Minimizing Errors in Acute Traumatic Spinal Cord Injury Trials by Acknowledging the Heterogeneity of Spinal Cord Anatomy and Injury Severity: An Observational Canadian Cohort Analysis

Marcel F. Dvorak,¹ Vanessa K. Noonan,^{2,25} Nader Fallah,^{2,25} Charles G. Fisher,¹ Carly S. Rivers,² Henry Ahn,³ Eve C. Tsai,⁴ A.G. Linassi,⁵ Sean D. Christie,⁶ Najmedden Attabib,^{7,8} R. John Hurlbert,⁹ Daryl R. Fournay,²⁴ Michael G. Johnson,¹⁰ Michael G. Fehlings,¹¹ Brian Drew,^{12,13} Christopher S. Bailey,¹⁴ Jérôme Paquet,¹⁵ Stefan Parent,^{16,17} Andrea Townson,¹⁸ Chester Ho,¹⁹ B.C. Craven,²⁰ Dany Gagnon,²¹ Deborah Tsui,²² Richard Fox,²³ Jean-Marc Mac-Thiong,¹⁶ and Brian K. Kwon²⁵

Abstract

Clinical trials of therapies for acute traumatic spinal cord injury (tSCI) have failed to convincingly demonstrate efficacy in improving neurologic function. Failing to acknowledge the heterogeneity of these injuries and under-appreciating the impact of the most important baseline prognostic variables likely contributes to this translational failure. Our hypothesis was that neurological level and severity of initial injury (measured by the American Spinal Injury Association Impairment Scale [AIS]) act jointly and are the major determinants of motor recovery. Our objective was to quantify the influence of these variables when considered together on early motor score recovery following acute tSCI.

Eight hundred thirty-six participants from the Rick Hansen Spinal Cord Injury Registry were analyzed for motor score improvement from baseline to follow-up.

In AIS A, B, and C patients, cervical and thoracic injuries displayed significantly different motor score recovery. AIS A patients with thoracic (T2-T10) and thoracolumbar (T11-L2) injuries had significantly different motor improvement. High (C1-C4) and low (C5-T1) cervical injuries demonstrated differences in upper extremity motor recovery in AIS B, C, and D. A hypothetical clinical trial example demonstrated the benefits of stratifying on neurological level and severity of injury. Clinically meaningful motor score recovery is predictably related to the neurological level of injury and the severity of the baseline neurological impairment. Stratifying clinical trial cohorts using a joint distribution of these two variables will enhance a study's chance of identifying a true treatment effect and minimize the risk of misattributed treatment effects.

¹Department of Orthopedics, ¹⁸Department of Physical Medicine and Rehabilitation, ²⁵University of British Columbia, Vancouver, British Columbia, Canada.

²Department of Research, Rick Hansen Institute, Vancouver, British Columbia, Canada.

³University of Toronto Spine Program, St. Michael's Hospital, Toronto, Ontario, Canada.

⁴Department of Surgery, The Ottawa Hospital, Ottawa Hospital Research Institute, University of Ottawa, Ontario, Canada.

⁵Department of Physical Medicine and Rehabilitation, ²⁴University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

⁶Department of Neurosurgery, Dalhousie University, Halifax, Nova Scotia, Canada.

⁷Horizon Health Network/Dalhousie University, St. John, New Brunswick, Canada.

⁸Saint John Regional Hospital, St John, New Brunswick, Canada.

⁹University of Calgary Spine Program, Calgary, Alberta, Canada.

¹⁰Department of Surgery, Orthopaedics and Neurosurgery, University of Manitoba, Winnipeg, Manitoba, Canada.

¹¹Division of Neurosurgery and Spinal Program, University of Toronto, Ontario, Canada.

¹²Hamilton General, Hamilton, Ontario, Canada.

¹³McMaster University, Hamilton, Ontario, Canada.

¹⁴Division of Orthopaedic Surgery, Western University, London, Ontario, Canada.

¹⁵Hôpital Enfant-Jésus/Laval University, Quebec City, Quebec, Canada.

¹⁶Department of Orthopaedic Surgery, Hôpital Ste-Justine, Montréal, Quebec, Canada.

¹⁷Université de Montréal, Montréal, Quebec, Canada.

¹⁹Department of Clinical Neurosciences, University of Calgary, Alberta, Canada.

²⁰University Health Network, Toronto Rehabilitation Institute, Toronto, Ontario, Canada.

²¹École de réadaptation, Université de Montréal, Montréal, Quebec, Canada.

²²Hamilton Health Sciences, Hamilton, Ontario, Canada.

²³Department of Surgery, Royal Alexandra Hospital, Edmonton, Alberta, Canada.

Clinical studies should stratify participants based on these factors and record the number of participants and their mean baseline motor scores for each category of this joint distribution as part of the reporting of participant characteristics. Improved clinical trial design is a high priority as new therapies and interventions for tSCI emerge.

Key words: bias; motor score; neurological recovery; spinal cord injury; statistical error

Introduction

REPORTING THAT A THERAPEUTIC EFFECT EXISTS WHEN it does not (type I, alpha error) has major repercussions in acute traumatic spinal cord injury (tSCI) because of the lack of recognized treatment,¹ the relative desperation of the population,² and the substantial public sympathy and media interest that is mobilized around tSCI research. Promising pharmaceutical, medical, surgical, and rehabilitation interventions identified in preclinical studies heighten the expectations for subsequent clinical trials. The associated risk of type II (beta) error, which usually results from underpowered studies, may condemn potentially useful therapies, thus denying patients access to potentially beneficial therapies while contributing to inefficiencies in translational research.^{3,4}

Despite numerous clinical trials over the past 20 to 30 years and the acknowledged lack of progress in identifying clinically effective treatments for tSCI,^{5,6} one has to question whether successful therapies have eluded us simply due to the ineffectiveness of these potential treatments, or whether it is possible that there is a flaw in the methodological approach taken by these trials.

Researchers who translate neuroprotective therapies in traumatic brain injury and stroke have experienced similar challenges in establishing efficacy in clinical trials and have attributed this to multiple etiologies, including a lack of acknowledgement of the pathoanatomic and pathophysiologic heterogeneity of the injury processes and a poor understanding of the most influential baseline prognostic indicators.^{7,8}

Controlling for bias requires the identification and quantification of variables that strongly influence outcome. Adjustment for imbalances in these strong predictive variables through stratification at the time of recruitment would increase the power of a study, something of substantial relevance in the neurotrauma literature.^{4,9} This will increase the power of studies, improve the chances of detecting a true therapeutic effect and most importantly for the high profile and vulnerable tSCI population, increase the likelihood that a statistically significant effect represents a true effect.

Randomized controlled trials (RCTs) in acute tSCI research are hampered by small sample sizes due to immense recruitment issues¹⁰; therefore, large observational cohort studies are likely to play a major role in future tSCI research. It is equally important to ensure balanced representation of confounding variables in observational study cohorts as in RCT's.

SCI has the benefit of a reproducible measure of neurological recovery: voluntary motor function. This is the most important recovery target for patients¹¹ and a measure of impairment that is strongly correlates with functional outcome.¹² Motor recovery can be reliably measured using the International Standards for Neurologic Classification of Spinal Cord Injury (ISNCSCI), a widely adopted grading scale accepted by the major international spinal cord societies.¹³ Using ISNCSCI motor scores as an outcome is complicated by the fact that injuries at different spinal levels will have differences in the potential measurable deficits and thus different potential for recovery. The ISNCSCI only measures between C5-T1 and L2-S1 neurological levels, so recovery of motor func-

tion is not equally and comprehensively detectable by the ISNCSCI motor examinations. Further, there is an increasingly prominent ceiling effect on neurological recovery as the level of injury moves more caudally. Finally, it is possible that the observed recovery of motor function is not comprehensively detected by ISNCSCI as recovery occurs at varying rates and degrees within the various anatomic regions of the spinal cord (upper vs. lower cervical and thoracic vs. thoracolumbar).

Despite these obvious limitations—the fact that motor score does not fully encompass global spinal cord function and the statistical idiosyncrasies of motor scores as described by numerous authors^{14–19}—we chose to analyze early motor recovery to identify and characterize those covariates that will most profoundly influence prognostic models and thus impact the design, conduct, and analysis of potential future clinical trials.

Although motor recovery following tSCI may be influenced by many variables,²⁰ we selected the two dominant predictive variables: the neurological level of injury and the baseline severity of the neurological impairment.^{19,21} To represent the neurological severity of the initial injury, we use the American Spinal Injury Association Impairment Scale (AIS) as a surrogate.¹⁹ Clinical trials have used AIS as a primary outcome measure.^{22,23} We categorized the neurological level of injury into four groups: high cervical (C1–C4); low cervical (C5–T1); thoracic (T2–T10); and thoracolumbar (T11–L2), basing these on the ISNCSCI last normal neurological level. There is no consensus in current or prior clinical trials on how to quantify and characterize the influence of neurological level and severity of injury,^{14,17,24} and we sought to investigate their influence on motor score recovery.

The purpose of this paper was to quantify the prognostic influence of neurological level and severity of injury on the magnitude of early motor recovery in a large observational cohort of acute tSCI patients to guide clinical scientists in recognizing the key prognostic variables for which tSCI clinical trials should stratify their participants. We proposed a standardized stratification and reporting template for clinical trials and provided a hypothetical example to demonstrate the impact of these prognostic variables on the design and analysis of tSCI clinical trials.

Methods

This Canadian multicenter observational study utilized data prospectively collected for the Rick Hansen Spinal Cord Injury Registry (RHSCIR). This ongoing cohort study recruited patients with acute tSCI presenting to one of 31 RHSCIR sites located in 16 cities representing nine of 10 Canadian provinces. Patients presenting to any RHSCIR site within one month from an acute tSCI are eligible for consent and enrollment. Non-traumatic etiologies of SCI (infection, neoplasm, iatrogenic, acute vascular) were excluded. The specific details of the data collection and registry structure is reviewed elsewhere.²⁵

The subjects were recruited to the RHSCIR from 2004 to December 2012. The final follow-up neurological examination was performed prior to March 31, 2013. We included patients who had a baseline ISNCSCI neurological assessment performed up to a

maximum of 30 d from injury, and a follow-up assessment at least 30 d after their baseline assessment. These broad inclusion criteria were selected for several reasons. Unlike a clinical trial where there are set times for assessments, registry data reflects clinical practice and the baseline neurological assessment is performed within a wider range, usually at the time of presentation to a specialized SCI center. Further, some clinical trials in SCI recruit in the very early acute phase while others recruit sub-acutely; thus, we included baseline examinations up to 30 d post-injury to enhance the generalizability of our results.

The change in upper extremity motor score (UEMS; maximum, 50) and change in total motor score (TMS; maximum, 100) were defined as the difference between the most recent follow-up and baseline respective motor scores.²⁶ Physicians, nurse practitioners, and physiotherapists who had been extensively trained in the ISNCSCI neurological examination performed detailed baseline and follow-up neurological assessments. The date and time of injury and each neurological assessment were recorded.

The neurological level of injury was defined as the most caudal segment with normal neurological function according to the ISNCSCI. The neurological level of injury was subdivided into high cervical (C1-C4), low cervical (C5-T1), thoracic (T2-T10), and thoracolumbar (T11-L2) based upon the neuroanatomical and physiological variability of these regions and the literature.^{19,21} The division of cervical into those with measurable motor efferent roots (C5-T1) and those without measured motor segments (C1-C4) is intuitive. Unlike the thoracic levels for which no measurable motor correlates exist, many segments through the thoracolumbar region (T11-L2) contain within the cord motor cell bodies whose function can be reliably measured using the ISNCSCI. Patients with a neurological level below L2 were excluded since these were likely to be predominantly cauda equina injuries.

The severity of neurological impairment was defined according to the AIS with grades from A (complete injury, no sensory or motor function is preserved in the sacral segments S4-5) to D (motor function is preserved below the neurological level). The AIS has been used as a primary outcome measure^{22,23} and has been proposed as a surrogate measure of injury severity.¹⁹

Statistical analysis

Parametric tests were used if variables had a normal distribution and non-parametric tests were used if either the data were not normally distributed or if there was a small sample size. Chi-square or Fisher's exact test were used for categorical data depending upon the number of the sample in each cell.

To evaluate the effect of the injury severity and the level of injury on changes in motor score, subgroup analyses were performed using a general linear model. A one-way between-subjects analysis of variance (ANOVA) was conducted when there were three or more groups. Post hoc comparisons using the Duncan and Scheffe's tests were used to detect any differences between the level/severity-based subgroups. Comparisons between two groups were analyzed using independent t-tests with Levene's test to assess the equality of variance. Mann-Whitney U and Kruskal-Wallis tests were used for comparing two groups or more than two groups, respectively, if the data were not normally distributed. We acknowledge the work of Geisler and his colleagues^{19,27} that described how motor and sensory scores do not follow the assumptions underlying normal-theory and that by pooling data from different strata may result in different probability distributions and give misleading results. We performed our statistical analyses specific to the distribution of data in each individual subgroup, thus minimizing heterogeneity and analyzing more statistically homogenous and clinically relevant subgroups. A *p* value of ≤ 0.05 was considered statistically significant. Data analysis was conducted using SPSS (version 20) and SAS (version 9.3).

A hypothetical SCI clinical trial scenario was created by randomly assigning participants from our RHSCIR study cohort to two equal groups (Group 1 and Group 2) using a computerized random number generator. The primary outcome in this hypothetical trial was an improvement in total motor score. The scenario was created not as a demonstration of the fundamental statistics and epidemiology of type I, or alpha error, but to demonstrate, using real world data, the methodological issue that we believe is plaguing many small SCI clinical trials. We used this example to highlight how acknowledging the heterogeneity in our proposed categories can explain imbalances that may lead to statistical error and how the reader, reviewer, and investigator can address this inadvertent selection bias through either a priori stratification or subsequent balancing of the heterogeneous participant population.

The hypothetical example also is presented to provide a real-world example of how conventional reporting of study participants is inadequate at exposing the imbalances in study populations. We demonstrate how the acknowledgment of our classification categories aids in appropriately powering trials.

Results

From a total population of 980 patients who met our inclusion criteria, 144 (14.7%) patients had one or more motor segments with missing individual motor scores, leaving a final cohort of 836 (85.3%) patients for our analysis. Table 1 describes the characteristics of the patient population and their neurological examinations. Institutional and university research ethics board approval was obtained at each of the centers.

The mean change in TMS and UEMS of all individuals in the cohort was sub-classified based upon the four neurological levels of injury and the initial AIS grade (A through D). This forms the basis of what we call the "Canadian Classification" and is seen in Table 2.

There was evidence of non-Gaussian distribution in motor score change; therefore, we used the Kruskal Wallis test to compare the four AIS groups. There were significant differences among the four groups ($p < 0.001$). Post hoc testing (Mann-Whitney U with Bonferroni correction) confirmed that the four groups should be considered separately. In each of the four AIS groups we also tested the normality of the data. With this particular data, AIS A and B are not normally distributed; therefore, the Kruskal Wallis test was used for comparison of the four anatomical regions within each AIS group. Thus, in AIS A and B participants, we used the Mann-Whitney U as a post hoc test with Bonferroni correction to adjust for the *p* value.

One-way ANOVA was used for comparing the four anatomical regions in AIS C and D participants, as the data distribution was normal. Two post hoc tests were used after getting a significant result in AIS C (Duncan and Scheffe).

A significant difference in mean total motor score recovery between the neurological levels of injury was identified for AIS A, B, and C ($p \leq 0.05$; Table 2). All AIS D injuries demonstrated similar TMS improvements with no significant differences among levels of injury ($p = 0.20$), although there was a larger motor score improvement in high cervical AIS D injuries.

A post hoc analysis of TMS recovery by neurological level and AIS demonstrated that when comparing high and low cervical levels, changes in TMS were not significantly different for AIS A, B, and C; however, significant differences in change in UEMS between high and low cervical levels were seen in AIS B (14.3 vs. 9.5), C (20.5 vs. 12.7) and D (11.4 vs. 7.5).

In AIS A thoracic (T2-T10) and thoracolumbar injuries (T11-L2), there was a significant difference in TMS improvement. No significant differences were observed between thoracic and thoracolumbar levels for the incomplete AIS categories (B, C, and D).

TABLE 1. STUDY PARTICIPANT DETAILS AS CONVENTIONALLY REPORTED IN CLINICAL TRIALS

	Baseline	Follow-up
Age: mean (range)	45 (14–91)	-----
Male gender: <i>n</i> (%)	665 (79.5%)	-----
Total hospital length of stay (days): mean (SD)	146.4 (92.1)	-----
Upper Extremity Motor Score (range 0 to 50): mean (SD)	30.9 (18.3)	37.9 (14.5)
Lower Extremity Motor Score (range 0 to 50): mean (SD)	11.2 (16.6)	19.4 (20.2)
Total Motor Score (0–100): mean (SD)	42.1 (24.9)	57.3 (27.1)
Neurological examination time (hours post-injury): mean (SD)	121 (168)	3125 (1875)
Baseline examination median	40	
Baseline interquartile range (IQR = Q3–Q1)	151	
Baseline examinations < 24 h post-injury	42.3%	
Baseline examination < 72 h post-injury	61.6%	
Follow-up examination mean days post-injury	130	
Follow-up examination median days post-injury	111	
AIS: number of individuals for each grade (%)		
A	388 (46.4)	303 (36.2)
B	105 (12.6)	87 (10.4)
C	180 (21.5)	95 (11.4)
D	163 (19.5)	335 (40.1)
E	-----	5 (0.6)
Anatomical region of injury number in each group (%)		
High cervical (C1–C4)	265 (31.7)	
Low cervical (C5–T1)	289 (34.6)	
Thoracic (T2–T10)	161 (19.3)	
Thoracolumbar (T11–L2)	121 (14.5)	
Glasgow Coma Scale: mean (SD)	14.2 (2.5)	-----
Steroids administered acutely: <i>n</i> (%)	72 (8.6)	-----
Surgery		-----
Surgery performed: <i>n</i> (%)	643 of 734 with complete data (87.6%)	
Time from injury to surgery (hours): mean (SD)	58 h (82.2)	

SD, standard deviation; AIS, American Spinal Injury Association Impairment Scale.

Hypothetical clinical trials and sample size

A computerized random number generator selected 64 participants, randomly assigning them to two groups in an attempt to mimic recruitment into a hypothetical trial similar in size to previously published trials.^{18,28–30} We compared these two groups for what are commonly reported characteristics of study participants—age, gender, AIS on admission, neurological level of injury, and mechanism of injury as outlined in Table 3—and found no significant differ-

ences. Groups 1 and 2 demonstrated a statistically significant ($p=0.05$) difference in motor score improvement. This is not a true effect but represents a type I error, which would be expected to occur one time out of twenty samplings despite randomized cohort selection. When we repeated this sampling process 20 times, two of the twenty occasions demonstrated “significant” differences in motor score improvement.

When the same hypothetical groups of patients are presented in the form of the Canadian Classification (Table 4) a significant

TABLE 2. MOTOR SCORE DISTRIBUTION BY SEVERITY AND NEUROLOGICAL LEVEL OF INJURY

AIS		Neurological Level				p value
		High cervical (C1–C4)	Low cervical (C5–T1)	Thoracic (T2–T10)	Thoracolumbar (T11–L2)	
A (<i>n</i> = 388)	Total motor score (SD)	10.4 (13.9)	9.0 (13.0)	1.15 (5.7)	5 (7.9)	<0.001
	<i>n</i>	88	127	122	51	
	Baseline motor score (SD)	6.4 (10.5)	28.3 (15.3)	50.4 (4.6)	53.9 (9.7)	
B (<i>n</i> = 105)	Total motor score (SD)	30.1 (22.6)	23.5 (23.6)	11.2 (16.5)	8.5 (7.6)	<0.001
	<i>n</i>	29	41	20	15	
	Baseline motor score (SD)	10.6 (12.1)	25.4 (15.6)	52.1 (7.9)	55.1 (5.6)	
C (<i>n</i> = 180)	Total motor score	41.3 (18)	35 (18.8)	20.55 (14.7)	17.5 (10.3)	<0.001
	<i>n</i>	72	61	11	36	
	Baseline motor score (SD)	26.0 (13.8)	37.6 (15.7)	60.4 (8.3)	61.5 (8.6)	
D (<i>n</i> = 163)	Total motor score	15.8 (15.4)	12 (14.8)	10.1 (10)	9.5 (7.2)	0.2
	<i>n</i>	76	60	8	19	
	Baseline motor score (SD)	66.8 (15.9)	74.2 (14.4)	87.6 (10.8)	83.7 (9.3)	

AIS, American Spinal Injury Association Impairment Scale; SD, standard deviation.

TABLE 3. PARTICIPANT CHARACTERISTICS OF A HYPOTHETICAL CLINICAL TRIAL DRAWN FROM OUR STUDY POPULATION

SCI variable	Group 1	Group 2	p value
Age	43.4 (19.9)	42.1 (19.5)	0.84
Gender, male (%)	27 (84.4%)	28 (87.5%)	0.72
Motor score baseline	36.9 (22.7)	43.4 (24.9)	0.25
AIS baseline: mean (SD)			
A	7 (21.9)	8 (25)	0.85
B	7 (21.9)	7 (21.9)	
C	13 (40.6)	10 (31.3)	
D	5 (15.6)	7 (21.9)	
Anatomical region: mean (SD)			0.83
C1-C4	13 (40.6)	10 (31.3)	0.58
C5-C8, T1	10 (31.3)	10 (31.3)	
T2-T10	5 (15.6)	6 (18.8)	
T11-T12, L1-L2	4 (12.5)	6 (18.8)	
Mechanism of injury: mean (SD)			
Fall	11 (34.4)	12 (37.5)	0.05
Sports	8 (25)	4 (12.5)	
Transport	10 (31.3)	11 (34.4)	
Other	3 (9.4)	5 (15.6)	
Motor score change: mean (SD)	32.8 (22.4)	21.2 (20.4)	

SCI, spinal cord injury; AIS, American Spinal Injury Association Impairment Scale; SD, standard deviation.

imbalance is visible, particularly in relation to those neurological level and injury severity subgroups where there is the greatest potential for neurological improvement; specifically AIS B and C high and low cervical categories.

Table 4 displays the numbers of patients with their anticipated mean motor score recovery for each AIS and anatomic region category in brackets for each category of the Canadian Classification for the two hypothetical study cohorts (Group 1 and Group 2). Anticipated motor recovery for each category in our 4×4 classification is shown in brackets and is obtained from Table 2. Im-

TABLE 4. HYPOTHETICAL RANDOMIZED CONTROLLED CLINICAL TRIAL WITH MEAN RECOVERY OF MOTOR SCORE (SD) DISPLAYED ACCORDING TO THE PROPOSED CLASSIFICATION SYSTEM

AIS	Neurological Level				Total number
	High cervical (C1-C4)	Low cervical (C5-T1)	Thoracic (T2-T10)	Thoracolumbar (T11-L2)	
A					
Group 1	1 (10)	1 (9)	4 (1)	1 (5)	7
Group 2	3 (10)	1 (9)	4 (1)	-	8
B					
Group 1	5 (30)	1 (24)	-	1 (9)	7
Group 2	1 (30)	2 (24)	1 (11)	3 (9)	7
C					
Group 1	4 (42)	6 (35)	1 (21)	2 (18)	13
Group 2	2 (42)	4 (35)	1 (21)	3 (18)	10
D					
Group 1	3 (16)	2 (12)	-	-	5
Group 2	4 (16)	3 (12)	-	-	7
Total Group 1	13	10	5	4	32
Total Group 2	10	10	6	6	32

balances that are not apparent in the far right column or bottom rows can be identified in the joint distribution, particularly in AIS B high cervical where there are four more patients in Group 1, compared with Group 2, each with a potential to improve by 29 motor points.

Group 1 has an additional two patients in each AIS C cervical group, each of which have 35–41 potential motor score improvements, thus explaining why the follow-up motor score in Group 1 is more than 10 points higher than that in Group 2. This demonstrates that small numerical imbalances within this joint 4×4 distribution (particularly in the subgroups with high potential for neurological motor improvement) can have a dramatic influence on neurological recovery in Group 1.

Further, we assessed the power of our hypothetical study in Figure 1 by performing sample size estimations. We calculated the sample size and power necessary to demonstrate a TMS improvement of 10. We estimated the standard deviation (SD) in the study groups based on our study population (SD = 15) and the literature (difference in SD = 10–35).¹⁶ The anticipated sample size to achieve a power of 0.8 would be 180 participants with random sampling of our cohort. When we repeated this exercise while stratifying for conventional characteristics (age, gender, AIS), we required 102 total participants. Finally, when we stratified subject selection based on the Canadian Classification, which considers the joint distribution of both neurological level and severity of injury together, 78 participants are required.

Discussion

We hypothesized that the magnitude of early motor recovery following acute tSCI can be more accurately estimated when both the neurological level and severity of the initial injury are used in a joint distribution to categorize these injuries. We selected four distinct neurological levels of injury—high cervical, low cervical, thoracic, and thoracolumbar—and used the patient's initial AIS grade as a measure of neurological severity of injury.

We analyzed the early motor recovery of 836 patients and verified that there are differences in early motor recovery between the four levels of injury that are unique to their initial AIS grades. Although it has been recognized that level and severity of injury are the two predominant predictors of neurological outcome,^{19,21} the strong influence of their joint distribution, our proposed categorization of neurological levels, and the impact this could have on RCT design and analysis of observational datasets have not been previously reported.

The significant impact of our study is that we identify the importance of the joint distribution of injury level and severity beyond what is commonly a simple univariate reporting of these characteristics in the conventional tables included under the category of study participant characteristics in most clinical trials. Unrecognized imbalances between categories within the Canadian Classification have plagued SCI translational efforts for many years.

We propose that all SCI trials should report on the level and severity of injury in our proposed joint distribution in their description of participant characteristics (Table 2).

While motor score improvements in some categories were very similar, patients with the same initial AIS grade, but a different neurological level had total motor score change differences that varied from a low of 5 motor points in AIS D to 20 motor points for AIS B and C (Table 2). This magnitude of difference in motor score improvement would be deemed to be clinically important.^{16,26} When the patient's initial AIS grade is A, B, or C, we have shown that mean early motor score improvements are significantly different for cervical and thoracic injuries.

FIG. 1. The sample size and power was calculated for a hypothetical study based on a 10-unit change in total motor score for three different sampling scenarios. For each graph, the associated standard deviations (SDs) for a ten unit change in motor score are reported based on data from the Rick Hansen Spinal Cord Injury Registry and other published papers. Curve 1 uses simple random sampling, Curve 2 stratifies for conventional patient characteristics (age, gender, American Spinal Injury Association Impairment Scale), and Curve 3 is based on stratifying patients based on the Canadian Classification.

We demonstrated that the high and low cervical levels of injury display different magnitudes of UEMS recovery in individuals with incomplete tSCI (AIS B, C, and D). Distinguishing between high and low cervical levels of injury reached significance for UEMS in AIS C and D, while a similar trend was seen in AIS B ($p=0.05$). High cervical injuries benefit from a larger number of segments available to recover motor function than would a low cervical injury of similar severity and this is particularly relevant for individuals with incomplete neurological injuries. The distinction between high and low cervical incomplete patients is not often specified in clinical trial reporting.^{22,23,29,31} Upper extremity motor recovery is a clinically and functionally important outcome³² and thus the significant differences in recovery potential between high and low cervical neurological levels should be distinguished and reported separately for each AIS, particularly when incomplete patients are included in a clinical trial or observational study.

We identified a significant difference in motor recovery between thoracic (T2-T10) and thoracolumbar (T11-L2) injuries when the initial AIS grade is A. This is not a surprising finding given the unique neuroanatomy of the region of the conus medullaris and cauda equina, and the way in which neurological recovery has been shown to be dependent upon this unique anatomy.³³

Our analysis defines the level of injury based on a validated and universally accepted standard, the ISCNSCI neurological level. We have subdivided the neurological level of injury into four groups that are intuitive based upon our understanding of neuroanatomy and physiology. We have used the validated and accepted AIS as a surrogate for severity of neurological injury.

We acknowledge that this study includes a wide range of baseline neurological examination times and does not report on long-term or final neurological outcome. There has been substantial debate on how early a baseline neurological examination can be reliably performed in neuroprotective trials, and this will be the subject of further

investigation by our group and others. Including the broader timing of baseline neurological examinations generalizes our results to acute neuroprotective and subacute neuroregenerative trials, although our intent was not to mimic a clinical trial, but to analyze registry data, which more closely represents actual clinical practice. It is estimated that over 70% of recovery occurs while the tSCI patient is still hospitalized, and the majority of rapid recovery occurs within the first three months following injury.^{16,34} Six-month neurological recovery has been used as a primary outcome in clinical trials and the majority of our patients had their follow-up neurological examination around the six month mark.

We also acknowledge that the time from injury to when the baseline neurological examination is performed can influence the magnitude and rapidity of motor recovery and this is the subject of a future study. This current analysis includes patients whose baseline examination is performed within 30 d of injury; however, we would anticipate that those with early neurological examinations (<24 h from injury) would demonstrate an even greater rate and magnitude of motor recovery than our current population. By including a broader range of baseline neurological examination times we have likely underestimated the effect of our classification on patients that are recruited into early intervention trials and have increased the generalizability of our results to acute and sub-acute intervention trials.

We believe that the national nature of our participant sample adds to the generalizability of our results, while the training of clinicians in the performance of the ISCNSCI standards adds to the reliability and accuracy of our results.³⁵ We also acknowledge that several of the categories in our classification are based on relatively small sample sizes. This is a consistent feature of all trials in SCI; despite national sampling, the numbers of some categories of tSCI, particularly incomplete injuries, are remarkably small. This reality is one that must be acknowledged by those who plan and design clinical trials.

We have provided an example of a hypothetical small clinical trial, similar in size to several reported clinical trials.^{18,28–30} Our hypothetical trial identified a significant difference in motor recovery between two groups by chance alone, a finding that is not at all surprising since we set our level of significance at 5% and thus we would expect one false positive result in every 20 tests. The fundamental reason for presenting this hypothetical trial is to demonstrate, using an example of “real-world data” drawn from our population, that by presenting the study data in the joint distribution of our proposed 4×4 classification of SCI, a reader, reviewer, or investigator is able to identify and quantify the imbalances between the two randomly selected cohorts. Essentially, we have identified subgroups of tSCI patients in which there are predictable and substantial differences in motor recovery. Between some subgroups, the differences are less profound. Our hypothetical example demonstrates the importance of the joint distribution of level and severity, and how imbalances between the cells of our proposed classification may not be readily apparent to clinical trialists when reporting study participants in the conventional fashion.

It is highly plausible that motor recovery occurs differently at the various levels of the spinal cord, and that the rate and magnitude of recovery also varies with the severity of initial injury. In addition to this variability, the tool that we use to measure this recovery (the ISNCSCI motor scale) records motor improvements differently depending upon the level and severity of injury. Given that the ISNCSCI only records ten motor segments each in the upper and lower extremities, a similar numerical magnitude of motor improvement may be distributed in very different ways in the same patient and between different patients, leading to a lack of predictability of what is clinically meaningful when a numerical change in motor score is reported.

Our hypothetical trial also suggests that the best strategy to avoid these imbalances is to stratify patients based on neurological level of injury and severity of neurological impairment. This will reduce the variability in the motor recovery outcome and also reduce the numbers needed for adequate power. Stratification may adversely influence the feasibility of recruitment, as well as the generalizability of a study. If clinical trials use this classification to report their study composition, then comparing datasets across clinical trials and observational studies would be facilitated.

It is likely that beneficial therapies may demonstrate an effect in only one or several of the anatomic and injury severity categories. Some therapies might demonstrate their effect in AIS A patients, while others may have a mechanism of action that is more effective for incomplete injuries. Further, with the varying contribution of grey and white matter to different neurological levels of spinal cord injury, a therapy directed at re-myelination would likely express a different clinical response than a cell-based or neuroprotective strategy.

It is only by acknowledging the heterogeneity of tSCI that clinicians will be able to personalize their approach to the treatment of tSCI. This study is simply a first step in understanding the heterogeneity inherent in tSCI. We anticipate that further observations, collaborations, validations, and replication will be required to refine these categories and how they are to be used to improve the methodology of tSCI observational and randomized clinical trials.

Acknowledgments

We would like to specifically acknowledge the two reviewers who provided substantial input that has improved the quality of this

manuscript. We would also like to acknowledge the Rick Hansen Spinal Cord Injury Registry Network, Karen Ethans, and Colleen O’Connell.

Author Disclosure Statement

Funding was provided by the Rick Hansen Institute and Health Canada. No competing financial interests exist.

References

1. Tator, C.H., Hashimoto, R., Raich, A., Norvell, D., Fehlings, M.G., Harrop, J.S., Guest, J., Aarabi, B., and Grossman, R.G. (2012). Translational potential of preclinical trials of neuroprotection through pharmacotherapy for spinal cord injury. *J. Neurosurg. Spine* 17(1 Suppl), 157–229.
2. Blight, A., Curt, A., Ditunno, J.F., Dobkin, B., Ellaway, P., Fawcett, J., Fehlings, M., Grossman, R.G., Lammertse, D.P., Privat, A., Steeves, J., Tuszynski, M., Kalichman, M., and Guest, J.D. (2009). Position statement on the sale of unproven cellular therapies for spinal cord injury The International Campaign for Cures of Spinal Cord Injury Paralysis. *Spinal Cord* 7, 713–714.
3. Bailey, C.S., Fisher, C.G., and Dvorak, M.F. (2004). Type II error in the spine surgical literature. *Spine* 29, 1146–1149.
4. Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., and Munafò, M.R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 65–76.
5. Steeves, J.D., Lammertse, D., Curt, A., Fawcett, J.W., Tuszynski, M.H., Ditunno, J.F., Dobkin, B.H., Havton, L.A., Ellaway, P.H., Fehlings, M.G., Privat, A., Grossman, R., Guest, J.D., Kleitman, N., Nakamura, M., Gaviria, M., and Short, D. (2006). Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: clinical trial outcome measures. *Spinal Cord* 45, 206–221.
6. Tator, C.H. (2006). Review of treatment trials in human spinal cord injury: issues, difficulties, and recommendations. *Neurosurgery* 59, 957–982.
7. Maas, A.I.R., Menon, D.K., Lingsma, H.F., Pineda, J.A., Sandel, M.E., and Manley, G.T. (2012). Re-orientation of clinical research in traumatic brain injury: report of an international workshop on comparative effectiveness research. *J. Neurotrauma* 29, 32–46.
8. (1998) Neuroprotection as initial therapy in acute stroke. Third Report of an Ad Hoc Consensus Group Meeting. The European Ad Hoc Consensus Group. *Cerebrovasc. Dis.* 8, 59–72.
9. Steyerberg, E.W., Bossuyt, P.M., and Lee, K.L. (2000). Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? *Am. Heart J.* 139, 745–751.
10. Lee, R.S.R., Noonan, V.K., Batke, J., Ghag, A., Paquette, S.J., Boyd, M.C., Fisher, C.G., Street, J., Dvorak, M.F., and Kwon, B.K. (2012). Feasibility of patient recruitment into clinical trials of experimental treatments for acute spinal cord injury. *J. Clin. Neurosci.* 19, 338–343.
11. Anderson, K.D. (2004). Targeting recovery: priorities of the spinal cord-injured population. *J. Neurotrauma* 21, 1371–1383.
12. Scivoletto, G., Tamburella, F., Laurenza, L., and Molinari, M. (2013). Distribution-based estimates of clinically significant changes in the international standards for neurological classification of spinal cord injury motor and sensory scores. *Eur. J. Phys. Rehabil. Med.* 49, 373–384.
13. Kirshblum, S.C., Burns, S.P., Biering-Sorensen, F., Donovan, W., Graves, D.E., Jha, A., Johansen, M., Jones, L., Krassioukov, A., Mulcahey, M.J., Schmidt-Read, M., and Waring, W. (2011). International standards for neurological classification of spinal cord injury (Revised 2011). *J. Spinal Cord Med.* 34, 535–546.
14. Tuszynski, M.H., Steeves, J.D., Fawcett, J.W., Lammertse, D., Kalichman, M., Rask, C., Curt, A., Ditunno, J.F., Fehlings, M.G., Guest, J.D., Ellaway, P.H., Kleitman, N., Bartlett, P.F., Blight, A.R., Dietz, V., Dobkin, B.H., Grossman, R., and Privat, A.; International Campaign for Cures of Spinal Cord Injury Paralysis. (2007). 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* 45, 222–231.
15. Steeves, J.D., Kramer, J.K., Fawcett, J.W., Cragg, J., Lammertse, D.P., Blight, A.R., Marino, R.J., Ditunno, J.F. Jr., Coleman, W.P., Geisler, F.H., Guest, J., Jones, L., Burns, S., Schubert, M., and van Hedel, H.J., Curt, A.; EMSCI Study Group. (2011). Extent of spontaneous motor

- recovery after traumatic cervical sensorimotor complete spinal cord injury. *Spinal Cord* 49, 257–265.
16. Fawcett, J.W., Curt, A., Steeves, J.D., Coleman, W.P., Tuszynski, M.H., Lammertse, D., Bartlett, P.F., Blight, A.R., Dietz, V., Ditunno, J., Dobkin, B.H., Havton, L.A., Ellaway, P.H., Fehlings, M.G., Privat, A., Grossman, R., Guest, J.D., Kleitman, N., Nakamura, M., Gaviria, M., and Short, D. (2007). 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* 45, 190–205.
 17. Lammertse, D., Tuszynski, M.H., Steeves, J.D., Curt, A., Fawcett, J.W., Rask, C., Ditunno, J.F., Fehlings, M.G., Guest, J.D., Ellaway, P.H., Kleitman, N., Blight, A.R., Dobkin, B.H., Grossman, R., Katoh, H., Privat, A., and Kalichman, M.; International Campaign for Cures of Spinal Cord Injury Paralysis. (2007). Guidelines for the conduct of clinical trials for spinal cord injury as developed by the ICCP panel: clinical trial design. *Spinal Cord* 45, 232–42.
 18. Geisler, F.H.F., Dorsey, F.C.F., and Coleman, W.P.W. (1991). Recovery of motor function after spinal-cord injury—a randomized, placebo-controlled trial with GM-1 ganglioside. *N. Engl. J. Med.* 324, 1829–1838.
 19. Coleman, W.P. and Geisler, F.H. (2004). Injury severity as primary predictor of outcome in acute spinal cord injury: retrospective results from a large multicenter clinical trial. *Spine J.* 4, 73–78.
 20. Al-Habib, A.F., Attabib, N., Ball, J., Bajammal, S., Casha, S., and Hurlbert, R.J. (2011). Clinical predictors of recovery after blunt spinal cord trauma: systematic review. *J. Neurotrauma* 28, 1431–1443.
 21. Steeves, J., Fawcett, J., and Tuszynski, M. (2004). Report of International Clinical Trials Workshop on Spinal Cord Injury February 20–21, 2004, Vancouver, Canada. *Spinal Cord* 42, 591–597.
 22. Geisler, F.H., Coleman, W.P., Grieco, G., and Poonian, D.; Sygen Study Group. (2001). The Sygen multicenter acute spinal cord injury study. *Spine* 26(24 Suppl), S87–S98.
 23. Fehlings, M.G., Vaccaro, A., Wilson, J.R., Singh, A.W., Cadotte, D., Harrop, J.S., Aarabi, B., Shaffrey, C., Dvorak, M., Fisher, C., Arnold, P., Massicotte, E.M., Lewis, S., and Rampersaud, R. (2012). Early versus delayed decompression for traumatic cervical spinal cord injury: results of the Surgical Timing in Acute Spinal Cord Injury Study (STASCIS). *PLoS ONE* 7, e32037.
 24. Blight, A.R. and Tuszynski, M.H. (2006). Clinical trials in spinal cord injury. *J. Neurotrauma* 23, 586–593.
 25. Noonan, V.K., Kwon, B.K., Soril, L., Fehlings, M.G., Hurlbert, R.J., Townson, A., Johnson, M., Dvorak, M.F.; RHSCIR Network. (2011). The Rick Hansen Spinal Cord Injury Registry (RHSCIR): a national patient-registry. *Spinal Cord* 50, 22–27.
 26. Furlan, J.C., Noonan, V., Singh, A., and Fehlings, M.G. (2011). Assessment of impairment in patients with acute traumatic spinal cord injury: a systematic review of the literature. *J. Neurotrauma* 28, 1445–1477.
 27. Geisler, F.H., Coleman, W.P., Grieco, G., and Poonian, D.; Sygen Study Group. (2001). Measurements and recovery patterns in a multicenter study of acute spinal cord injury. *Spine* 26(24 Suppl), S68–S86.
 28. Fehlings, M.G., Theodore, N., Harrop, J., Maurais, G., Kuntz, C., Shaffrey, C.I., Kwon, B.K., Chapman, J., Yee, A., Tighe, A., and McKerracher, L. (2011). A phase I/IIa clinical trial of a recombinant rho protein antagonist in acute spinal cord injury. *J. Neurotrauma* 28, 787–796.
 29. Casha, S., Zygun, D., McGowan, M.D., Bains, I., Yong, V.W., and Hurlbert, R.J. (2012). Results of a phase II placebo-controlled randomized trial of minocycline in acute spinal cord injury. *Brain* 135, 1224–1236.
 30. Lammertse, D.P., Jones, L.A., Charlifue, S.B., Kirshblum, S.C., Apple, D.F., Ragnarsson, K.T., Falci, S.P., Heary, R.F., Choudhri, T.F., Jenkins, A.L., Betz, R.R., Poonian, D., Cuthbert, J.P., Jha, A., Snyder, D.A., and Knoller, N. (2012). Autologous incubated macrophage therapy in acute, complete spinal cord injury: results of the phase 2 randomized controlled multicenter trial. *Spinal Cord* 50, 661–671.
 31. Bracken, M.B., Shepard, M.J., Collins, W.F., Holford, T.R., Young, W., Baskin, D.S., Eisenberg, H.M., Flamm, E., Leo-Summers, L., Maroon, J., Marshall, L.F., Perot, P.L., Jr., Piepmeier, J., Sonntag, V.K.H., Wagner, F.C., Wilberger, J.E., and Winn, H.R. (1990). 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.* 322, 1405–1411.
 32. Marino, R.J. and Graves, D.E. (2004). Metric properties of the ASIA motor score: subscales improve correlation with functional activities. *Arch. Phys. Med. Rehab.* 85, 1804–1810.
 33. Kingwell, S.P., Noonan, V.K., Fisher, C.G., Graeb, D.A., Keynan, O., Zhang, H., and Dvorak, M.F. (2010). Relationship of neural axis level of injury to motor recovery and health-related quality of life in patients with a thoracolumbar spinal injury. *J. Bone Joint Surg. Am.* 92, 1591–1599.
 34. Pollard, M.E. and Apple, D.F. (2003). Factors associated with improved neurologic outcomes in patients with incomplete tetraplegia. *Spine* 28, 33–39.
 35. Schuldt, C., Wiese, J., Franz, S., Putz, C., Stierle, I., Smoor, I., and Weidner, N.; EMSCI Study Group, Rupp R. (2013). Effect of formal training in scaling, scoring and classification of the International Standards for Neurological Classification of Spinal Cord Injury. *Spinal Cord* 51, 282–288.

Address correspondence to:
 Marcel F. Dvorak, MD, FRCS
 Department of Orthopaedics
 University of British Columbia
 6th Floor, Blusson Spinal Cord Center
 818 West 10th Avenue
 Vancouver, British Columbia, Canada V5Z 1M9
 E-mail: Marcel.Dvorak@ubc.ca