



Validation of the AO Spine CROST (Clinician Reported Outcome Spine Trauma) in the clinical setting

Said Sadiqi¹ · Erin E. A. de Gendt¹ · Sander P. J. Muijs¹ · Marcel W. M. Post^{2,3} · Lorin M. Benneker⁴ · Martin Holas⁵ · Jin W. Tee^{6,7} · Christoph E. Albers⁴ · Sonja Häckel⁴ · Juraj Svac⁵ · Richard J. Bransford⁸ · Mohammad M. El-Sharkawi⁹ · Frank Kandziora¹⁰ · Shanmuganathan Rajasekaran¹¹ · Klaus J. Schnake¹² · Alexander R. Vaccaro¹³ · F. Cumhur Oner¹

Received: 27 June 2023 / Revised: 19 September 2023 / Accepted: 13 January 2024
© The Author(s) 2024

Abstract

Purpose To evaluate feasibility, internal consistency, inter-rater reliability, and prospective validity of AO Spine CROST (Clinician Reported Outcome Spine Trauma) in the clinical setting.

Methods Patients were included from four trauma centers. Two surgeons with substantial amount of experience in spine trauma care were included from each center. Two separate questionnaires were administered at baseline, 6-months and 1-year: one to surgeons (mainly CROST) and another to patients (AO Spine PROST—Patient Reported Outcome Spine Trauma). Descriptive statistics were used to analyze patient characteristics and feasibility, Cronbach's α for internal consistency. Inter-rater reliability through exact agreement, Kappa statistics and Intraclass Correlation Coefficient (ICC). Prospective analysis, and relationships between CROST and PROST were explored through descriptive statistics and Spearman correlations.

Results In total, 92 patients were included. CROST showed excellent feasibility results. Internal consistency ($\alpha = 0.58$ – 0.70) and reliability (ICC = 0.52 and 0.55) were moderate. Mean total scores between surgeons only differed 0.2 – 0.9 with exact agreement 48.9 – 57.6% . Exact agreement per CROST item showed good results (73.9 – 98.9%). Kappa statistics revealed moderate agreement for most CROST items. In the prospective analysis a trend was only seen when no concerns at all were expressed by the surgeon (CROST = 0), and moderate to strong positive Spearman correlations were found between CROST at baseline and the scores at follow-up ($r_s = 0.41$ – 0.64). Comparing the CROST with PROST showed no specific association, nor any Spearman correlations ($r_s = -0.33$ – 0.07).

Conclusions The AO Spine CROST showed moderate validity in a true clinical setting including patients from the daily clinical practice.

Keywords Outcome instrument · Clinician perspective · AO Spine CROST · Spine trauma · Health · Function

Introduction

The influence of spine fractures on patients' functioning, including social and financial situation, is considered very significant compared to other injuries [1]. Currently, the decision-making between non-operative management and surgical care is far from settled for various types of spine fractures. In this perspective, measurement of outcomes is relevant in order to compare different treatment options, and thereby develop more rational choices for treatment strategies [2].

To address this void, the AO Spine Knowledge Forum Trauma developed the first disease-specific outcome

measure for spine trauma patients, the Patient Reported Outcome Spine Trauma (AO Spine PROST) [3]. An important note is that there may be discrepancies when comparing patients' perspective with clinicians' perspective on what is considered as a good outcome of a specific treatment [4, 5]. It is imperative to also capture the perspective of the clinicians in a simple, reliable and quick to administer tool. Including the most relevant clinical and radiological parameters, this tool would be able to evaluate and predict clinical outcomes of spine trauma patients. This led to the development of a separate, unique tool that is rated by clinicians: the Clinician Reported Outcome Spine Trauma (AO Spine CROST) [6].

An initial reliability study, using anonymized clinical cases from daily clinical practice through an online system,

Extended author information available on the last page of the article

showed moderate results [6]. It was hypothesized that a more adequate evaluation of the CROST would be possible when patients were seen and assessed by the clinician in a true clinical setting. Therefore, the aim of the current study was to evaluate the feasibility, internal consistency, interrater reliability, and prospective validity of the CROST in the clinical setting. Also, the correlation between the clinician reported CROST and patient reported PROST was investigated.

Materials and methods

Study design

An international multicenter cross sectional study with prospective follow-up until 1-year post-trauma was performed in four centers, recruited through the AO Spine Knowledge Forum (KF) Trauma. The participating centers included trauma hospitals from Australia (The Alfred Hospital, National Trauma Research Institute, Monash University, Clayton), the Netherlands (University Medical Center, Utrecht), Slovakia (Slovak Medical University, F. D. Roosevelt University General Hospital, Banska Bystrica), and Switzerland (Inselspital, University of Bern). Data were gathered through the online system REDCap, using study identification codes. According to the Medical Ethics Committee of the participating centers, this protocol did not need ethical approval under the scope of the Medical Research Involving Human Subjects Act because participants were not subjected to procedures, nor were they required to follow any specific protocol.

Surgeons

Two spine surgeons with at least 3 years of experience in spine trauma care participated from each center. Surgeon 1 was a member of the AO Spine KF Trauma, and was considered as the most experienced among these two surgeons. Surgeon 2 was recruited by Surgeon 1 at each center.

Patients

Adult patients (≥ 18 years) sustaining traumatic spine fractures and within 3 months post-trauma were included. They had to have mild or no neurological deficit (American Spinal Injury Association (ASIA) Severity score (AIS) C, D or E) at the time of discharge from hospital. In line with the target patient population in previous validation studies of PROST, patients with motor complete paralysis (AIS A or B) and

hospitalized patients were excluded [3]. The desired sample size was 100 patients (25 per center), based on recommendations for this type of study [7].

Instruments

Two separate questionnaires were administered: one to the surgeons and another to the patients.

Surgeons completed CROST for each patient at their center. As shown in Appendix 1, this tool consists of 10 parameters. Eight parameters are rated for both surgically and nonsurgically treated patients, while 2 parameters are only applicable to surgically treated patients ('Wound healing' and 'Implants'). Each parameter is rated both for the short-term (<12 months) and long-term (≥ 12 months). A 'yes'-answer provides 1 point, and expresses any expected problems or adverse events for the parameters. The total recorded score is the sum of the 'yes'-answers with a maximum achievable score being 8 points for non-surgically and 10 points for surgically treated patients. A higher score indicates worse expected outcome.

Additionally, surgeons were also asked to complete patients' background data, as well as evaluation questions in order to assess the feasibility: time to complete CROST, if it was considered as an easy and useful tool, if any difficulties were encountered when filling out, and if there were any redundant or missing parameters. Finally, the AO Spine KF Trauma surgeon was asked to assess the overall patient outcome in various prospective time points.

The patient part of the questionnaire consisted of PROST, which includes 19 questions on a broad range of aspects of functioning [3, 8–12]. Each item has a 0–100 Numeric Rating Scale, with 0 indicating no function at all and 100 the pre-injury level of function. The item "Work/Study" is optional. The total score is calculated by the mean of the answered questions. A higher score indicates improved outcome.

Study procedures

Eligible patients were identified and screened either just before discharge from hospital or at their first outpatient clinic appointment. Patients were enrolled in the study after informed consent was given. They were seen at three time points: baseline (i.e., the first outpatient clinic visit), 6-months, and 1-year after the trauma that caused their spine injury. At all these time points, patients were asked to complete PROST.

In order to assess the reliability of CROST, the two surgeons located at the same center independently made clinical assessments, and completed the tool for the same patient at the baseline visit.

Concerning the prospective evaluation, CROST was also scored at 6-months and 1-year visits. At these time points, the questionnaire was only completed by Surgeon 1 (i.e., the AO Spine KF Trauma member). This surgeon was also asked to judge the overall outcome of the patient at 6-months and 1-year with a binary definition: ‘same or better outcome than expected’ or ‘worse outcome than expected’. A ‘same or better outcome than expected’ was scored if the treatment goals were achieved, and ‘worse outcome than expected’ if they were not. For example, conversion of a conservatively treated patient to a surgical case, a surgically treated patient that undergoes a re-operation, or a patient highly dysfunctional in daily activities could be considered as ‘worse outcome than expected’.

Statistical analysis

Descriptive statistics were used to analyze patient characteristics and the feasibility of CROST. The internal consistency of the tool was analyzed by calculating Cronbach’s α . An $\alpha > 0.70$ is accepted as satisfactory result [7].

Inter-rater reliability analysis was performed both for individual CROST items as well as for the total score. Kappa statistics was used for the individual CROST items, with < 0 values indicating poor agreement, 0.00–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement [13]. The Intraclass Correlation Coefficient (ICC) was used for total CROST score, with an ICC of 0.70–0.85 and > 0.85 indicating good and excellent reliability, respectively [7].

The prospective analysis was performed by comparing outcomes as assessed at the baseline to the outcomes at 6-months and 1-year follow-up. The CROST scores at baseline were compared to the actual outcomes (same/better *versus* worse outcome) at 6-months and 1-year follow-up. Also, Spearman correlation coefficients (r_s) between CROST scores at baseline and the scores at 6-months and 1-year follow-up were analyzed. The r_s ranges from +1 to –1, with +1 indicating a perfect association, 0 no association, and –1 perfect negative association [7].

Finally, correlations between the clinician-reported CROST scores and patient-reported PROST scores were explored. Descriptive statistics were used to correlate CROST scores at baseline to PROST scores at different prospective time points. The change in CROST and PROST scores over time was analyzed using Spearman correlations. Also, the association between the ‘actual’ binary outcome (same/better *versus* worse outcome) was compared to PROST scores at 6-months and 1-year follow-up.

Results

Patient characteristics

A total of 92 patients were included in the study: 24 (26.1%) from Australia, 27 (29.3%) Dutch patients, 15 (16.3%) from Slovakia, and 26 (28.3%) Swiss patients. Table 1 shows the overall patient characteristics, as well as stratified for the provided treatment and per participating center.

Feasibility

The questions concerning the feasibility of the CROST were completed by 7 surgeons. Five surgeons stated that it took less than 5 min to complete the tool; while, two surgeons mentioned 5–10 min. All agreed the tool was easy to use and no difficulties were experienced in completing. No parameter was deemed difficult, redundant or missing. All surgeons expected that the CROST would be a useful tool in the clinical setting.

Internal consistency

As shown in Table 2, the internal consistency of CROST total score was moderate with Cronbach’s α ranging from 0.58 and 0.70.

Inter-rater reliability

The inter-rater reliability results for the total CROST scores as well as for each item are shown in Tables 3 and 4, respectively.

Moderate reliability results were found for the total scores, both for the short-term anticipated scores (ICC = 0.55) and long-term anticipated scores (ICC = 0.52). Subanalysis showed better reliability results for conservatively treated patients (ICC = 0.59–0.81) compared with surgically treated patients (ICC = 0.34–0.39).

As shown in Table 4, analyses of the mean scores per CROST item showed very good exact agreement results ranging from 73.9% (‘Range of motion impairment’) to 98.9% (‘Sagittal alignment problems’) for the short-term anticipated scores. Comparable results were seen for the long-term anticipated scores: 81.5% (range of motion impairment) to 100.0% (wound healing problems). Additional analysis including Kappa values showed somewhat varying results. Except poor agreement for ‘Implants adverse events’ ($\kappa = -0.4$ both for the short-term and long-term anticipated scores), most other CROST items showed moderate

Table 1 Patient characteristics. Also stratified per type of treatment as well as per center

	Total (n=92)	Conservative (n=48)	Surgical (n=44)	Australia (n=24)	Netherlands (n=27)	Slovakia (n=15)	Switzerland (n=26)
Male (%)	64.1	62.5	65.9	62.5	66.7	66.7	61.5
Age, mean (SD)	52.0 (19.7)	54.1 (20.1)	49.7 (19.2)	60.1 (16.7)	44 (20.5)	51.5 (16)	53.1 (20.9)
Trauma cause (%)							
Traffic	19.6	22.9	15.9	33.3	18.5	6.7	15.4
Fall	56.5	62.5	50.0	41.7	66.7	73.3	50.0
Sports	22.8	14.6	31.8	25	14.6	13.3	34.6
Other	1.1	0	2.3	0	0	6.7	0
Surgical treatment (%)	47.8	NA	NA	8.3	55.6	13.3	96.2
Fracture levels							
C spine	22.8	20.8	25.0	37.5	18.5	0	26.9
T spine (T1-T10)	19.6	10.4	29.5	12.5	33.3	6.7	19.2
TL spine (T11-L2)	51.1	60.4	40.9	41.7	44.4	86.7	46.2
L spine (L3-L5)	6.5	8.3	4.5	8.3	3.7	6.7	7.7
Fracture type ^a (%)							
A	80.2	94.9	66.7	77.8	88.0	100	62.5
B	33.3	2.6	61.9	16.7	44.0	0	54.2
C	2.5	0	4.8	0	0	0	8.3
F	4.9	2.6	7.1	5.6	0	0	12.5
Related injuries ^b (%)	50.0	58.3	41.7	45.8	74.1	6.7	53.8
≥ 1 comorbidities ^c (%)	47.8	52.1	43.2	62.5	55.6	20.0	42.3
ASIA at discharge (%)							
A-C	0	0	0	0	0	0	0
D	9.8	4.2	15.9	12.5	3.7	0	19.2
E	90.2	95.8	84.1	87.5	96.3	100	80.8

^afracture type according to the AO Spine Injury Classification system (F=facet fracture in cervical spine). More than one option possible, displayed are percentages of cases

^be.g., costal fractures, soft tissue lesions to extremities, abdominal or thorax trauma, trauma to the skull/cerebrum

^ce.g., pulmonary, cardiovascular, neurological diseases, diabetes, alcohol abuse, osteoporosis

NA not applicable

agreement; while, ‘Sagittal alignment problems’ showed an almost perfect agreement ($\kappa=0.85$).

Prospective analysis

The CROST scores at baseline were divided into 3 scoring subcategories: 0, 1, and ≥ 2 . As shown in Table 5, none of those subcategories showed a specific correlation to the

actual assessed outcomes at the follow-up. Nevertheless, a trend was seen when CROST was scored 0 (indicating no concerns at all), in which the vast majority of patient outcomes (87.0–93.8%) were classified as ‘same or better than expected’. Moderate to strong positive Spearman correlations were found between CROST scores at baseline and the scores at 6-months and 1-year follow-up, with significant r_s values ranging from 0.41 to 0.64 (Table 6).

Table 2 Internal consistency results (Cronbach’s α), shown for the AO Spine CROST scores at different study time points (baseline, 6-months, and 1-year follow-up), stratified for the short-term (<12 months) and long-term (≥ 12 months) anticipated CROST scores, as well as stratified for surgeons (Surgeon 1 and Surgeon 2). Results are shown for all patients (conservative and surgically treated patients, i.e., 8 CROST items scored) and only surgically treated patients (i.e., 10 CROST items scored)

Study time point	Surgeon 1		Surgeon 2	
	All patients (8 items) (n=92)	Surgical only (10 items) (n=44)	All patients (8 items) (n=92)	Surgical only (10 items) (n=44)
<i>Baseline</i>				
Short-term (<12 m) CROST scores	0.63	0.60	0.64	0.44
Long-term (≥ 12 m) CROST scores	0.61	0.57	0.69	0.49
<i>6-months follow-up</i>				
Short-term (<12 m) CROST scores	0.67	0.71		
Long-term (≥ 12 m) CROST scores	0.70	0.75		
<i>1-year follow-up</i>				
Short-term (<12 m) CROST scores	0.59	0.68		
Long-term (≥ 12 m) CROST scores	0.58	0.56		

Table 3 Descriptive and agreement statistics for AO Spine CROST total score, between Surgeon 1 and Surgeon 2 at baseline study time point

	Surgeon 1—Mean (SD)	Surgeon 2—Mean (SD)	% Exact agreement	ICC (95% CI)	
<i>Short-term (<12 m) CROST scores</i>	1.0 (1.4)	0.7 (1.2)	48.9	0.55 (0.38–0.68)	
	0.7 (1.2)	0.9 (1.4)	64.6	0.81 (0.68–0.89)	
	Total sample (n=92)	1.4 (1.5)	0.5 (0.9)	31.8	0.34 (0.02–0.58)
	Conservative (n=48)				
Surgical (n=44)					
<i>Long-term (≥ 12 m) CROST scores</i>	0.9 (1.3)	0.4 (1.0)	57.6	0.52 (0.31–0.67)	
	1.0 (1.4)	0.5 (1.2)	64.6	0.59 (0.34–0.76)	
	Total sample (n=92)	0.8 (1.2)	0.4 (0.8)	50.0	0.39 (0.11–0.61)
	Conservative (n=48)				
Surgical (n=44)					

Table 4 Descriptive and agreement statistics for each AO Spine CROST item, between Surgeon 1 and Surgeon 2 at baseline study time point (n=92)

CROST item	Short-term (<12m) CROST scores				Long-term (≥ 12 m) CROST scores			
	Surgeon 1, yes-answer (%)	Surgeon 2, yes-answer (%)	% Exact agreement	Kappa	Surgeon 1, yes-answer (%)	Surgeon 2, yes-answer (%)	% Exact agreement	Kappa
1 Neurological deterioration	2	0	97.8	X**	2	0	97.8	X
2 Sagittal alignment problems	4	3	98.9	.85	14	5	89.1	.40
3 Bone quality adverse events	16	10	86.9	.43	21	10	84.8	.42
4 Mechanical stability adverse events	10	4	92.4	.43	8	4	92.4	.33
5 Range of motion impairment	15	13	73.9	.23	18	4	81.5	.13
6 General physical condition	10	13	88.0	.41	9	10	92.4	.55
7 General psychological condition	2	7	93.5	.23	2	2	95.7	-.02
8 Functional recovery problems	22	15	82.6	.43	18	4	83.7	.23
9* Wound healing problems	5	2	97.8	.66	0	0	100	X
10* Implants adverse events	14	2	84.1	-.04	2	7	90.1	-.04

*Only surgical patients (n=44)

**Kappa cannot be computed because one or both variables are a constant

Correlation AO Spine CROST and PROST

No specific correlation was observed between the clinician-reported CROST scores at baseline as compared to the patient-reported PROST scores at different time points (baseline, 6-months, and 1-year follow-up). Higher CROST scores (i.e., more concerned from clinical perspective) did not result in worse PROST scores nor were the differences statistically significant (Table 7). As shown in Table 8, no Spearman correlations were found between the change in CROST scores and change in PROST scores when compared at the baseline relatively to the scores at 6-months and 1-year follow-up ($r_s = -0.33 - 0.07$). Finally, there seemed to be a statistically significant correlation between the PROST score and the assessed outcome by the surgeon (same/better versus worse outcome than expected). Table 9 reflects this with worse patient-reported PROST scores when the overall outcome is assessed as worse than expected.

Discussion

This study investigated the validation of the AO Spine CROST (Clinician Reported Outcome Spine Trauma) in the clinical setting. In contrast to a previous validation study that included online cases [6], the current study was performed in an actual clinical setting including patients from daily clinical practice. Excellent feasibility and acceptable internal consistency results were found. This indicates that the tool is deemed useful in the clinical setting and that its content measures the intended concept of assessing clinical outcomes from the perspective of the clinicians.

The inter-rater reliability analysis showed moderate results. Although only minor differences were found for the total CROST scores between Surgeon 1 and Surgeon 2 (0.2–0.9 difference), the agreement percentages were relatively low (48.9–57.6%). This may be explained by the high amount of variations in scoring the same exact score with a total ranging from 0 to 10. Additional subanalysis per CROST item showed very good exact agreement results (73.9–100.0%). On the other hand, varying Kappa values were found with the most agreements being moderate. These Kappa results may be skewed, and not fully representative, due to the very high number of CROST items that were responded with a ‘no’-answer (i.e., no concerns were expected with those items).

Prospective evaluation analysis of the CROST scores did not show a specific correlation to the overall outcomes as assessed by the surgeon at follow-up time points (same/better versus worse than expected). It is interesting to explore the clinicians’ perspective relative to the patients’ perspective on health and functioning. In the case of the treatment of spinal

Table 5 AO Spine CROST scores for the short-term (<12-months) as scored at baseline study time point compared to assessed outcomes (‘same/better’ versus ‘worse’ outcome than expected) at the follow-up: 6-months ($n = 89$) and at 1-year ($n = 75$) study time points. Results are shown for the total patient sample as well as for conservatively and surgically treated patients

CROST score	Total sample				Conservative				Surgical			
	6M follow-up, n (%)		1Y follow-up, n (%)		6M follow-up, n (%)		1Y follow-up, n (%)		6M follow-up, n (%)		1Y follow-up, n (%)	
	Same/better ($n = 76$)	Worse ($n = 13$)	Same/better ($n = 63$)	Worse ($n = 12$)	Same/better ($n = 38$)	Worse ($n = 8$)	Same/better ($n = 31$)	Worse ($n = 4$)	Same/better ($n = 38$)	Worse ($n = 5$)	Same/better ($n = 32$)	Worse ($n = 8$)
0	40 (87.0)	6 (13)	34 (87.2)	5 (12.8)	25 (83.3)	5 (16.7)	20 (87.0)	3 (13)	15 (93.8)	1 (6.3)	14 (87.5)	2 (12.5)
1	17 (89.5)	2 (10.5)	15 (78.9)	4 (21.1)	6 (75.0)	2 (25.0)	6 (85.7)	1 (14.3)	11 (100)	0 (0)	9 (75.0)	3 (25.0)
≥2	19 (79.2)	5 (21.8)	14 (82.4)	3 (17.6)	7 (87.5)	1 (12.5)	5 (100)	0 (0)	12 (75)	4 (25)	9 (75.0)	3 (25.0)

Table 6 Spearman correlations (r_s) between AO Spine CROST scores at baseline and 6-months and 1-year follow-up study time points. Results are shown for the total patient sample as well as for conservatively and surgically treated patients

	Total sample		Conservative		Surgical	
	6M follow-up (n = 89)	1Y follow-up (n = 75)	6M follow-up (n = 46)	1Y follow-up (n = 35)	6M follow-up (n = 43)	1Y follow-up (n = 40)
Short-term (<12 m) CROST scores	.61***	.41***	.58***	.28	.63***	.46***
Long-term (≥12 m) CROST scores	.58***	.41***	.51***	.20	.64***	.62***

*** $p < .001$

Table 7 Relationships between short-term (<12 months) AO Spine CROST as scored at baseline study time point, in comparison to AO Spine PROST scores at baseline and follow-up study time points (6-months and 1-year). Results are shown for the total patient sample as well as for conservatively and surgically treated patients

Short-term (<12 m) CROST scores	Total sample			Conservative			Surgical		
	PROST baseline (n = 84)	PROST at 6M (n = 74)	PROST at 1Y (n = 67)	PROST baseline (n = 48)	PROST at 6M (n = 42)	PROST at 1Y (n = 32)	PROST baseline (n = 36)	PROST at 6M (n = 32)	PROST at 1Y (n = 35)
0	59.4 (19.3)	79.5 (17.6)	84.5 (15.4)	54.9 (19.6)	76.9 (19.3)	82.1 (17.4)	69.1 (14.8)	85.2 (12.0)	88.0 (11.6)
1	63.7 (19.4)	83.0 (10.1)	87.1 (12.1)	62.5 (19.5)	81.0 (10.6)	82.7 (15.2)	64.7 (20.2)	84.6 (10.1)	90.4 (8.8)
≥2	54.1 (20.4)	72.3 (21.6)	78.7 (17.9)	51.0 (20.2)	76.0 (17.6)	78.4 (19.8)	56.6 (21.2)	69.7 (24.5)	78.8 (17.9)
All	59.1 (19.6)	78.6 (17.5)	83.6 (15.5)	55.6 (19.6)	77.4 (17.6)	81.6 (16.9)	63.9 (18.8)	63.9 (18.8)	85.4 (14.1)

NB: none of the differences in PROST scores are statistically significant

Table 8 Spearman correlations between change in AO Spine CROST scores and change in AO Spine PROST scores as compared between baseline to 6-months and 1-year follow-up study time points. Results are shown for the total patient sample as well as for conservatively and surgically treated patients

	Total sample		Conservative		Surgical	
	Baseline to 6M	Baseline to 1Y	Baseline to 6M	Baseline to 1Y	Baseline to 6M	Baseline to 1Y
Short-term (<12 m) CROST scores	0.07	-0.30	0.02	-0.33	0.06	-0.30

Table 9 Relationships between assessed outcomes ('same/better' versus 'worse' outcome than expected) in comparison to mean AO Spine PROST scores (SD) at 6-months and at 1-year follow-up study time points. Results are shown for the total patient sample as well as for conservatively and surgically treated patients

Assessed outcome	Total sample		Conservative		Surgical	
	PROST 6M (n = 73)	PROST 1Y (n = 65)	PROST 6M (n = 42)	PROST 1Y (n = 32)	PROST 6M (n = 31)	PROST 1Y (n = 33)
Same/better	82.6 (13.3)	87.5 (11.9)	81.3 (14.4)	84.3 (14.4)	84.3 (12.2)	90.1 (7.3)
Worse	51.0 (20.2)*	60.0 (15.8)*	53.8 (19.5)*	55.4 (19.1)*	45.5 (24.8)*	62.8 (15.1)*

* $p < .001$

trauma patients, several clinical and radiological parameters are generally used by treating surgeons to evaluate treatment results. The most relevant parameters among spine trauma patients were identified in two preparatory studies in the developmental process of CROST [14, 15]. An estimation of any expected problems with respect to those parameters are made by the treating surgeons in order to determine the further course of treatment. The surgeon's assessment may differ substantially from the patient's perception [16, 17]. These discrepant views have also been addressed for a variety of other diseases, including metastatic diseases [18], multiple sclerosis [19], rheumatoid arthritis [20], and peripheral artery diseases [21]. The current study substantiates the discrepant views, and therefore the need for the clinician-reported CROST.

The patient-reported PROST analysis was not the main focus of the current study and, therefore, not further detailed in the Results section. Nevertheless, it is worth to mention that during the follow-up a gradual increase is seen in the mean PROST scores, indicating gradual recovery of the patients over time. This is in line with previous validation studies in which the PROST was cross-culturally translated and validated in the Dutch, English, German, Nepali and Slovak versions [8–11, 12]. A very recent publication states that translations have been, or are being, performed in a total of 17 languages [22]. This facilitates a worldwide use of the patient-reported outcome measure. As the clinician-reported CROST is assessed by the treating surgeons or clinicians, the authors recommend no additional translations besides the original English version.

This study has several limitations. The intra-rater reliability was not assessed due to the study procedures, as it was considered very challenging to see patients back at multiple additional time points across 4 different centers. Secondly, the number of included patients was lesser than initially anticipated, and the contribution of included patients from the 4 centers was not equal. The different amount of spine trauma exposure and local practical difficulties at the centers contributed to this limitation. Also, the patient population was somewhat heterogeneous. Finally, the binary outcome as assessed by the treating surgeon may be somewhat arbitrary. However, we believe this is a valid strategy to assess clinical outcomes, as judged by a highly experienced spine trauma surgeon.

In conclusion, the AO Spine CROST showed moderate results in the current validation study in a true clinical setting including patients from the daily clinical practice. In future studies, the validation will be further investigated among larger patient and clinician samples. With its unique approach as a clinician-rated outcome measure, this tool has the potential to be valuable for use in clinics and research.

Appendix 1: AO Spine CROST

AO Spine CROST (Clinician Reported Outcome Spine Trauma)

The AO Spine CROST is applied after the initial treatment, and allows you as the treating surgeon to evaluate and predict clinical outcomes of spine trauma patients

Surgeon's Name:
Center:
Date (MM/DD/YY): ____ / ____ / ____

Please rate the following parameters:

		In the next 12 months	From 12 months onwards
1. Neurological status	Do you expect a neurological deterioration?	<input type="checkbox"/> No <input type="checkbox"/> Yes	<input type="checkbox"/> No <input type="checkbox"/> Yes
2. Radiographic sagittal alignment	Do you expect clinically relevant problems from sagittal alignment?	<input type="checkbox"/> No <input type="checkbox"/> Yes	<input type="checkbox"/> No <input type="checkbox"/> Yes
3. General bone quality	Do you expect adverse events related to the general bone quality?	<input type="checkbox"/> No <input type="checkbox"/> Yes	<input type="checkbox"/> No <input type="checkbox"/> Yes
4. Stability of the injured spine level	Do you expect adverse events related to mechanical instability of the injured spinal level(s)?	<input type="checkbox"/> No <input type="checkbox"/> Yes	<input type="checkbox"/> No <input type="checkbox"/> Yes
5. Spinal column mobility	Do you expect a functionally relevant impairment related to spinal column range of motion?	<input type="checkbox"/> No <input type="checkbox"/> Yes	<input type="checkbox"/> No <input type="checkbox"/> Yes
6. General physical condition	Do you expect the clinical outcome to be negatively affected by the general physical condition?	<input type="checkbox"/> No <input type="checkbox"/> Yes	<input type="checkbox"/> No <input type="checkbox"/> Yes
7. General psychological condition	Do you expect the clinical outcome to be negatively affected by the general psychological condition?	<input type="checkbox"/> No <input type="checkbox"/> Yes	<input type="checkbox"/> No <input type="checkbox"/> Yes

8. Functional recovery Do you expect problems in functional recovery? No Yes No Yes

Please rate parameters 9 and 10 *only if the patient is treated surgically*:

9. Wound healing Do you expect problems with wound healing or persistent infection? No Yes No Yes

10. Implants Do you expect any implant related adverse events? No Yes No Yes

Each 'yes'-answer provides 1 point. The total score is the sum of each 'yes'-answer with a maximum of 8 points for nonsurgically and 10 points for surgically treated patients. A higher score indicates worse expected outcome. The score guides the treating surgeon in anticipating on a change in the current treatment plan. Currently, studies are being prepared or performed to define cutoff points for the scoring algorithm

Declarations

Conflict of interest This study was organized and funded by AO Spine through the AO Spine Knowledge Forum Trauma, a focused group of international Trauma experts. AO Spine is a clinical division of the AO Foundation, which is an independent medically guided not-for-profit organization. Study support was provided directly through AO Network Clinical Research.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.


References

- Oner C, Rajasekaran S, Chapman JR, Fehlings MG, Vaccaro AR, Schroeder GD et al (2017) Spine trauma-what are the current controversies? J Orthop Trauma 31(Suppl 4):S1–S6
- Oner C, Sadiqi S, Lehr AM, Schroeder GD, Vaccaro AR (2017) The need of validated disease-specific outcome instruments for spine trauma. J Orthop Trauma 31(Suppl 4):S33–S37
- Sadiqi S, Lehr AM, Post MW, Dvorak MF, Kandziora F, Rajasekaran S et al (2017) Development of the AOSpine Patient Reported Outcome Spine Trauma (AOSpine PROST): a universal disease-specific outcome instrument for individuals with traumatic spinal column injury. Eur Spine J 26(5):1550–1557
- Nygaard OP, Kloster R, Dullerud R, Jacobsen EA, Mellgren SI (1997) No association between peridural scar and outcome after lumbar microdiscectomy. Acta Neurochir (Wien) 139(12):1095–1100
- Witt I, Vestergaard A, Rosenklint A (1984) A comparative analysis of x-ray findings of the lumbar spine in patients with and without lumbar pain. Spine (Phila Pa 1976) 9(3):298–300
- Sadiqi S, Muijs SPJ, Renkens JJM, Post MW, Benneker LM, Chapman JR et al (2020) Development and reliability of the AOSpine CROST (Clinician Reported Outcome Spine Trauma): a tool to evaluate and predict outcomes from clinician's perspective. Eur Spine J 29(10):2550–2559
- Terwee CB, Bot SDM, de Boer MR, van der Windt DIAWM, Knol DL, Dekker J, et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60(1):34–42
- Dhakal GR, Sadiqi S, Dhakal R, Dhungana S, Yadav PK, Shah G et al (2022) Reliability and validity of the adapted Nepali version of the AO spine patient reported outcome spine trauma. J Nepal Health Res Counc 19(4):730–739
- Sadiqi S, Dvorak MF, Vaccaro AR, Schroeder GD, Post MW, Benneker LM, et al (2020) Reliability and validity of the English version of the AOSpine PROST (Patient Reported Outcome Spine Trauma). Spine (Phila Pa 1976) 45(17):E1111–E1118
- Hackel S, Oswald KAC, Koller L, Benneker LM, Benneker LA, Sadiqi S, et al (2023) Reliability and Validity of the German Version of the AO Spine Patient Reported Outcome Spine Trauma Questionnaire. Global Spine J 21925682231156124
- Sadiqi S, Post MW, Hosman AJ, Dvorak MF, Chapman JR, Benneker LM et al (2021) Reliability, validity and responsiveness of the Dutch version of the AOSpine PROST (Patient Reported Outcome Spine Trauma). Eur Spine J 30(9):2631–2644
- Holas M, Gajdos R, Svac J, Holasova J, Valihorova M, Alberty R (2023) Translation, intercultural adaptation, and validation of the Slovak version of AOSpine patient reported outcome for spinal trauma tool. Bratisl Lek Listy 124(4):273–276
- McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22(3):276–282
- Sadiqi S, Verlaan JJ, Lehr AM, Dvorak MF, Kandziora F, Rajasekaran S, et al (2016) Surgeon reported outcome measure for spine trauma: an international expert survey identifying parameters relevant for the outcome of subaxial cervical spine injuries. Spine (Phila Pa 1976) 41(24):E1453–E1459
- Sadiqi S, Verlaan JJ, Mechteld Lehr A, Dvorak MF, Kandziora F, Rajasekaran S et al (2017) Universal disease-specific outcome instruments for spine trauma: a global perspective on relevant parameters to evaluate clinical and functional outcomes of thoracic and lumbar spine trauma patients. Eur Spine J 26(5):1541–1549
- Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS (1994) Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 331(2):69–73
- Remes VM, Lamberg TS, Tervahartiala PO, Helenius IJ, Osterman K, Schlenzka D et al (2005) No correlation between patient outcome and abnormal lumbar MRI findings 21 years after posterior or posterolateral fusion for isthmic spondylolisthesis in children and adolescents. Eur Spine J 14(9):833–842
- Wilson KA, Dowling AJ, Abdolell M, Tannock IF (2000) Perception of quality of life by patients, partners and treating physicians. Qual Life Res 9(9):1041–1052
- Rothwell PM, McDowell Z, Wong CK, Dorman PJ (1997) Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. BMJ 314(7094):1580–1583

20. Kwok CK, O'Connor GT, Regan-Smith MG, Olmstead EM, Brown LA, Burnett JB et al (1992) Concordance between clinician and patient assessment of physical and mental health status. *J Rheumatol* 19(7):1031–1037
21. Vossen RJ, Ras D, Vahl AC, Leijdekkers VJ, Montauban van Swijndregt AD, Wisselink W, et al (2022) Correlation of patient-reported outcome measures and the ankle-brachial index in patients who underwent revascularization for peripheral artery disease. *Vasc Med* 1358863X221138879
22. Sadiqi S, Oner FC (2022) A disease-specific patient reported outcome instrument for spine trauma is developed, validated and available! Re: Andrzejowski et al. Measuring functional outcomes in major trauma: can we do better? *Eur J Trauma Emerg Surg* 49(3):1607

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Said Sadiqi¹  · Erin E. A. de Gendt¹ · Sander P. J. Muijs¹ · Marcel W. M. Post^{2,3} · Lorin M. Benneker⁴ · Martin Holas⁵ · Jin W. Tee^{6,7} · Christoph E. Albers⁴ · Sonja Häckel⁴ · Juraj Svac⁵ · Richard J. Bransford⁸ · Mohammad M. El-Sharkawi⁹ · Frank Kandziora¹⁰ · Shanmuganathan Rajasekaran¹¹ · Klaus J. Schnake¹² · Alexander R. Vaccaro¹³ · F. Cumhur Oner¹

✉ Said Sadiqi
s.sadiqi-3@umcutrecht.nl

¹ Department of Orthopaedics, University Medical Center Utrecht, Utrecht, The Netherlands

² Rehabilitation Center 'De Hoogstraat', Utrecht, The Netherlands

³ Department of Rehabilitation Medicine, Center for Rehabilitation, University Medical Center Groningen, Groningen, The Netherlands

⁴ Department of Orthopaedic Surgery and Traumatology, Inselspital, University Hospital Bern, University Bern, Bern, Switzerland

⁵ Department of Trauma Surgery, Slovak Medical University, F. D. Roosevelt University General Hospital, Banska Bystrica, Slovakia

⁶ Department of Neurosurgery, Alfred Hospital, Melbourne, VIC, Australia

⁷ Department of Surgery, Central Clinical School, Monash University, Melbourne, VIC, Australia

⁸ Department of Orthopaedics and Sports Medicine, University of Washington/Harborview Medical Center, Seattle, WA, USA

⁹ Department of Orthopaedic and Trauma Surgery, Faculty of Medicine, Assiut University, Assiut, Egypt

¹⁰ Unfallklinik Frankfurt Am Main, Frankfurt, Germany

¹¹ Department of Orthopaedics and Spine Surgery, Ganga Hospital, Coimbatore, Tamil Nadu, India

¹² Center for Spinal and Scoliosis Surgery, Malteser Waldkrankenhaus St. Marien, Erlangen, Germany

¹³ Department of Orthopaedic Surgery, Rothman Institute, Thomas Jefferson University, Philadelphia, PA, USA