ISCoS
The International
Spinal Cord Society

ARTICLE

Clinical judgment is a cornerstone for validating and using clinical prediction rules: a head-to-head study on ambulation outcomes for spinal cord injured patients

Rémi Pelletier-Roy[1,2] · Andréane Richard-Denis [1,2] · Stéphanie Jean[1,3] · Étienne Bourassa-Moreau[1,2] · Jean Fleury[1,3] · Geneviève Beauchamp-Vien[2] · Jean Bégin[2] · Jean-Marc Mac-Thiong [1,2]

Abstract

Study design Retrospective comparative study.
Objective Clinical prediction rules (CPRs) are an effervescent topic in the medical literature. Recovering ambulation after a traumatic spinal cord injury (tSCI) is a priority for patients and multiple CPRs have been proposed for predicting ambulation outcomes. Our objective is to confront clinical judgment to an established CPR developed for patients with tSCI.
Settings Level one trauma center specialized in tSCI and its affiliated rehabilitation center.
Method In this retrospective comparative study, six physicians had to predict the ambulation outcome of 68 patients after a tSCI based on information from the acute hospitalization. Ambulation was also predicted according to the CPR of van Middendorp (CPR-vM). The success rate of the CPR-vM and clinicians to predict ambulation was compared using criteria of 5% for defining clinical significance, and a level of statistical significance of 0.05 for bilateral McNemar tests.
Results There was no statistical difference between the overall performance of physicians (success rate of 79%) and of the CPR-vM (81%) for predicting ambulation. The differences between the CPR-vM and physicians varied clinically and significantly with the level of experience, clinical setting, and field of expertise.
Conclusion Confronting CPRs with the judgment of a group of clinicians should be an integral part of the design and validation of CPRs. Head-to-head comparison of CPRs with clinicians is also a cornerstone for defining the optimal strategy for translation into the clinical practice, and for defining which clinician and specific clinical context would benefit from using the CPR.

Introduction

Clinical prediction rules (CPRs) are an effervescent topic in many medical and non-medical fields. Comparing CPRs to unstructured clinician judgment alone is often recommended to measure the impact of CPR [1]. However, the optimal methodology for comparing CPRs to clinical judgment remains unclear such that very few CPRs have

been confronted with clinical judgment [2]. Sanders et al. [2] showed that only 25 diagnostic CPRs have been compared to clinical judgment. They also underlined the high heterogeneity in the methodology used to compare CPRs to clinical judgment [2]. Furthermore, the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) checklist on the development and validation of CPR only put little emphasis on comparing CPR to clinicians' performance [3].

These shortcomings are ubiquitous in some fields such as traumatic spinal cord injury (tSCI) where there is no report of CPR compared with unstructured clinical judgment. Prediction of long-term prognosis is of paramount importance after tSCI since it is a devastating event affecting up to 500,000 individuals annually worldwide [4]. TSCI involves a heterogeneous group of patients with unpredictable outcomes with variable lifelong limitations in motor, sensory, bladder, bowel, cardiovascular, and respiratory functions.

✉ Jean-Marc Mac-Thiong
jean-marc.mac-thiong@umontreal.ca

1 Université de Montréal, Faculty of Medicine, Montréal, QC, Canada

2 Hôpital du Sacré-Cœur de Montréal, Montréal, QC, Canada

3 Institut de réadaptation Lindsay-Gingras de Montréal, Montréal, QC, Canada

Several prognostic CPRs have been proposed [5–10], particularly to predict ambulation outcomes since it is the main priority for patients [11, 12]. Van Middendorp et al. [5] have proposed a CPR with excellent discrimination for predicting the ability of patients to walk independently one year after a tSCI. This CPR is considered as a reference model for predicting walking recovery, and its statistical performance has been further validated by independent authors [13, 14]. However, it has never been confronted with clinical judgment at any stage of its development thus limiting its translation to the medical community.

Our objective is to compare the ability of clinicians to predict independent household ambulation after severe tSCI with the CPR developed by van Middendorp et al. (CPR-vM). Our hypothesis is that the CPR-vM is more accurate than physicians, and should be used routinely in clinical practice.

Methods

Study cohort

The study cohort was derived from a prospective database of 458 tSCI patients treated at a single Level-1 trauma center specialized in acute SCI care between April 2010 and December 2018. The following inclusion criteria were used to identify eligible patients: (1) age 16 years or older, (2) severe tSCI with American Spinal Injury Association impairment scale (AIS) grade A to C, (3) neurological level of injury between C1 and L2, and (4) household ambulatory status assessed from item 12 of the 3rd version of the Spinal Cord Independence Measure (SCIM) one year after the tSCI. Of the 152 eligible patients, a study cohort of 68 patients was randomly retrieved for analysis. AIS grade D patients were excluded to prevent inadequate high prediction accuracy from clinicians. Many AIS grade D patients could already be considered as independent walkers at the time of injury and <1% AIS grade D patients were not independent walkers at 1 year after the injury. Also, excluding levels of injury from L3 to S5 was done to remove cauda equina syndrome patients. Cauda equina injuries are distinctive in terms of inferior motor neuron deficits and prognosis and were excluded from this present study.

Participating clinicians

Six physicians who do not routinely use the CPR-vM in their practice and involved in communicating the prognosis to tSCI patients were enrolled for this study. We chose a group that represents the clinicians involved in both acute and subacute care of our spinal cord injured patients in

order to reflect the typical attending physicians involved with SCI patients within the 1st month after the injury. In our center, this represents mainly the orthopaedic and physical medicine and rehabilitation (PM&R) clinicians. Our idea to select different levels of experience within each group (resident, junior staff, and senior staff) was to identify if there were any difference in prediction accuracy that would be the best address with the use of the clinical prediction rule:

1. Senior attending (>10 years of practice) in physical medicine and rehabilitation (PM&R) specialized in inpatient functional rehabilitation after tSCI,
2. Junior attending (<10 years of practice) in PM&R specialized in the acute care of tSCI prior to functional rehabilitation,
3. Postgraduate year-4 resident in PM&R with training in inpatient functional rehabilitation after tSCI, and to a lesser extent in the acute care of tSCI,
4. Senior attending (>10 years of practice) in orthopaedic surgery specialized in the surgical treatment of tSCI,
5. Junior attending (<10 years of practice) in orthopaedic surgery specialized in the surgical treatment of tSCI, and
6. Postgraduate year-2 resident in orthopaedic surgery.

Data collection

All participating physicians were consulted in order to identify the information from the acute hospitalization that they use to establish the long-term prognosis on ambulation. To reflect their actual practice, all physicians were therefore provided with: the initial consultation notes from spine surgery, PM&R and occupational therapy (including pre-tSCI functional status), patient's age, comorbidities and past medical/surgical history, surgical protocol, preoperative and early postoperative (within one week after surgery) International Standards for Neurological Classification of SCI (ISNCSCI) worksheet detailing the initial neurological examinations, and all imaging reports pertaining to the spine and/or spinal cord (MRI, CT scans, X-rays) at admission to acute care. Medical chart reprints were collected and anonymized by a research assistant not involved in the study design, assessments nor data analyses.

Main outcomes measures

The main outcome measures consisted of the patient's ability to walk indoors independently without supervision (independent household ambulation) 1 year after the tSCI. The actual ability of patients to walk indoors was obtained

from the patient's answer to item 12 of the SCIM (Indoors Mobility) at the one-year follow-up visit. Item 12 includes nine possible answers scored from zero (Requires total assistance) to eight (Walks without walking aids). In accordance with van Middendorp et al. [5], independent walking was defined as a score between four (Walks with a walking frame or crutches) and eight, as opposed to a score between zero and three (Requires supervision while walking) for patients unable to walk or dependent on assistance for walking.

CPR-vM is a validated clinical prediction rule on ambulation outcome 1 year after tSCI, which includes age and four neurological tests. Age is dichotomized at 65-years-old and score either zero or one. Motor grades of the quadriceps femoris muscle (L3) and of the gastrocsoleus muscle (S1) are scored from zero to five. Light touch sensation of L3 and S1 dermatomes are scored from zero to two. Every score has a weighted coefficient, which gives a final score between -10 and 40. This score is used in CPR-vM equation and gives a percentage of having a SCIM item 12 score of four or more, which corresponds to independent walking. For example, a 60-year-old patient with motor grades of three and light touch sensory grades of one for both L3 and S1 would have a 35% probability of having a SCIM item 12 equal to or above four, thus would be considered as not an independent walker ($<50\%$). Predicting independent walking from the CPR-vM was performed by an independent statistician based on patient's age, as well as the four neurological variables retrieved from the early postoperative ISNCSCI worksheet.

All participating physicians predicted the ability of patients to recover independent walking based solely on unstructured clinical judgment, using the medical chart reprints and the criteria previously described for item 12 of the SCIM.

Statistics

The performance of the CPR-vM and physicians to correctly predict independent walking was assessed from the percentage of accurate predictions ((number of true positive + number of true negative)/68 cases). The performance of physicians was calculated individually and the overall physicians' performance was obtained by averaging all individual physician's performance. Descriptive statistics were performed to characterize the study cohort as well as the performance of the CPR-vM and physicians.

We used bilateral McNemar tests to compare the performance of CPR-vM and physicians, using a statistical level of significance of 0.05. In addition, a threshold of 5% was used to define a clinically significant difference in performance between the CPR-vM and physicians'

judgment. This clinical level of significance was determined by consensus among all physicians involved in this study in order to carry out the sample size calculation. This threshold of 5% was defined by clinicians as the minimal improvement in performance for which they would consider adopting a CPR in their practice.

The sample size of 68 patients was calculated to provide 80% power to detect a clinically significant difference of 5% in performance between CPR-vM and physicians' judgment, using bilateral McNemar tests and a statistical level of significance of 0.05. The sample size calculation was performed using G*Power 3.1.9.6 (Düsseldorf, Germany), while statistical tests were done using IBM SPSS (Version 25.0, NY, USA).

Results

The characteristics of the study cohort are presented in Table 1. Physicians accurately predicted ambulation status one year after tSCI in 79% of the cases. This was not statistically different from the performance of the CPR-vM of 81%. The individual performance of physicians ranged between 71% and 85% for predicting all the cases (Table 2). The orthopaedic group was clinically more accurate than the PM&R group (83% vs. 76%). The performance was similar between the CPR-vM and all individual orthopaedic

Table 1 Baseline characteristics of study cohort.

	Study Cohort ($n = 68$)
Settings	Level one trauma center specialized in tSCI
Inclusion period	April 2010 to December 2018
Sex male n (%)	54 (79%)
Mean age at injury years (±SD)	44 (±18)
Severity of initial neurological deficit n (% patients)	
AIS grade A	28 (41%)
AIS grade B	21 (31%)
AIS grade C	19 (28%)
Level of initial neurological deficit n (% patients)	
High cervical (C1-C4)	24 (35%)
Low cervical (C5-T1)	8 (12%)
High thoracic (T2-T10)	19 (28%)
Low thoracic (T11-L2)	17 (25%)
Outcomes measurements	
SCIM item 12 mean score (±IQR)	3.6 (±4)
Independent household ambulation n (% patients)	25 (37%)

tSCI traumatic spinal cord injury, *AIS* American spinal injury association (ASIA) impairment scale, *SD* standard deviation.

Table 2 Accuracy of clinical prediction rule and clinicians stratified by AIS grade.

	Clinical prediction rule	PM&R resident	PM&R junior attending	PM&R senior attending	Orthopaedic surgery resident	Orthopaedic surgery junior attending	Orthopaedic surgery senior attending
AIS grade A	89%	71%	86%	86%	89%	89%	89%
AIS grade B	76%	71%	81%	81%	86%	81%	86%
AIS grade C	74%	68%	84%	42%	63%	84%	68%
All grades	81%	71%[a]	84%[a]	72%[a]	81%	85%	82%

AIS American spinal injury association (ASIA) impairment scale, *PM&R* physical medicine and rehabilitation.

[a]$p < 0.05$ for comparison with clinical prediction rule.

physicians, and there was no difference among the orthopaedic group with regard to the level of experience (junior resident vs. junior attending vs. senior attending). The CPR-vM was clinically and statistically more accurate than the PM&R physicians with predominant expertise in functional rehabilitation, who were respectively 10% and 9% less accurate than the CPR-vM (respective *p* values of 0.007 and 0.04). On the contrary, the PM&R attending specialized in acute tSCI care working in a different center was statistically more accurate than the CPR-vM (84% vs. 81%, $p = 0.039$), but the difference did not reach clinical significance.

The performance of the CPR-vM and physicians was also stratified based on the AIS grade of patients (Table 2). The performance was consistently highest for patients with the most severe tSCI (AIS grade A) and lowest for those with the least severe tSCI (AIS grade C), for the CPR-vM and all physicians. The largest discrepancies occurred between the CPR-vM and physicians for the prediction of AIS grade C patients, with differences ranging from a 32% increase for the CPR-vM over the PM&R attending working at the rehabilitation facility, to a 10% decrease for the CPR-vM compared to junior attendings in PM&R and orthopaedic surgery working at the acute care facility.

Discussion

Our study is the first to confront unstructured clinical judgment to a CPR intended for the tSCI population, supporting the prime importance of validating CPRs through a head-to-head comparison with clinicians to facilitate their translation into practice. While comparing with unstructured clinical judgment is generally recommended in the later stages of the development of CPRs following external statistical validation [1], assessing unstructured clinical judgment can also be invaluable during the initial stage for identifying the need for a CPR and orienting the subsequent stages of development. By no means this paper intends to assess the relevance, performance, or accuracy of the van Middenddorp rule, but instead, this paper intends to lay the ground for the adequate methodological aspects to consider

for validating and using any clinical prediction rule, as well as for identifying whether clinicians should use dedicated prediction rules or undergo additional training.

The results show that the performance of the CPR-vM was similar to the overall performance of a team of physicians involved in communicating prognosis to tSCI patients in their clinical practice. Physician performance was mainly dependent on the field of expertise and clinical setting. The contrasting performances between all physicians involved in this study strongly suggest that the CPR-vM could represent a benchmark for establishing standards of practice and tailoring the training needs of clinicians, in addition, to target clinical knowledge users best suited for the CPR-vM. Physicians with a predominant practice in a functional rehabilitation facility and thus not involved in the acute care of tSCI did not perform at the level of the CPR-vM. It is possible that the physicians specialized in functional rehabilitation and ambulation training are strongly influenced by their actual practice that predominantly involves patients with severe limitations, and is, therefore, less likely to follow patients with favorable outcomes in the long term. In addition, physicians from functional rehabilitation facilities are typically more concerned with community ambulation than household ambulation in their practice. Surprisingly, the experience level was not as impactful as the field of expertise in this study, since physicians working at the same institution demonstrated similar performances regardless of their experience level.

Previous studies showed that the performance of the CPR-vM was dependent on the AIS grade [15]. Accordingly, we have decided to exclude patients with AIS grade D tSCI in this study because it is well known that the great majority of these patients will walk indoors without supervision (<1% AIS D not walking in our database). Our results showed a lower performance of both CPR-vM and physicians for cases with severe but incomplete tSCI, with a particular decline and variability in performance for AIS grade C lesions. Accordingly, future developments of CPR predicting ambulation in tSCI patients should preferably target patients with sensory and non-functional motor lesions, respectively AIS grades B and C tSCI.

Strengths and limitations

This work highlights key methodological concepts for minimizing the heterogeneity of studies assessing unstructured clinical judgment. In summary, we propose a systematic approach comprising these three fundamental steps.

1. **Build a representative study cohort with adequate sample size**

    As a first step, it is imperative to obtain an adequately powered sample of patients that is representative of the patient population targeted by the CPR. To achieve this task, clinicians—with existing literature if present—should be consulted in order to define the minimal performance improvement provided by the CPR (level of clinical significance) that would promote the clinical implementation of the CPR.

2. **Assemble a representative group of clinicians**

    A representative group of clinicians prone to use the CPR should be assembled. At least these three variables need to be accounted for when selecting participating clinicians: their clinical setting (or institution), field of expertise (or prior training), and their level of experience. In addition to defining the level of clinical significance, clinicians also have a key role in identifying the information (e.g., medical chart, imaging studies, etc.) that will be given to participants to achieve the predictions in order to replicate the actual clinical practice.

3. **Assess performance based on clinical and statistical significance**

    While comparing the CPR to the overall performance for the entire group of clinicians is required, it is also important to assess the individual performance of clinicians. This step is important for measuring the clinical validity of the CPR based on variables such as the field of expertise, clinical setting, and level of experience, and thereby determining the applicability of the CPR for clinical translation. Although different metrics can be used to compare the performance of CPRs to clinical judgment, one convenient method is to assess the percentage of accurate predictions ((number of true positive + number of true negative)/ total number of cases). When using this metric, it becomes easier for clinicians to define a clinical level of significance based on the clinical needs, target population and predicted outcome. The performance of the CPR and clinicians should also be stratified according to clinically relevant subgroups of patients for identifying the optimal areas for clinical translation of the CPR.

A key weakness is the small number of physicians participating in the study, thus limiting the variability in setting, the field of expertise and clinical experience. However, we were still able to draw conclusions on the need for tailored use of the CPR-vM for a specific subgroup of physicians. The small number of patients is another recognized limitation of the study although the sample size was estimated a priori for reaching adequate power, and was sufficient to observe significant differences between the CPR-vM and physicians' judgment. Moreover, the study cohort was randomly retrieved from our local tSCI database and was representative of the general tSCI population [16]. Accordingly, there was a larger number of AIS grade A patients, and a similar proportion of AIS grades B and C patients, which allowed to assess the performance of the CPR-vM and physicians based on the AIS grade. A potential information bias would be clinicians referencing the items used in CPR-vM. As clinicians working in specialized SCI centers, all our clinicians are well aware of the principles of the CPR-vM, but they do not use the clinical prediction rule per se in their practice. We agree that the risk of a selective bias for referencing medical chart information (i.e., selecting the information related to the CPR-vM) is present, but we think it is minor since the results were significantly different from the CPR-vM. Also, even knowing the items of the prediction rule doesn't allow the clinician to calculate the ambulation prediction easily. In fact, because of the mathematics associated with CPR-vM, even knowing the items don't translate into prediction percentage. Regarding the comparability of our cohort and the one of van Middendorp, we acknowledge the difference between our inclusion criteria set at 16 years or older and the inclusion criteria of van Middendorp set at 18 years or older. Our patients were randomly chosen within our entire database of traumatic spinal cord injured patients treated at our level one trauma center. Our adult trauma center admits patients aged 16 years and older because they are presumed to have the capacity to make health care decisions and consent to care, such that parental or tutor consent is not necessary [17]. In our institution, individuals aged between 16 and 18 years old undergo the same treatments and continuum care as patients 18 years and older. There was one patient aged 16 years 4 months old at the time of injury included in our cohort. This patient was physically and skeletally mature upon admission to our institution. He has been managed with the same treatments and continuum care as all patients admitted to our institution. Our study design included a sample size calculation and a statistical plan—an important take-home message of the study—, and removing this subject from our analysis or replacing it a posteriori by another subject would inherently introduce a bias. Although we agree that clinicians will not be tempted to use the van

Middendorp rule for a patient aged 16 years and 4 months, this should not distract the reader from the principal objective of the study which is not to assess the relevance, performance or accuracy of the van Middenddorp rule. Finally, another weakness could be the fact that we didn't analyze explicitly which factors increased clinician accuracy or which information each clinician individually used in what was provided in the medical chart. We agree that it should be evaluated in another subsequent study for the benefit of the scientific community, but should not distract from the principal aim of our study to lay the ground for the adequate methodological aspects to consider for validating and using any clinical prediction rule, as well as for identifying whether clinicians should use dedicated prediction rules or undergo additional training.

Conclusions

Despite the excellent discriminative ability of the CPR-vM on a statistical basis, this landmark CPR developed in the field of tSCI was not consistently superior to the clinical judgment of physicians for predicting ambulation outcomes. The benefit of the CPR-vM over clinical judgment was mainly dependent on the clinical setting and field of expertise of physicians, in addition to the severity of the tSCI. The findings of this study underline the prime importance of comparing CPRs with unstructured clinical judgment using a systematic approach. While head-to-head comparison between CPR and clinicians' judgment is an integral part of the development and clinical translation of CPRs, it can also be used to establish standards of practice and tailor the training of clinicians.

Data availability

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

revision and final acceptance of the paper. JB participated in the study design, data collection, statistical analysis, and revision and final acceptance of the paper. EBM, JF, and SJ participated in data measurement, data analysis, and revision and final acceptance of the paper. GBV participated in data analysis and revision and final acceptance of the paper.

Compliance with ethical standards

References

1. Cowley LE, Farewell DM, Maguire S, Kemp AM. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. Diagn Progn Res. 2019;3:16.
2. Sanders S, Doust J, Glasziou P. A systematic review of studies comparing diagnostic clinical prediction rules with clinical judgment. PLoS One. 2015;10:e0128233.
3. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
4. World Health Organization, International Spinal Cord Society, International Perspectives on Spinal Cord Injury. Geneva, Switzerland: World Health Organization. 2013.
5. van Middendorp JJ, Hosman AJF, Donders ART, Pouw MH, Ditunno JF Jr, Curt A, et al. A clinical prediction rule for ambulation outcomes after traumatic spinal cord injury: a longitudinal cohort study. Lancet.2011;377:1004–10.
6. Dalkilic T, Fallah N, Noonan VK, Salimi Elizei S, Dong K, Belanger L, et al. Predicting injury severity and neurological recovery after acute cervical spinal cord injury: a comparison of cerebrospinal fluid and magnetic resonance imaging biomarkers. J Neurotrauma. 2018;35:435–45.

7. Wilson JR, Grossman RG, Frankowski RF, Kiss A, Davis AM, Kulkarni AV, et al. A clinical prediction model for long-term functional outcome after traumatic spinal cord injury based on acute clinical and imaging factors. J Neurotrauma. 2012;29: 2263–71.

8. Kaminski L, Cordemans V, Cernat E, M'Bra KI, Mac-Thiong JM. Functional outcome prediction after traumatic spinal cord injury based on acute clinical factors. J Neurotrauma. 2017;34:2027–33.

9. Facchinello Y, Beausejour M, Richard-Denis A, Thompson C, Mac-Thiong JM. The use of regression tree analysis for predicting the functional outcome following traumatic spinal cord injury. J Neurotrauma. 2017; https://doi.org/10.1089/neu.2017.5321.

10. Belliveau T, Jette AM, Seetharama S, Axt J, Rosenblum D, Larose D, et al. Developing artificial neural network models to predict functioning one year after traumatic spinal cord injury. Arch Phys Med Rehabil. 2016;97:1663–8.e3.

11. Simpson LA, Eng JJ, Hsieh JT, Wolfe DL. The health and life priorities of individuals with spinal cord injury: a systematic review. J Neurotrauma. 2012;29:1548–55.

12. Goulet J, Richard-Denis A, Thompson C, Mac-Thiong JM. Relationships between specific functional abilities and health-related quality of life in chronic traumatic spinal cord injury. Am J Phys Med Rehabil. 2019;98:14–19.

13. Hicks KE, Zhao Y, Fallah N, Rivers CS, Noonan VK, Plashkes T, et al. A simplified clinical prediction rule for prognosticating independent walking after spinal cord injury: a prospective study from a Canadian multicenter spinal cord injury registry. Spine J. 2017;17:1383–92.

14. van Silfhout L, Peters AE, Graco M, Schembri R, Nunn AK, Berlowitz DJ. Validation of the Dutch clinical prediction rule for ambulation outcomes in an inpatient setting following traumatic spinal cord injury. Spinal Cord. 2016;54:614–8.

15. Phan P, Budhram B, Zhang Q, Rivers CS, Noonan VK, Plashkes T, et al. Highlighting discrepancies in walking prediction accuracy for patients with traumatic spinal cord injury: an evaluation of validated prediction models using a Canadian Multicenter Spinal Cord Injury Registry. Spine J. 2019;19:703–10.

16. NSCISC National Spinal Cord Injury Statistical Center. Spinal Cord Injury Model Systems. Annual Report—Complete Public Version. 2019.

17. Coughlin KW. Medical decision-making in paediatrics: infancy to adolescence. Paediatr Child Health. 2018;23:138–46.