvement in neurological status be correlated with a functional (meaningful) benefit.

Spinal Cord (2015) **53**, 84–91; doi:10.1038/sc.2014.232; published online 16 December 2014

INTRODUCTION TO OUTCOME MEASUREMENTS IN SCI CLINICAL TRIALS

Statistical significance may not equal clinical significance

Currently, no valid pivotal phase III clinical trial has provided an effective pharmacological or cell transplant treatment for improving neurological or functional recovery in spinal cord injury (SCI), and thus there are no ‘gold standard’ treatments, with demonstrated clinical benefit, to serve as possible comparators in trials of new interventions. Thus, we often rely on comparisons against study end points statistically estimated from historical data sets tracking spontaneous recovery patterns after SCI.^{1–5} However, there are a number of emerging neural protective, neural repair and neuroplasticity strategies, which require sensitive and accurate assessment for the reliable interpretation of their clinical importance.^{6–8}

Various outcome instruments, such as elements of the international standards for neurological classification of spinal cord injury (ISNCSCI), spinal cord independence measure (SCIM), the functional independence measure and short form 36 health survey, have been suggested as potential tools to measure aspects of recovery.^{2–4,9–15} Obviously, the data from such study valuations are dependent on the level and severity of SCI, the heterogeneity within each study cohort, the therapeutic target, the nature of the therapy, any confounding influences or comorbidities, the assessment times relative to the date of injury, the outcome measurement instrument and the clinical end-point threshold used to assess any change in tissue activity, functional behaviour or quality of life (QoL).

Even if statistically significant differences can be established, this finding does not guarantee that the experimental therapeutic provides

¹ICORD, Blusson Spinal Cord Centre, Vancouver General Hospital, University of British Columbia, Vancouver, British Columbia, Canada; ²Department of Spinal Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China; ³Spinal Cord Injury Center, University Hospital Balgrist, University of Zurich, Zurich, Switzerland; ⁴Epidemiology, Biostatistics and Prevention Institute, University Zurich, Zurich, Switzerland; ⁵Craig Hospital, Englewood, CO, USA; ⁶Acorda Therapeutics Inc. Hawthorne, NY, USA; ⁷Shepherd Center, Crawford Research Institute, Atlanta, GA, USA; ⁸Spinal Unit, IRCCS Fondazione S. Lucia, Rome, Italy; ⁹Craig H. Neilsen Foundation, Encino, CA, USA; ¹⁰Kessler Institute for Rehabilitation, West Orange, NJ, USA; ¹¹Trauma Center Bayreuth, Bayreuth, Germany; ¹²John van Geest Centre for Brain Repair, University of Cambridge, Cambridge, UK; ¹³Miami Project to Cure Paralysis, University of Miami Miller School of Medicine, Miami, FL, USA; ¹⁴Asubio Pharmaceuticals Inc. Edison, NJ, USA; ¹⁵Trauma Center Murnau, Murnau, Germany; ¹⁶Departments of Neurology and Physiology, Emory University and VA Medical Center, Atlanta, GA, USA and ¹⁷Spinal Cord Injury Center, Heidelberg University Hospital, Heidelberg, Germany

Correspondence: Professor JD Steeves, ICORD, Blusson Spinal Cord Centre, Vancouver General Hospital, University of British Columbia, 818 West 10th Avenue, Vancouver, British Columbia, Canada V5Z 1M9.

E-mail: steeves@icord.org

Received 11 August 2014; revised 16 November 2014; accepted 19 November 2014; published online 16 December 2014

a person living with SCI an improved capacity for functional independence and/or an increased QoL. After completing all phases of a clinical trial programme, registration of any new intervention by a regulatory agency as an approved treatment for a disorder normally requires demonstration of both statistical significance and clinical (that is, functional) benefit for experimental participants. The minimum threshold for demonstrating such a clinical benefit is sometimes referred to as a minimal clinically important difference (MCID), which is not to be confused with the statistical concept called minimum detectable difference (MDD) that determines the smallest real change in outcome, beyond measurement error. More importantly, MCID and MDD can have different meanings in disparate contexts or from differing perspectives.

Heterogeneity of SCI challenges measurement of treatment outcomes and any estimation of MCID

During recovery after SCI, neurological and functional characteristics are not constant across all levels and severities of injury. For example, after a fifth cervical (C5) sensorimotor complete SCI, an improvement of 10 ISNCSCI motor points within the cervical segments may be highly significant and clinically meaningful, whereas the same number of motor points recovered and distributed equally across all caudal muscles of the upper and lower extremities may not correlate with meaningful functional recovery. Thus, a meaningful clinical benefit or MCID is not addressed by statistical differences in neurological outcomes alone. If the biological target of a therapeutic is the injured spinal cord, it is most desirable that any improvement in neurological status be correlated with a functional benefit. Along with improvements in neurological impairment and advances in independent activities of daily living (ADL), as outlined in functional assessments, increased QoL, as described by patient-reported outcomes (PROs), has also been suggested as an important goal for study participants, as well as a possible clinical end point.

Currently, investigators recognise the heterogeneity of SCI and the variability in the rates and extent of recovery across differing levels and severity of SCI. It is widely accepted that there are also an increasing number of uncontrolled independent variables that can influence the

interpretation of study results as assessments move across the domains of the International Classification of Functioning, disability and health (or ICF) from body structure/function, through activity measures, to participation outcomes (Tables 1 and 2). Thus, the SCI field faces a dilemma until some novel therapeutic is validated as a new standard of treatment and can then be used as a comparator in future studies.

In this paper, we discuss the following: (1) concepts and constraints for the determination of MCID, (2) compare and contrast MCID with the statistical concept of MDD, (3) evaluate the different domains influencing MCID, (4) examine the roles of clinical investigators in defining MCIDs and (5) the implementation of MCID in acute versus chronic studies.

DEFINING MCID

In 1989, Jaeschke *et al.*¹⁶ introduced the concept of MCID to determine whether a difference in treatment effect between experimental and control participants are judged to be of value to people living with the disorder. They defined the MCID 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’.¹⁶ By this definition, MCID is only defined by the minimal amount of change (treatment effect) that is important to people living with SCI; the experienced wisdom of clinical investigators is excluded.

The ability to define a stable, universal MCID threshold value for a particular instrument is an attractive and useful concept, especially for chronic studies (for example, chronic pain or cancer) in which people living with a disease have developed an awareness of what improvements provide personal benefit (that is, increased independence in ADLs and improved QoL). Yet, it is not clear how MCIDs should be defined for each SCI outcome measure and how an acute trial MCID might differ from an MCID for a chronic study.

As mentioned above, MCID is defined as the smallest difference that people living with SCI perceive as beneficial. Several arguments suggest that this concept has merit. First, it ties the magnitude of an observed change to subsequent treatment decisions in clinical practice.

Table 1 Selected outcome measurement tools for SCI

<i>Outcome tools for different ICF domains</i>		
<i>Body structure and function</i>	<i>Activity</i>	<i>Participation</i>
ISNCSCI scores (for example, UEMS and LEMS points and motor levels)	10-m walk test(10MWT) 6-min walk test (6MWT)	Craig handicap and reporting technique (CHART) Perceived handicap questionnaire (PHQ)
Quantitative sensory testing (QST)	Timed up and go (TUG)	Participation survey for mobility-limited people (PARTS/M)
Contact heat-evoked potentials (CHEPs)	Spinal cord injury measure (SCIM)	Quality of Well-Being scale (QWB)
Sensory and motor-evoked potentials (SEPs and MEPs)	Functional independence measure (FIM)	Qualiveen
Autonomic function testing (for example, orthostatic tolerance using blood pressure for cervical and T2–T5 only)	Graded and redefined assessment of strength, sensibility and prehension (GRASSP)	Short Form 36
Imaging assessments (MRI, CT)	Walking Index for Spinal Cord Injury (WISCI)	Short Form 12
Modified ashworth scale (for spasticity)	Quadriplegia Index of Function (QIF)	Quality-of-Life Index (QLI)
	Modified Barthel Index (MBI)	Quality-of-Life Profile for adults with physical Disabilities (QOLP-(D)
		Satisfaction With Life Scale (SWLS)
		Sense of Well-Being Index (SWBI)
		World Health Organisation Quality-of-Life-BREF scale (WHOQOL-BREF)

Abbreviations: CT, computed tomography; ICF, International Classification of Functioning, disability and health; ISNCSCI, international standards for neurological classification of spinal cord injury; LEMS, lower extremity motor score; MRI, magnetic resonance imaging; SCI, spinal cord injury; UEMS, upper extremity motor score. Also view <http://www.rehabmeasures.org> and <http://www.scireproject.com> for further details of tests.

Table 2 Outcome measurement tools as they relate to ICF domains and potential independent variables that could influence the accuracy of the measurement and/or the appropriate interpretation of data derived from clinical therapies targeted to the CNS

<i>International Classification of Functioning, disability and health (ICF) as ICF relates to CNS-targeted therapies for spinal cord injury</i>			
<i>Level or domain</i>	<i>Body structure and function</i>	<i>Activity</i>	<i>Participation</i>
What change in outcome is measured?	Improvement (or impairment), that is, Neurological or physiological	Functional capacity (or limitation)	Quality of life (or restriction)
Outcome tools (examples)	ISNCSCI scores, electrophysiology	SCIM, 10 m walk	CHART, SF-36
What factors influence outcomes within each domain? ^{2a}	CNS integrity Neural circuits (sensory/motor)	CNS integrity Neural circuits Adaptive behaviours ^a Rehabilitation ^a Psychology/motivation ^a	CNS integrity Neural circuits Adaptive behaviours ^a Rehabilitation ^a Psychology/motivation ^a Community/family Support ^a Finances ^a Work/school ^a

Abbreviations: CHART, craig handicap and reporting technique; CNS, central nervous system; ISNCSCI, international standards for neurological classification of spinal cord injury; SF-36, short form 36; SCIM, spinal cord injury measure.

Increasing number of independent variables.

^aPotential independent variable.

Second, MCID helps with the determination of effect size, which is important to trial design. Third, it emphasises the primacy of the perspective of the person living with SCI, which hopefully correlates with that of the clinician or caregiver. In fact, investigators should take an active role in defining MCID where and when participants might not be sufficiently informed to make an adequate determination independently (for example, no one can perceive an improved bone mass).

Best practices in clinical care increasingly emphasise shared decision-making in determining treatment goals.¹⁷ Although the MCID concept appears to be easily understood by clinicians and trial investigators,¹⁸ it does involve the inherent risk of a trial participant setting an unrealistic trial end point, whether it is too high or too low.¹⁹ It should be noted that MCID has also been termed the minimal important difference, as well as clinically important change and minimal important change when used to describe longitudinal changes over time within the same patient.²⁰

During the initial stages of a clinical trial programme (phase I and II), there is no need to declare a primary clinical end point or to specify an MCID for the determination of a clinical benefit by an experimental treatment. In fact, the opposite is encouraged. phase II studies were developed to encourage investigators to explore various end points and outcomes measures so as to determine what would be most effective in a pivotal phase III study. In addition, historical longitudinal data sets can be used to model various recovery patterns for a prospective control cohort, and this information can be used to guide the determination of an appropriate clinical end point. Thus, in early-phase trials, statistical differences alone are acceptable between the experimental and control study cohorts. However, what statistical options in phase II should be explored when developing an MCID concept for pivotal phase III trials?

STATISTICAL CONSIDERATIONS, MDD AND MCID

Continuous scales

Continuous scales, such as quantitative measurements of time and distance (for example, time to walk a short distance or the distance walked within a short time), are the most readily and confidently interpreted outcomes of an SCI clinical trial. Numerous parametric

statistical tests have been developed and repeatedly validated to accurately discriminate outcomes between experimental and control groups.

MDD, also known as smallest real difference or minimum detectable change, is the smallest change that can be considered beyond any measurement error with a given level of confidence (usually at the 95% confidence level²¹). A valid MCID cannot be less than the MDD and preferably should exceed the value of the observed MDD.^{19,21,22}

The few continuous SCI outcome scales, such as timed or distance-measured assessments of walking, can be used as possible clinical end points, but this is limited to a subset of study participants who have some independent ambulatory capability. The measures include the 10-m walk, the 2- or 6-min walk or the timed up and go. For the 10-metre walk, the MDD and/or MCID has been suggested to lie within a range of improvement from 0.06^{23,24} to 0.13 m s⁻¹.²⁵ However, these estimates relied on grouping heterogeneous incomplete SCI subjects, which may have limited the accuracy of the estimate for the range of functional abilities. For the 6-min walk, a minimal detectable difference (MDD) has been suggested as 45.8 m (~150 feet), but this also involved a heterogeneous population of patients with incomplete SCI.²⁵ For the 2-min walk, an MDD of 4 m (13 feet) has been proposed for people with incomplete SCI, who were not community ambulators.²⁶ For timed up and go, van Hedel *et al.*²⁷ reported an MDD of 10.8 s. The uncertainty for these estimates emphasises the heterogeneity in the level and severity of SCI, which can influence an individual's functional capacity. There is also heterogeneity in the assessment time point, in relation to time since SCI, which can also alter the determination of a universal MDD and any inference of a MCID. In brief, MDD, and thus MCID values, may be distinct when starting from different baseline scores and also dependent on the neurological characteristic or functional activity being assessed.^{19,28} Nevertheless, what has not been agreed upon is whether an increase in the distance walked is more meaningful than an increase in walking speed.

Therefore, the heterogeneity of the SCI population is the most significant challenge to defining both an appropriate clinical end point for a trial and what might be a reasonable MCID. The selection of

relatively homogeneous subject cohorts would make any of these tasks more manageable.

Ordinal scales

Unfortunately, many widely used SCI clinical scales, either neurological or functional (for example, ISNCSCI, SCIM), are actually ordinal in nature, with the separation in assigned values (for example, integers) neither equal nor involving a consistently quantifiable change in value. Statistically, ordinal scales are much more challenging than continuous scales for confidently interpreting accurate outcomes. The statistical complexity is further compounded by the fact that most SCI-specific outcome tools (for example, ISNCSCI motor scores, SCIM items) are composite variables that generate cumulative sum scores, which are a sum of several independent ordinal values. The latter point may not be a trivial concern, and it depends on the magnitude of the statistical insufficiencies when a formula designed for continuous data is applied to ordinal data. Thus, concepts associated with normally distributed data, such as s.e.m., s.d., effect size or MDD, lose accuracy when applied to an ordinal variable that may not be normally distributed. In this context, even more robust measures of location and distribution (for example, median and box plots), as well as tests that rank ordinal outcomes, still may not be ideal.

Estimates of MDD and MCID from ordinal scales

Over the past 30 years, there have been several attempts to define optimal MDD and MCID algorithms, and several approaches have been reported for calculating MCID from ordinal data. In general, methodological approaches can be classified into two broad groups—distribution-based and anchor-based—and most of the calculations are based on parametric formula. There are several papers available for a more in-depth review of these methods.^{18,19,23,29–31} The wide range of available calculation methods is a limitation, as it can produce a wide range of available MCID values for any outcome assessment. This myriad of methods, as well as a plethora of synonymous terms (describing the same statistical approach), leads to problems of interpretation when deciding the validity of any MDD or MCID value.³²

A limitation of the MDD is that it assumes that detectable changes are uniform throughout the outcome scale being used, which may not be true of SCI ordinal scales. Thus, for ordinal data, MDD will likely not be a straightforward calculation, and the use of formulas for continuous data could lead to errors.

Distribution-based approaches focus on the statistical characteristics of the patient sample and compare an observed change with an index of variability to determine whether the change is meaningful. Such approaches have used an array of variability measures, including *P*-values of <0.05 (below 5%) as calculated from simple parametric tests, the s.e.m. and the s.d. Some authors have suggested that 0.5 s.d.³³ or 1 s.e.m.³⁴ may approximate an MDD and thus an MCID for some instruments. However, statistical significance does not necessarily describe the clinical benefit of the therapeutic effect.

A better statistic is probably effect size, which quantifies the size of the difference between two groups and is thought to be a more accurate measure for a significant difference.³⁵ Effect size is the standardised mean difference between two groups (that is, between control and treatment groups or baseline and end-point scores). It is determined by dividing the difference between group mean scores by the s.d. (variance) of the mean scores in both groups at baseline. Calculated values of 0.2, 0.5 and 0.8 have been suggested to correspond to small, moderate and large magnitude changes,

respectively, but interpretation of the meaningful benefit associated with effect size is arbitrary.³⁵

Samsa *et al.*³⁶ have suggested that an effect size of 0.2 could serve as an appropriate MCID value, whereas others have argued that an effect size of 0.5 (a half s.d.) is a more reasonable MCID threshold.³³ A limitation of the above formula is that the effect size will decrease if the variability of the total study population, at baseline, is large. This is another argument for using relatively homogeneous study cohorts in clinical trials. Nevertheless, one of the distinct advantages of effect size is that different treatments can be directly compared with each other for their relative significance (if the two study populations and protocols are similar).

Anchor-based methods compare outcome changes with an anchor measure that is itself interpretable—that is, it has a known relevance to study investigators and/or study participants. For example, a global rating of change or status on an important and easily understood measure of function, including the presence of symptoms, change in scores of participants with a similar severity of the disorder, prognosis of future events such as community participation or participant perception, could be appropriate anchors, depending on the therapeutic intervention and the prospective trial end point.³⁷

Anchor-based methods require at least moderate correlation ($r > 0.3$) between the anchor and the change observed by the outcome instrument.³⁸ After trial completion, global perception of treatment efficacy by participants has been an anchor used to develop MCID values. In the article by Jaeschke *et al.*,¹⁶ one example of an MCID was defined by the change in PRO scores from a group of participants selected according to their answers on a Likert-type scale. Participants rated their change on a 15-point global scale ($-7 =$ 'much worse' through $0 =$ 'no change' to $+7 =$ 'much better'). The MCID was defined as the average change of the participants who exhibited a small change (that is, who scored themselves with $\pm 1, 2$ or 3).¹⁶ However, the number of levels influences the determination of MCID threshold: the larger the number of levels, the smaller the projected MCID value.¹⁹ However, fewer levels along an ordinal scale may mean that the gap between adjacent ranks may be too large and thus might miss detecting a subtle but meaningful change. The selection of anchor criteria is rarely based on statistical analysis alone, and it is often made on the basis of the knowledge and experience of the clinical investigator as to what is clinically meaningful, a PRO or a consensus-based process.³⁹ It has been suggested that an MCID determined by an anchor-based method should be confirmed and validated by a distribution-based method.⁴⁰

In summary, statistical significance alone is valuable for early-phase clinical trials (for example, phase II), and it provides support for continued investigation and confirmation in a pivotal phase III study. In SCI clinical trials, in which sample sizes may be small, the possibility of a type II error (not rejecting the null hypothesis that there is no difference between an experimental group and a control group when this difference actually exists) is at least as problematic as failing to meet a *P*-value of ≤ 0.05 . There has also been growing recognition that a statistically significant difference between the experimental group and the control group (in a phase III trial) does not guarantee that the therapeutic effect has provided a clinically meaningful benefit. In short, statistical significance may not correlate with clinical significance.^{17,41} Nevertheless, a strong correlation between a positive structure/function change and an improved activity outcome may be sufficient and is currently the best interim strategy, while greater SCI trial experience is obtained.

DETERMINING MCID WITHIN THE REALM OF ICF DOMAINS

Objective outcome measurements and precise interpretation are critical to both SCI clinical practice and SCI clinical trials. As outlined in Tables 1 and 2, the ICF lists types of outcome assessments within three broad domains (categories) of body structure and function, activity and participation.^{42,43} Within the domain of body structure and function, assessments are often aimed at describing the structural and functional changes associated with the tissue targeted by the therapeutic. A change in neurological function or impairment of the injured spinal cord, regardless of the ability of the patient to functionally use any change in neural connectivity for improved independent ADLs and mobility, is one type of body structure/function assessment. Examples here would include elements of the ISNCSCI assessment, electrophysiological measurements, aspects of muscle strength, spasticity, joint range of motion and the newly emerging imaging technologies.

Within the activity domain, assessments focus on the capacity of a person with SCI to perform ADLs and mobility tasks, as measured and scored by global instruments such as the SCIM.^{44,45} In addition, in an effort to increase the sensitivity and accuracy for detecting therapeutic effects, more specific activity or targeted functional capacity measures have also been developed. For example, SCIM is divided into sub-scales, in which the self-care sub-scale is directed to changes in upper extremity activities, whereas ambulatory items within the mobility sub-scale of SCIM can track changes in lower extremity function. In addition, there is the Graded Redefined Assessment of Strength, Sensibility and Prehension (GRASSP) for upper extremity capacity,⁴⁶ as well as various ambulatory measurement tools for the lower extremity, such as the 10-m walk and the 6-min walk.¹²

Surveys, questionnaires and subject interviews are often used to determine participation within the community, as well as QoL and perceived pain. Outcome tools include the Craig Handicap Scale and Reporting Technique,⁴⁷ Perceived Handicap Questionnaire,⁴⁸ short form 36, and its variations, satisfaction with life survey and numerous other tools that are too lengthy to discuss here (for further description, see <http://www.parqol.com>). Many of the participation outcome tools are ordinal and semi-quantitative, similar to many of the tools within the ICF structure-function and activity domains. Evaluations, such as QoL surveys, rely more on subjective PROs. Table 2 attempts to summarise some of the outcome tools that are currently used in evaluating subject recovery within the three main ICF domains. There are several articles that can be consulted for more details about the appropriate use of these outcome measurement tools.^{10,12-14}

At present, no MCID values for SCI have been conclusively established for any ICF domain or for any SCI outcome measurement tool, and there are at least three considerations when determining an MCID. First, calculating an MDD is not trivial for ordinal data, and it should be completed before moving to an estimate of an MCID. Second, there are a number of independent (that is, uncontrolled) variables that could be a source of uncertainty or 'noise' in any activity or participation measure and thus challenge the accurate determination of a therapeutic benefit from the collected data. The number of potential independent variables escalates as you progress across the ICF domains (see Table 1). Therefore, relying only on an ICF participation measure or QoL survey, as a primary trial end point, could result in an erroneous determination of a treatment effect. Third, it would be better if changes observed in a valid structure and function assessment of neurological impairment (often the biological target of a therapeutic) correlated strongly with changes measured by another instrument at the level of activity or participation, including a PRO.

Cervical sensorimotor complete SCI as an example within the structure/function domain

Good correlations across ICF domains are highly desirable and possibly indicative of what threshold value might be implemented and accepted as an MCID.^{2,3} The ISNCSCI assessment is a standardised and internationally adopted classification scheme for patients with traumatic SCI, and it has been widely adopted as a measure of neurological impairment.^{15,49,50} Establishing a functionally meaningful ISNCSCI motor score or motor level change is assisted by knowledge of the natural history for spontaneous recovery, and this depends on both the level and severity of SCI.^{2,3,5} In addition, any proposed MCID value for a particular subject population should be greater than the MDD (that is, degree of measurement error) for the outcome value.

For example, a change of two motor points has been previously reported as the MDD for the assessment of an acute upper extremity motor score (UEMS) from the ISNCSCI.⁴⁹ More recently, the MDD was estimated to be three motor points for cervical sensorimotor complete SCI.³ Thus, there is some agreement on an approximate MDD for UEMS measurements, although the validity of the parametric statistical formula might be questioned, as the data are ordinal in nature. However, any value that might be used as a clinical end point requires consideration for the magnitude of change in the outcome over the study duration, which must include consideration for the degree of spontaneous recovery within that specific study population.

Acute cervical sensorimotor complete SCI is a relatively more homogeneous subset of SCI. Several analyses of SCI data sets have reported an approximate 10 motor-point improvement in the spontaneous recovery of the UEMS, with six points being the average over the first 6 months.^{1-3,9} Thus, adding the value of an MDD to that of the anticipated spontaneous recovery in UEMS might provide a conservative value for demonstrating a therapeutic benefit (for example, ~10 UEMS points). However, in terms of any suggested MCID, the segmental distribution of any increased UEMS score is more important than the absolute value of the UEMS change.

Significant statistical changes are most useful for guiding early clinical phase studies, but may not be satisfactory for a pivotal phase III study. As suggested above, any UEMS change that exceeds the prospectively defined threshold may still not adequately track or document a functional clinical benefit (that is, MCID), especially if the UEMS improvement is distributed across many different spinal cord segments within the upper extremities. Thus, a good correlation between ISNCSCI motor level changes and SCIM self-care scores has implied that a recovery of two motor levels (on either side) within the cervical cord might serve as a clinically meaningful end point, but just for acute cervical sensorimotor complete SCI.¹⁻³

Depending on the cervical segment, however, it could be argued that even one motor level of recovery could be clinically meaningful and be an appropriate MCID threshold (for example, recovery of independent respiration could be accomplished with an improvement in the motor level from C4 to C5). However, examination of natural spontaneous recovery indicates that about 66% of people with cervical sensorimotor complete SCI will gain one motor level within the first year after SCI, whereas only 25% will spontaneously recover two motor levels.^{2,3} The high percentage of individuals exhibiting one motor level of spontaneous recovery suggests that investigators may encounter a statistical 'ceiling' effect, unless a very large sample size (>500) is enrolled, which is pragmatically difficult. Estimates of spontaneous recovery are important to appropriate trial design, and modelling exercises comparing and contrasting neurological and

functional outcomes can be highly instructive and should be completed for other levels and severities of SCI.

SCIM as an example within the activity domain

Within the activity domain of ICF, the SCIM scale was developed specifically for people living with SCI as a means to evaluate their rehabilitation progress, their functional capabilities and preferences with regard to ADLs.⁴⁴ Although most studies have focused on the psychometric validity and reliability of SCIM,^{45,51} a few have examined SCIM for its value to determining functional MCID values. Scivoletto *et al.*⁴ and colleagues used distribution-based approaches to calculate a clinically significant change of SCIM for people living with SCI and concluded that an improvement of at least 4 points in the total SCIM was needed to obtain an MCID. This result is based on parametric statistical assumptions that may not have been achieved, and it relies on the psychometric properties of SCIM, instead of the informed opinion of clinical investigators and/or trial participants. Further investigation is needed to develop MDD and MCID for SCIM. It is not known whether one value is accurate for both the total SCIM score or any SCIM sub-scale, as well as all types of SCI and time points after SCI. The generation of a universal MCID value may not be possible, but a series of MDD and MCID values could be developed with sufficient longitudinal recovery data.

Pain as a classic example of a PRO

Chronic pain has long been known as a major ongoing health concern after SCI. Because the discomfort of pain can only be perceived and appreciated by the individuals themselves, PROs are the most clinically relevant assessments available. Even if a functional physiological measure correlated with a neurological mechanism for pain, the perceived level of distress will vary between individuals and also change according to the person's emotional state and circumstances under which the assessments were taken.⁵² PROs provide the 'gold standard' in assessing pain outcomes, because they reveal the individual's inherent perception of pain, which is most important when evaluating the clinical benefit of a pain therapy.

This is probably the clearest example of where a PRO is likely to supersede any physiological measure. Although several clinical scales have been developed to better characterise pain after SCI,^{53,54} the changes in perceptual pain rating remain the primary outcome from which to assess clinical efficacy. To address MCID, a panel of experts recommended that a 30% reduction in a self-reported pain rating could be considered clinically beneficial.⁵⁵ Originally unsupported by empirical evidence, a recent study has shown that such a threshold is associated with statistically significant improvements in objective functional outcomes (for example, lifting capacity⁵⁶). Although further validation is required, a recent phase III study in SCI was able to detect a 30% reduction in individuals treated with pregabalin,⁵⁷ thereby demonstrating the feasibility of this percent change.

DEVELOPING MCIDS FOR SCI: A CHALLENGING AND EVOLVING PROCESS

None of the approaches, outlined above, take into account the cost and the risks of the experimental treatment, which are important considerations for the merits of a treatment, if and when an MCID has been achieved. In brief, a small improvement may be acceptable for a low-risk therapeutic, especially if it were also a low-cost treatment.¹⁷ However, high-risk strategies, with significant potential for causing adverse events or unwanted side effects, may require a more rigorous demonstration of significant functional improvement to justify the risk. Similarly, a participant might agree that he/she has improved, but

based on the cost healthcare payers (private or government insurer) may consider that the benefit is not worth the cost of treatment.⁵⁸

It is important to emphasise that any of the above listed MDD and MCID values cannot be expected to remain constant for all SCI levels (for example, thoracic and lumbosacral injuries) or for the highly variable neurological impairments across the spectrum of incomplete SCI (AIS-B through to AIS-D). In fact, what is functionally meaningful for a person who is C5 motor complete, but sensory incomplete, may be different from what is meaningful for a person who is C5 or C7 motor and sensory incomplete. Further analyses and focus-group discussions between experienced clinicians, scientists and people living with SCI should be instructive. Knowledge of the MDD and spontaneous recovery rates are only initial steps to define an MCID value, but they provide a starting point for discussion. Thus, for each outcome measurement tool, the calculation of an MDD, as well as the spontaneous recovery rate, should be completed for each distinct SCI population. In addition, for any given SCI outcome measurement tool, the MCID value for acute SCI studies is likely to be different from its chronic equivalent.

Acute SCI

In acute clinical situations, almost all people living with SCI show some amount of spontaneous recovery, especially after incomplete SCI.^{1,3,9,10} During early time points after SCI, a person will usually lack the appropriate information and experience to judge the amount of functional change (trial end point) that would eventually have a meaningful impact on their life. Thus, in comparison with clinical investigators, newly injured SCI individuals have been observed to overestimate the magnitude of change that they expect to be meaningful and available from any current therapeutic intervention.⁵⁹ As recovery progresses, expectations for therapeutic benefits are likely to change, but no careful longitudinal survey of the expectations of people living with SCI has been completed to date.

Therefore, in the best interests of people living with SCI, clinical experience and expertise is often helpful to define and refine reasonable MCID thresholds. Nevertheless, the shared responsibility for the determination of MCID means that the reasoned expectations of people living with SCI must be sought and discussed when determining reasonable MCID values.

Chronic SCI

People who have lived with chronic SCI have had time to appreciate the natural history of SCI, in themselves and others they have encountered. They are more likely to appreciate how even small functional changes can be beneficial. Most people living with chronic SCI have a relatively stable baseline of function;⁹ thus, they can more readily perceive the benefits of even modestly effective interventions. Even though there is a lack of empirical data, it is intuitive to suggest that people living with chronic SCI are likely to have insightful expectations for improvement and could serve as patient advocates for suggesting MCID thresholds, as well as consultants to individuals participating in an acute SCI intervention study.

In addition, chronic SCI may have different trial protocols with more interest focused on longitudinal changes within each individual, rather than between-group differences. An obscure but important distinction in the MCID literature is that between-group differences versus longitudinal, within-an-individual, changes are likely to provide different MCID values.³¹

Summary

At this time, the best strategy that can be offered for determining a therapeutic benefit must rely on the correlation between a structure/function measure and a relevant activity outcome, as well as statistically significant differences between experimental and control study cohorts. At present, MCID estimates are very difficult. In the future and if they are based on systematic research and relevant anchors, they could provide one basis for evaluating SCI clinical trial data. This would assist all stakeholders to better understand the therapeutic benefits and limitations of experimental treatments, including regulatory agencies, healthcare payers, clinicians, scientists and people living with SCI. An MCID, based on PROs, is often included in the evaluation of clinical trial outcomes, but it is only one value that enhances our interpretation of trial data. A valid MCID cannot be less than the MDD, and it should exceed the value of the observed MDD, which in itself can be difficult to estimate. Thus, MCID estimates require careful consideration, cautious inference from modelling of large longitudinal SCI databases, as well as clinician and/or participant experience. Finally it has to be accepted that any MCID definition will, and should, evolve over time with increasing study experience.

DATA ARCHIVING

There were no data to deposit.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are grateful for the support and encouragement of the funders and contributing members of SCOPE (Spinal Cord Outcomes Partnership Endeavor, <http://www.scope-sci.org>) and EMSCI (European Multicenter Study about spinal Cord Injury, <http://www.emsci.org>).

- Steeves JD, Kramer JK, Fawcett JW, Cragg J, Lammertse DP, Blight AR *et al*. Extent of spontaneous motor recovery after traumatic cervical sensorimotor complete spinal cord injury. *Spinal Cord* 2011; **49**: 257–265.
- Steeves JD, Lammertse DP, Kramer JL, Kleitman N, Kalsi-Ryan S, Jones L *et al*. Outcome measures for acute/subacute cervical sensorimotor complete (AIS-A) spinal cord injury during a phase 2 clinical trial. *Top Spinal Cord Inj Rehabil* 2012; **18**: 1–14.
- Kramer JL, Lammertse DP, Schubert M, Curt A, Steeves JD. Relationship between motor recovery and independence after sensorimotor complete cervical spinal cord injury. *Neurorehabil Neural Repair* 2012; **26**: 1064–1071.
- Scivoletto G, Tamburella F, Laurenza L, Molinari M. The spinal cord independence measure: how much change is clinically significant for spinal cord injury subjects. *Disabil Rehabil* 2013; **28**: 1808–1813.
- 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**: 507–515.
- Kwon BK, Okon E, Hillyer J, Mann C, Baptiste D, Weaver LC *et al*. A systematic review of non-invasive pharmacologic neuroprotective treatments for acute spinal cord injury. *J Neurotrauma* 2011; **28**: 1545–1588.
- Mothe AJ, Tator CH. Review of transplantation of neural stem/progenitor cells for spinal cord injury. *Int J Dev Neurosci* 2013; **31**: 701–713.
- Varma AK, Das A, Wallace G, Barry J, Vertegel AA, Ray SK *et al*. Spinal cord injury: a review of current therapy, future treatments and basic science frontiers. *Neurochem Res* 2013; **38**: 895–905.
- Fawcett JW, Curt A, Steeves JD, Coleman WP, Tuszynski MH, Lammertse D *et al*. Guidelines for the conduct of clinical trials for spinal cord injury as developed by the ICCP panel: spontaneous recovery after spinal cord injury and statistical power needed for therapeutic clinical trials. *Spinal Cord* 2007; **45**: 190–205.
- Steeves JD, Lammertse D, Curt A, Fawcett JW, Tuszynski MH, Ditunno JF *et al*. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: clinical trial outcome measures. *Spinal Cord* 2007; **45**: 206–211.
- Itzkovich M, Gelernter F, Catz A. The spinal cord independence measure (SCIM) version III: reliability and validity in a multi-center international study. *Disabil Rehabil* 2007; **29**: 1926–1933.

- Alexander MS, Anderson KD, Biering-Sorenson F, Blight AR, Brannon R, Bryce TN *et al*. Outcome measures in spinal cord injury: recent assessments and recommendations for future directions. *Spinal Cord* 2009; **47**: 582–591.
- Hill MR, Noonan VK, Sakakibara BM, Miller WC. SCIRE Research Team. Quality of life instruments and definitions in individuals with spinal cord injury: a systematic review. *Spinal Cord* 2010; **48**: 438–450.
- Furlan JC, Noonan V, Fehlings MG. Assessment of disability in patients with acute traumatic spinal cord injury: a systematic review of the literature. *J Neurotrauma* 2011; **28**: 1413–1430.
- Kirshblum SC, Waring W, Biering-Sorensen F, Burns SP, Johansen M, Schmidt-Read M *et al*. Reference for the 2011 revision of the international standards for neurological classification of spinal cord injury. *J Spinal Cord Med* 2011 **34**: 547–554.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; **10**: 407–415.
- Tuszynski MH, Steeves JD, Fawcett JW, Lammertse D, Kalichman M, Rask C *et al*. Guidelines for the conduct of clinical trials for spinal cord injury as developed by the ICCP Panel: clinical trial inclusion/exclusion criteria and ethics. *Spinal Cord* 2007; **45**: 222–231.
- Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; **77**: 371–383.
- Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine* 2007; **7**: 541–546.
- Gatchel RJ, Lurie JD, Mayer TG. Minimal clinically important difference. *Spine J* 2010; **35**: 1739–1743.
- Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res* 2001; **10**: 571–578.
- Schuck P, Zwingmann C. The “smallest real difference” as a measure of sensitivity to change: a critical analysis. *Int J Rehabil Res* 2003; **26**: 85–91.
- Musselman KE. Clinical significance testing in rehabilitation research: What, why and how? *Physical Ther Rev* 2007; **12**: 287–296.
- Musselman KE, Yang JF. Walking tasks encountered by urban-dwelling adults and persons with incomplete spinal cord injuries. *J Rehabil Med* 2007; **39**: 567–574.
- Lam T, Noonan VK, Eng JJ. SCIRE Research Team. A systematic review of functional ambulation outcome measures in spinal cord injury. *Spinal Cord* 2008; **46**: 246–254.
- Field-Fote EC, Roach KE. Influence of a locomotor training approach on walking speed and distance in people with chronic spinal cord injury: a randomized clinical trial. *Phys Ther* 2011; **91**: 48–60.
- 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.
- Wang YC, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. *Phys Ther* 2011; **91**: 675–688.
- Scott SC, Goldberg MS, Mayo NE. Statistical assessment of ordinal outcomes in comparative studies. *J Clin Epidemiol* 1997; **50**: 45–55.
- Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; **56**: 395–407.
- King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011; **11**: 171–184.
- Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther* 2012; **20**: 160–166.
- Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; **41**: 582–592.
- Wyrwich KW, Tierney WM, Wolinsky F. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; **52**: 861–873.
- Cohen J. *Statistical power analysis for the behavioral sciences*, 2nd edn. Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.
- Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important difference in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics* 1999; **15**: 141–155.
- Brozek JL, Guyatt GH, Schunemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual Life Outcomes* 2006; **4**: 69.
- Revicki D, Hays SD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; **61**: 102–109.
- Gatchel RJ, Mayer TG, Choi Y, Chou R. Validation of a consensus-based minimal clinically important difference (MCID) threshold using an objective functional external anchor. *Spine J* 2013; **13**: 889–893.
- Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C *et al*. What is clinically meaningful change on the functional assessment of cancer therapy (FACT-L) questionnaire? Results from Eastern Cooperative Oncology Group (ECOG). *J Clin Epidemiol* 2002; **55**: 285–295.
- Lammertse D, Tuszynski MH, Steeves JD, Curt A, Fawcett JW, Rask C *et al*. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP Panel: Clinical trial design. *Spinal Cord* 2007; **45**: 232–242.

- 42 World Health Organization. *International Classification of Functioning, Disability and Health: ICF*. WHO: Geneva, Switzerland, 2001.
- 43 Kirchberger I, Sinnott A, Charlifue S, Kovindha A, Lüthi H, Campbell R *et al*. Functioning and disability in spinal cord injury from the consumer perspective: an international qualitative study using focus groups and the ICF. *Spinal Cord* 2010; **48**: 603–613.
- 44 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.
- 45 Anderson KD, Acuff ME, Arp BG, Backus D, Chun S, Fisher K *et al*. United States (US) multi-center study to assess the validity and reliability of the spinal cord independence measure (SCIM III). *Spinal Cord* 2011; **49**: 880–885.
- 46 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**: 905–914.
- 47 Whiteneck GG, Charlifue SW, Gerhart KA, Overholser JD, Richardson GN. Quantifying handicap: a new measure of long-term rehabilitation outcomes. *Arch Phys Med Rehabil* 1992; **73**: 519–526.
- 48 Tate D, Forchheimer M, Maynard F, Dijkers M. Predicting depression and psychological distress in persons with spinal cord injury based on indicators of handicap. *Am J Phys Med Rehabil* 1994; **73**: 175–183.
- 49 Marino RJ, Jones L, Kirshblum S, Tal J, Dasgupta A. Reliability and repeatability of the motor and sensory examination of the international standards for neurological classification of spinal cord injury. *J Spinal Cord Med* 2008; **31**: 166–170.
- 50 Furlan JC, Fehlings MG, Tator CH, Davis AM. Motor and sensory assessment of patients in clinical trials for pharmacological therapy of acute spinal cord injury: psychometric properties of the ASIA standards. *J Neurotrauma* 2008; **25**: 1273–1301.
- 51 Bluvshstein V, Front L, Catz A, Aidinoff E, Gelernter I, Hart J *et al*. SCIM III is reliable and valid in a separate analysis for traumatic spinal cord lesions. *Spinal Cord* 2011; **49**: 292–296.
- 52 Lumley MA, Cohen JL, Borszcz GS, Cano A, Radcliffe AM, Porter LS *et al*. Pain and emotion: A biopsychosocial review of recent research. *J Clin Psychol* 2011; **67**: 942–968.
- 53 Jensen MP, Dworkin RH, Gammaitoni AR, Olaleye DO, Oleka N, Galer BS. Assessment of pain quality in chronic neuropathic and nociceptive pain clinical trials with the Neuropathic Pain Scale. *J Pain* 2005; **6**: 98–106.
- 54 Hallstrom I, Norrbrink C. Screening tools for neuropathic pain: can they be of use in individuals with spinal cord injury? *Pain* 2011; **152**: 772–779.
- 55 Dworkin RH, Turk DC, Farrar JT, Olaleye DO, Oleka N, Galer BS. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005; **113**: 9–19.
- 56 Gatchel RJ, Mayer TG, Chou R. What does/should the minimum clinically important difference measure? A reconsideration of its clinical value in evaluating efficacy of lumbar fusion surgery. *Clin J Pain* 2012; **28**: 387–397.
- 57 Cardenas DD, Nieshoff EC, Suda K, Goto S, Sanin L, Kaneko T *et al*. A randomized trial of pregabalin in patients with neuropathic pain due to spinal cord injury. *Neurology* 2013; **80**: 533–539.
- 58 Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. *Pharmacoeconomics* 2000; **18**: 419–423.
- 59 van Hedel HJ, Dokladal P, Hotz-Boendermaker S. Mismatch between investigator-determined and patient-reported independence after spinal cord injury: Consequences for rehabilitation and trials. *Neurorehab Neural Repair* 2011; **25**: 855–864.