REVIEW ARTICLE

A systematic review identifying outcome measures used in evaluating adults sustaining cervical spine fractures

Phillip C. Copley[1,2,3] · Daniel Tadross[4] · Nadia Salloum[1] · Julie Woodfield[1,2] · Ellie Edlmann[5,6] · Michael Poon[1,2,3,7] · Sadaquate Khan[1,2,3] · Paul M. Brennan[1,2,3]

Abstract

Objective To assess the outcome measures used in studies investigating cervical spine fractures in adults, with or without associated spinal cord injury, to inform development of a core outcome set.

Methods Medline, Embase and Scopus were searched for relevant studies until May 28, 2022, without a historic limit on study date. Study characteristics, population characteristics and outcomes reported were extracted and analyzed.

Results Our literature search identified 536 studies that met criteria for inclusion, involving 393,266 patients. Most studies were single center (87.3%), retrospective studies (88.9%) and involved a median of 40 patients (range 6–167,278). Treatments assessed included: surgery (55.2%), conservative (6.2%), halo immobilization (4.9%), or a mixture (33.2%). Median study duration was 84 months (range 3–564 months); the timing of clinical and/or radiological follow-up assessment after injury was reported in 56.7%. There was significant heterogeneity in outcomes used, with 79 different reported outcomes measures. Differences in use were identified between smaller/larger, retro-/prospective and single/multicenter cohorts. Over time, the use of radiological outcomes has declined with greater emphasis on patient-reported outcome measures (PROMs). Studies of conservative management were more likely to detail PROMs and mortality, whereas surgical studies reported Frankel/ASIA grade, radiological fusion, complication rates, duration of hospital stay and re-operation rates more frequently. In studies assessing the elderly population (> 65 years), use of PROMs, mortality, hospital stay and discharge destination were more common, whereas fusion was reported less often. Response rates for outcome assessments were lower in studies assessing elderly patients, and studies using PROMs.

Conclusions We have classified the various outcome measures used for patients with cervical spine fractures based on the COMET outcome taxonomy. We also described the contexts in which different outcomes are more commonly employed to help guide decision-making when designing future research endeavors.

Abbreviations

AISA	American Spinal Cord Injury Association
AS	Ankylosing spondylitis
COS	Core outcome set
COMET	Core outcome measure in effectiveness trials
CSOQ	Cervical Spine Outcome Questionnaire
DISH	Diffuse idiopathic skeletal hyperostosis
EQ-5D	European Quality of Life Five Dimension
FIM	Functional independence measure score
JOA	Japanese Orthopaedic Association Scoring System
LEDCs	Less economically developed countries

✉ Phillip C. Copley
Phillip.copley2@nhs.scot

1 Department of Clinical Neurosciences, Royal Infirmary of Edinburgh, Edinburgh, UK

2 Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

3 Department of Clinical Neurosciences, Edinburgh Spinal Surgery Outcome Studies Group, Edinburgh, UK

4 Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

5 Southwest Neurosurgical Centre, Derriford Hospital, Plymouth, UK

6 Faculty of Health: Medicine, Dentistry and Human Sciences, University of Plymouth, Plymouth, UK

7 Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, UK

MEDCs	More economically developed countries
mJOA	Modified Japanese Orthopaedic Association Scoring System
mRS	Modified Rankin score
NDI	Neck Disability Index
NPDI	Neck pain driving index
NRS	Numeric rating scale
ODI	Oswestry disability scale
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PROMs	Patient-reported outcome measures
PSI	Patient satisfaction index
PTNC	Post-traumatic neck score
SCI	Spinal cord injury
SCIM	Spinal cord independence measure
SF-36	36-Item short form survey
VAS	Visual analog scale
WISCI-2	Walking Index for Spinal Cord Injury 2 Score

Introduction

The incidence of traumatic cervical spine fractures is estimated at 15–65/100,000 hospital admissions annually [1]. Most cervical spine fractures are not associated with spinal cord injury (SCI). One prospective population-based study from Norway identified SCI in only 10% of cases of cervical fractures [2]. Symptoms resulting from cervical spine fractures vary in severity and in their impact on quality of life. A meaningful and relevant measure of outcomes from cervical fractures is important to understand the personal, population, healthcare and economic impact of these injuries [3–5].

Young patients tend to sustain cervical spine injuries from high-energy trauma. In less economically developed countries (LEDCs), this remains the most common mechanism of injury [6, 7]. However, in more economically developed countries (MEDCs), cervical spine fractures due to low-impact trauma are becoming more common in older people [8]. A nationwide database study 1 from 2005 to 2013 found an approximate 32% increase in the incidence of cervical fractures in American patients, along with an increase in the average age at which the fracture was sustained, from 51 to 59 years. The proportion of patients injured in falls rather than motor vehicle accidents also increased.

These heterogeneous patient demographics present a challenge to selecting outcome measures that are applicable and relevant to all groups, reliable, and can be compared between studies [9]. Consequently, outcome measures are often selected for specific contexts, such as younger or older patients, or those with or without an associated neurological deficit. There is therefore a need to identify a common core outcome set (COS) for cervical spine fractures to provide consistency in reporting of studies and facilitate comparisons across studies [10–13].

Outcome domains can be classified as per the COMET (Core Outcome Measure in Effectiveness Trials) outcome taxonomy of core areas [14]. COMET is an initiative aimed at guiding the development and application of COSs for use in specific conditions as a minimum set of outcomes that should be reported in clinical trials. Their taxonomy aims to classify the different types of outcomes used in trials, providing key 'areas' that aid development of COSs. We aimed to identify all outcome measures used in studies of cervical spine fractures in adults and classify them into these core areas in order to help inform future development of a COS.

Methods

The protocol for this systematic review was published before starting the review in PROSPERO (CRD42020172311).

Eligibility criteria

We included studies that reported original data from clinical research involving adult human subjects (> 18 years) with fractures of the cervical spine related to trauma. Included studies must have recorded at least one outcome measure. There were no restrictions on year of publication, location of the study or study design.

Exclusion criteria included: non-English language studies, case reports/series involving < 5 patients, studies with > 50% sustaining associated arterial injuries, studies where < 50% patients sustained a traumatic injury, studies where < 50% injuries were cervical spine fractures, studies with > 50% of cases comprising children (< 18 years).

Information sources/Search strategy

We searched Ovid Medline, Ovid Embase, and Scopus, on March 4, 2020, using the search strategies documented in the protocol (see supplementary material for full search strategies). Searches of the Cochrane Central Register of Controlled Trials, and Trials registries of ongoing trials (ClinicalTrials.gov, ISRCTN, EU clinical trials registry) revealed no further relevant studies. The search was updated on May 28, 2022.

Selection process

After removal of duplicates, two reviewers independently reviewed each record during abstract screening and full-text eligibility assessment. Disagreements were resolved through discussion with a third reviewer.

Data collection process

Data were extracted from each included paper. Data items included: study characteristics (year of publication, whether prospective/retrospective, whether single/multi-center, country/countries the study was based in, study recruitment period), patient characteristics (number of included patients, patient age, presence of SCI, whether the study assessed only patients > 65 years old, whether the study included > 50% of patients with AS/DISH), fracture characteristics (mechanism of injury, level of cervical fracture [upper = C1-2, subaxial = C3-7], proportion of patients with multiple cervical spine fractures, proportion of patients with concomitant thoracolumbar spine fractures), treatment modality, follow-up, timing of outcome assessment following injury and outcomes measured (including: a validated pain score, Frankel grade, American Spinal Cord Injury Association [ASIA] Impairment score, radiological evidence of fusion, radiological evidence of stability, any other radiological parameter recorded, mortality rates, complications, patient-reported outcome measures [PROMs], other assessments of functional outcomes [including any assessment of ambulatory status, bowel/bladder function, or employment status], length of hospital stay, discharge destination and re-operation rate).

Risk of bias

Risk of bias was not examined, as our aim was to assess the type of outcome measures used and not to perform summary effect measures from included studies.

Effect measures

For each outcome identified, the measure used to define the outcome was recorded.

Synthesis measures

Results were grouped according to study characteristics (decade of publication, whether prospective/retrospective, whether single/multi-center, continent the study was based in), patient characteristics (whether the study included greater/fewer patients than the median number of all studies, whether the study assessed only patients > 65 years old, whether the study included > 50% of patients with AS/DISH, presence/absence of SCI), level of cervical fracture and treatment modality. Heatmaps of the summary outcomes were generated using a graded color scale in order to highlight variance in reporting of outcomes across different studies. Each clinical outcome identified was classified into a core area, as per the COMET outcome taxonomy of core areas [14]. Statistical analysis was performed using SPSS version 24. Nonparametric data were assessed using the Mann–Whitney U test to assess for difference between two independent groups.

Results

Search findings and study characteristics

Our search identified 536 eligible studies involving 393,266 patients (Fig. 1). Articles included were published between 1979 and 2022. There was an increase over time in the number of publications identified per year. Table 1 outlines study characteristics. Most studies were single center ($n = 468$; 87.3%), and retrospective in nature ($n = 480$; 88.7%). The median number of patients involved in each study was 40, range 6–167,278 (interquartile range; Q1 = 20, Q3 = 82). Studies assessed surgical intervention ($n = 296$; 55.2%), a mixture of treatment modalities ($n = 178$; 33.2%), conservative treatments ($n = 33$; 6.2%), or management in a halo brace ($n = 26$; 4.9%). The upper cervical region alone (C1-2) was the focus of the publication in 310 studies (57.8%), the subaxial cervical spine (C3-7) in 88 (16.4%) and the whole cervical spine in 138 (25.7%). In 318 studies (68.4%), cohorts included patients admitted with clinical features attributable to SCI.

Median or mean age was reported in 461 studies (86.0%) with an age range reported in 353 (65.9%). In total, 80 (14.9%) studies included only patients over 65 years of age. Mechanism of injury was reported in 365 studies (59.3%), and of these, 99 (18.5%) reported > 50% of included cases being caused by low energy injuries or falls. In this group of studies mostly assessing patients with low energy injuries, 66 (66.7%) assessed only C1/2 fractures and 32 (32.3%) included only patients > 65 years of age. The number of patients with more than one cervical spine fracture was reported in 36.6% studies and whether there were other non-cervical vertebral fractures in 24.8%. Median study duration was 84 months (range 3–564 months [interquartile range; Q1 = 48, Q3 = 120]), with clear documentation of mean/median/minimum follow-up reported in 464 studies (86.6%).

Overall, the most frequently reported outcomes across all studies were: complications (both fracture-related and treatment-related) in 400 (74.6%), radiological assessment of fusion in 365 (68.1%), mortality rate in 325 (60.6%), radiological assessment of spinal stability in 209 (40%), at least one patient reported outcome in 157 (29.3%), other assessment of functional outcome (as defined in our methods) in 150 (28%), a validated pain score in 118 (22%), ASIA grade in 144 (26.9%), another radiological outcome in 104 (19.4%), Frankel grade in 68 (12.7%), length of stay in hospital in 93 (17.4%), re-operation rates in 69 (12.9%),

Fig. 1 Preferred reporting items for systematic reviews and meta-analyses (PRISMA) diagram
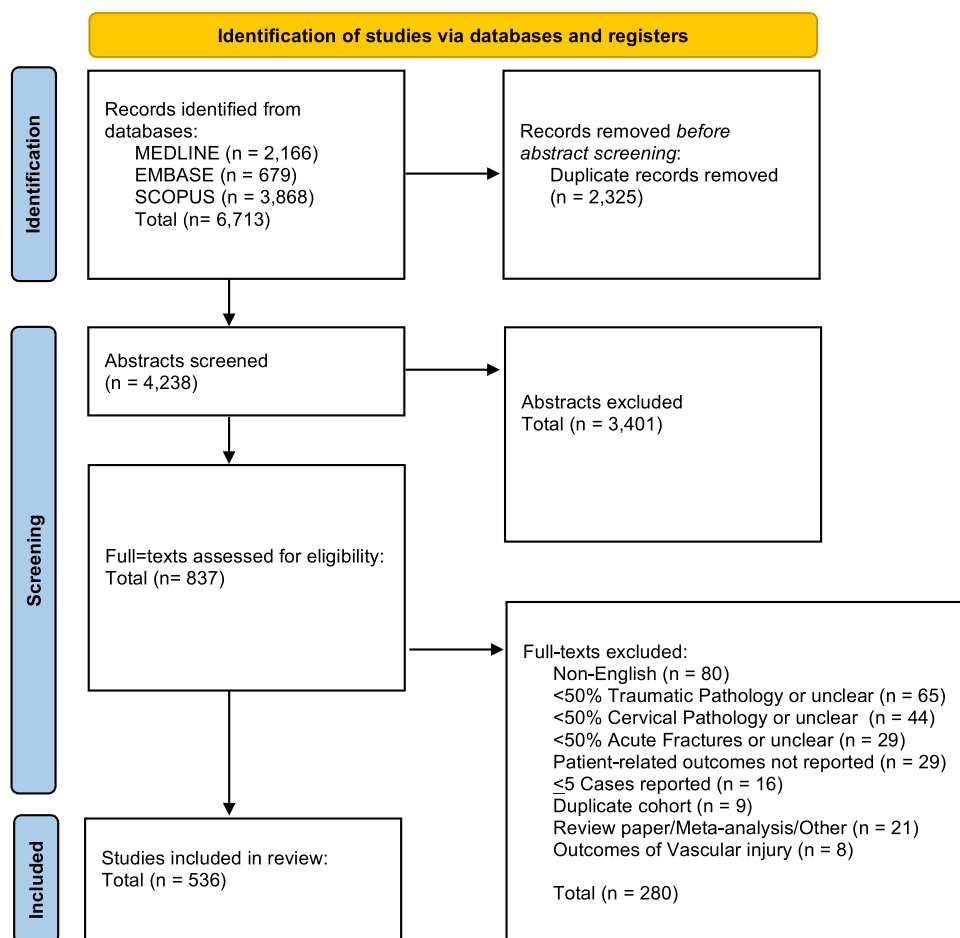
discharge destination in 26 (4.9%). Most studies did not define a single primary outcome.

The PROMs most commonly used were a validated pain score in 118 (22%), Neck Disability Index (NDI) in 67 (12.5%), 36-Item Short Form Survey (SF-36) in 18 (3.4%), European Quality of Life Five Dimension (ED-5Q) in 8 (1.5%), Cervical Spine Outcome Questionnaire (CSOQ) in 8 (1.5%), Patient Satisfaction Index (PSI) in 7 (1.3%), Neck Pain Driving Index (NPDI) in 3 (0.6%), and Post-Traumatic Neck Score (PTNC) in 2 (0.4%).

Definition of outcome types

Outcomes were measured using both validated and non-validated tools. Table 2 shows the manner in which outcomes assessed were recorded in the included studies. Some outcomes were uniformly measured across all studies, for example mortality as a percentage of participants. In contrast, recovery of neurological function was measured in some studies using a validated tool such as the ASIA or Frankel grade to depict improvement over follow-up, while other studies recorded a narrative depiction of how patients' neurological function evolved over the study. Some studies

defined their own criterion for outcome assessment. For example, 'treatment success/failure' was often defined by the authors. Table 2 shows the outcomes used in studies included in our review, as they relate to the COMET outcome taxonomy.

Outcomes measured in different contexts

Tables 3, 4 and 5 identify the outcomes used according to study population, fracture pattern and treatment modality. The use of PROMs was greater in multicenter studies compared to single center studies (34.4% vs. 28.6%), and in prospective studies (32.4% vs. 28.8%) but similar in studies assessing larger and small patient cohorts. Functional assessments were used more commonly in larger studies (29.5% vs. 26.5%) and in prospective studies (33.8% vs. 27.1%). Reporting of hospital stay and discharge destination was greater in larger studies (22.6% and 6.5%) greater compared to smaller studies (12.4% and 3.3%). This was true of multicenter studies (34.4% and 10.9%), compared to single-center studies (15% and 4%). Reporting of radiological parameters was more common in smaller, retrospective and single-center studies. While mortality rates were more

Table 1 Study characteristics

Characteristic						
Decade of publication	1970–79 (n=1, 0.2%)	1980–89 (n=15, 2.8%)	1990–99 (n=65, 12.1%)	2000–09 (n=124, 23.1%)	2010–19 (n=256, 47.8%)	2020–22 (n=76, 14.2%)
Single/multicenter study	Single (n=468, 87.3%)	Multicenter (n=68, 12.7%)				
Prospective/retrospective study	Prospective (n=56, 11.3%)	Retrospective (n=480, 88.7%)				
Study duration (months)	Median 84 months (range 3–564[interquartile range; Q1 48, Q3 120])					
Geographical location of study	Asia (n=179, 33.4%)	North America (n=178, 33.2%)	Europe (n=143, 26.7%)	Africa (n=15, 2.8%)	Oceania (n=11, 2.1%)	South America (n=7, 1.3%)	Multiple continents (n=3, 0.6%)
Number of patients	Range 6–167,278 patients (median 40 [interquartile range; Q1 20, Q3 82])					
Mean/median age recorded	Yes (n=461, 86%)	No (n=75, 14%)				
Age range recorded	Yes (n=353, 65.9%)	No (n=152, 34.1%)				
Level of cervical fracture	Upper/C1-2 (n=310, 57.8%)	Subaxial/C3–7 (n=88, 16.4%)	Any (n=138, 25.7%)			
Mechanism of injury recorded	Yes (n=365, 68.1%)	No (n=171, 31.9%)				
Patients with SCI included within study	Yes (n=318, 59.3%)	No (n=218, 40.7%)				

Table 1 (continued)

Studies with > 50% patients having AS/DISH	Yes (n = 29, 5.4%)			No (n = 507, 94.6%)
Patients with multiple cervical spine fractures recorded	Yes (n = 196, 36.6%)			No (n = 340, 63.4%)
Patients with concomitant thoracolumbar spine fractures recorded	Yes (n = 133, 24.8%)			No (n = 037, 75.2%)
Treatment modality	Conservative (n = 33, 6.2%)	Halo (n = 26, 4.9%)	Surgery (n = 296, 55.2%)	Mixture (n = 178, 33.2%)
Follow-up duration recorded	Yes (n = 464, 86.6%)			No (n = 72, 13.4%)
Timing of assessments post-injury recorded	Yes (n = 304, 56.7%)			No (n = 232, 43.3%)

Table 2 Outcomes measured according to the COMET outcome taxonomy of core areas, showing units of measure used in the studies in brackets

Core area	Outcomes used (unit of measurement)
Death	Mortality (%)
Physiological /clinical	Patient symptoms: Persistence of symptoms (no/%), Change in neurological function (non-validated score used), Frankel grade (A–E), ASIA grade (A–E), Zancolli's score (I–III) Change in pain (non-validated classification used), change in use of pain medications (no/%), VAS (1–10), NRS (1–10) Neck range of movement/neck stiffness on recovery (non-validated classification used) Dysphagia/dysphonia (non-validated classification used), Bazaz score for dysphagia (none/mild/moderate/severe) Radiological: Radiological fusion (%), radiological stability (%), radiological deformity (change in degree of kyphosis/Cobb angle/ degree of angulation; other non-validated classifications used), radiological reduction (%), loss of muscle on CT imaging (non-validated score), increase in cord diameter on imaging (%) Operative: Incision length (mm), Intra-operative bleeding (mL), requirement for transfusion (%), duration of operation (mins), changes in neuromonitoring during operation (% of participants experiencing this), radiation exposure (Gy) Hardware-related: Screw accuracy (non-validated classification used), Hojo classification (grade 0–2), subsidence rate (%), screw/rod breakage post-operatively (%), pseudoarthrosis post-operatively (%) Success of intervention: Treatment failure/success (defined by author), Offer of further surgery (no/%), re-operation rate (%)
Life impact	Clinician completed scores: JOA (JOA 0–17; mJOA 0–18), Odom's functional assessment (4 point scale from poor to excellent), Nurick/modified Nurick score (grade 0–5), Smiley-Webster functional outcome (4 point scale from poor to excellent), Katz index of independence in activities of daily living (0–6), Barthel index (0–100), SCIM (0–100), mRS (0–6), FIM (18–126), WISCI-2 score (0–20) PROMs: NDI (0–50), EQ-5Q-3L (3 level responses across 5 categories), EQ-5Q-3L (5 level responses across 5 categories), SF-12 and SF-36 (0–100), CSOQ (0–100), NPDI (9 item score with 3 subsections), ODI (0–100), PTNC (0–100), patient satisfaction (non-validated classification used), PSI (grade 1–4) Other assessments of Activities of Daily living: Ambulatory status (% walking independently), assessment of bowel/bladder function or assessment of gait (non-validated classification used), requirement for PEG (%), % voiding without catheter Other: Return to work (%), employment status, marital status, return to activities (non-validated classification used)
Resource use	Time to intervention (time), duration of bed-rest (time) Length of stay in ICU/hospital (days), discharge destination (location), hospital expenses (currency), unplanned re-admission rate (%), re-operation rate (%)
Adverse events	Complications of treatment (%), re-operation (%)

VAS, Visual analog scale; NRS, Numeric rating scale; JOA, Japanese Orthopaedic Association scoring system; SF-36, 36-Item Short Form Survey; CSOQ, cervical spine outcome questionnaire; NPDI, neck pain driving index; PTNC, post-traumatic neck score; PSI, patient satisfaction index; WISCI-2, Walking Index for Spinal Cord Injury 2 score; FIM, functional independence measure score; mRS, modified Rankin score; SCIM, spinal cord independence measure; ODI, Oswestry disability scale

often reported in larger and multicenter studies, complication rates were similar across all studies of different size and regardless of whether prospective and multicenter. Re-operation rates were more commonly reported with prospectively collected data (3.94% vs. 8.8%).

When comparing studies including patients of any age with studies assessing only patients over the age of 65, those assessing the older population only more commonly reported PROMs (35% vs.28.5%), functional assessments (36.3% vs.26.8%), mortality (92.5% vs.55.3%), hospital stay (23.8% vs.16.4%) and discharge destination (11.3% vs.3.8%). The ASIA/Frankel scales to classify neurological recovery, and radiological parameters, were less commonly used in patients over 65 than in the studies that assessed

patients of any age. Bony fusion was reported in 61.3% of studies only assessing patients over 65 years old, compared to 69.2% of other studies.

Studies assessing SCIs less commonly used NDI (14.6% vs. 10.7%), a validated pain score (25.8% vs 19.2%), PROMs (34.7% vs. 25.5%) and functional assessments (31.9% vs. 25.5%) when compared to those studies not including SCI patients. This phenomenon was accentuated in the AS/DISH population of studies, which all included patients with SCI and in which ASIA score and mortality were the most cited outcomes, 51.7% and 82.8%, respectively. Despite 59.3% of all papers describing inclusion of patients with SCI, the reporting of Frankel/ASIA in these was only 65.1%, with the rest of the papers not classifying their patients' neurological

Table 3 Heatmap showing percentage use of outcomes measured for included studies based on study design/study population

Outcome Measures	Number of patients*		Single/Multicentre		Retro-/prospective		Age **		SCI ***		AS/DISH		Type of Injury			Type of Treatment **			
	<40 (n=275)	>40 (n=261)	Single (n=472)	Multiple (n=64)	Retrospective (n=465)	Prospective (n=71)	Any (n=452)	>65yrs only (n=80)	Present (n=318)	Absent (n=213)	Present (n=29)	Absent (n=507)	C1-2 (n=310)	Subaxial (n=88)	Mixed (n=138)	Conservative (n=33)	Halo (n=26)	Surgery (n=296)	Mixed (n=178)
NDI	9.50%	15.70%	12.30%	14.10%	11.60%	18.30%	11.10%	21.30%	10.70%	14.60%	0.00%	13.20%	16.50%	12.50%	3.60%	30.30%	0.00%	11.50%	12.90%
EQ-5D	0.70%	2.30%	1.10%	4.70%	1.10%	4.20%	0.90%	5.00%	0.90%	1.40%	0.00%	1.60%	1.30%	2.30%	1.40%	6.10%	0.00%	1.40%	1.10%
Pain score	21.80%	22.20%	22.00%	21.90%	21.50%	25.40%	22.10%	22.50%	19.20%	25.80%	0.00%	23.30%	27.70%	18.20%	11.60%	30.30%	19.20%	25.30%	15.70%
Frankel	15.30%	10.00%	13.60%	6.30%	14.20%	2.80%	14.40%	3.80%	20.10%	1.90%	10.30%	12.80%	9.70%	23.90%	12.30%	12.10%	15.40%	15.20%	8.40%
ASIA	26.20%	27.60%	28.80%	12.50%	24.70%	40.80%	30.50%	7.50%	45.00%	0.50%	51.70%	25.40%	18.40%	48.90%	31.90%	12.10%	7.70%	33.80%	20.80%
Radiological Fusion	73.80%	62.10%	71.20%	45.30%	69.50%	59.20%	69.20%	61.30%	67.60%	69.00%	34.50%	70.00%	77.70%	63.60%	49.30%	60.60%	65.40%	73.00%	62.40%
Radiological Stability	45.10%	32.60%	41.70%	18.80%	38.30%	43.70%	40.90%	28.80%	37.70%	40.40%	13.80%	40.40%	42.30%	46.60%	26.80%	48.50%	38.50%	42.60%	31.50%
Radiological - Other	20.00%	18.80%	20.30%	12.50%	19.10%	21.10%	21.50%	7.50%	19.50%	18.30%	10.30%	19.90%	16.50%	31.80%	18.10%	15.20%	15.40%	22.60%	15.70%
Mortality	56.70%	64.80%	58.50%	76.60%	60.20%	63.40%	55.30%	92.50%	60.70%	61.00%	82.80%	59.40%	62.30%	56.80%	59.40%	75.80%	50.00%	54.40%	69.10%
Complications	75.60%	73.60%	74.60%	75.00%	74.60%	74.60%	74.80%	73.80%	74.50%	75.60%	72.40%	74.80%	78.10%	72.70%	68.10%	60.60%	69.20%	82.80%	64.00%
PROM	28.70%	29.90%	28.60%	34.40%	28.80%	32.40%	28.50%	35.00%	25.50%	34.70%	6.90%	30.60%	35.20%	27.30%	17.40%	36.40%	26.90%	30.40%	27.00%
Functional Assessment	26.50%	29.50%	28.60%	23%	27.10%	33.80%	26.80%	36.30%	25.20%	31.90%	27.60%	28.00%	33.20%	27.30%	16.70%	36.40%	11.50%	28.70%	27.50%
Hospital stay	12.40%	22.60%	15.00%	34.40%	15.50%	29.60%	16.40%	23.80%	16.40%	19.20%	13.80%	17.60%	17.70%	12.50%	19.60%	12.10%	15.40%	15.20%	22.50%
Discharge Destination	3.30%	6.50%	4.00%	10.90%	4.50%	7.00%	3.80%	11.30%	5.00%	4.70%	10.30%	4.50%	4.20%	1.10%	8.70%	6.10%	0.00%	3.00%	8.40%
Re-operation	10.90%	14.90%	12.70%	14.10%	8.80%	39.40%	13.70%	8.80%	15.10%	8.90%	20.70%	12.40%	11.90%	23.90%	8.00%	0.00%	3.80%	17.60%	9.00%

*One study was a protocol and included no patients; **in three studies this was not defined; ***in five studies this was unclear

function into a validated system. Re-operation rates were more commonly described in the SCI cohorts (15.1% vs. 8.9%).

When comparing studies assessing different types of injuries, the use of a validated pain score was more prevalent in studies assessing upper cervical injuries (27.7% vs. 11.6% in those assessing subaxial injuries only). However, studies assessing only subaxial injuries more commonly used the Frankel/ASIA grade for neurological outcome compared to studies assessing only upper cervical injuries, 23.9% vs. 9.7% for Frankel and 48.9% vs.18.4% for ASIA. Reporting of radiological fusion was more frequent in studies of upper cervical spine rather than subaxial cervical spine (77.7% vs.63.6%). Mortality, complication rates, PROMs, functional assessments, length of hospital stay and discharge destination were also reported more frequently in studies of upper rather than subaxial cervical spine fractures. Re-operation rates, however, were more often reported in subaxial injuries (23.9% vs.11.9%).

Studies reporting outcomes for patients treated purely conservatively more commonly reported PROMs (36.4%), functional outcomes (36.4%), pain scores (30.3%) and mortality (75.8%) than those assessing the use of only halo brace immobilization or surgery. Studies assessing patients managed surgically more frequently reported Frankel/ASIA grade (15.2%/33.8%) for neurological outcome, radiological fusion rates (73%), complication rates (82.8%), duration of hospital stay (15.2%), and re-operation rates (17.6%).

There was a dearth of studies from LEDCs, especially Africa (*n* = 15) and South America (*n* = 7), when compared to the output of North America (*n* = 178), Asia (*n* = 179) and Europe (*n* = 143). Table 4 shows PROMs, and functional assessments were more commonly used in Asia, Europe and Oceania than other continents. Specifically, NDI was more used in Asia (16.2%) and Europe (15.4%), compared to Oceania (9.1%), South America (12.9%), and Africa (6.7%). The EQ-5D score was chiefly used in European studies (4.2%). Studies from Asia more commonly reported a validated pain score (31.8%), ASIA/Frankel grade (54.8%), radiological fusion (79.3%), radiological stability (50.8%) and complications (79.3%) than European or North American studies. Duration of hospital stay was more commonly reported in North American studies (26.4%) than European (18.2%) or Asian (7.8%) studies. Discharge destination was more commonly documented in North American studies (8.4%) than European (4.9%) and Asian studies (0.6%).

Over time, there has been an increase in the reporting of validated pain scores (28% in 2020s vs.13.3% in 1980s), ASIA/Frankel grade (50.7% in 2020s vs. 26.7% in 1980s), complications (78.7% in 2020s vs.60% in 1980s), PROMs (34.7% in 2020s vs.13.33% in 1980s) and functional assessments (36% in 2020s vs.13.33% in 1980s). The use of

Table 4 Heatmap showing percentage use of outcomes measured for included studies based on geographical location of study

Outcome Measures	Africa (Total = 15)	Asia (Total = 179)	Europe (Total = 143)	N America (Total = 178)	Oceania (Total = 11)	S America (Total = 7)
NDI	6.70%	16.20%	15.40%	7.30%	9.10%	12.90%
EQ-5D	0.00%	0.60%	4.20%	0.60%	0.00%	0.00%
Pain score	26.70%	31.80%	22.40%	12.90%	9.10%	0.00%
ASIA/Frankel	46.70%	54.80%	38.50%	27.00%	0.00%	42.90%
Radiological Fusion	66.70%	79.30%	70.60%	58.40%	27.30%	57.10%
Radiological Stability	33.30%	50.80%	35.00%	32.00%	27.30%	28.60%
Radiological - Other	40.00%	19.00%	18.90%	19.10%	9.10%	28.60%
Mortality	60.00%	48.00%	70.60%	64.00%	72.70%	71.40%
Complications	86.70%	79.30%	74.80%	69.10%	63.60%	85.70%
PROM	26.70%	36.90%	34.30%	18.50%	27.30%	14.30%
Functional Assessment	20.00%	35.80%	35.70%	14.60%	45.50%	14.30%
Hospital stay	6.70%	7.80%	18.20%	26.40%	27.30%	28.60%
Discharge Destination	0.00%	0.60%	4.90%	8.40%	18.20%	14.30%
Re-operation	6.70%	13.40%	19.60%	8.40%	9.10%	0.00%

radiological parameters has declined with 62.7% of studies from the 2020s reporting radiological fusion compared to 86.7% in the 1980s and a similar decline in the reporting of radiological stability and other radiological outcomes (Table 5).

Timing of outcome measures

A clear record of schedule of timing of clinical and/or radiological assessment following injury was recorded in 304 (56.7%) of studies. The remainder of studies recorded outcomes at last follow-up with resultant variation in timing between patients, or did not state when the outcomes were measured. Where recorded, the median first timing of assessment following discharge was 3 months (interquartile range 3–6 months).

Response rates

Where available, the response rates of outcome measures used were analyzed. There were statistically significantly greater response rates in the studies assessing patients of any age compared to those assessing only patients over 65 years old (U = 6729.5, p = < 0.001). There were statistically significantly lower response rates in the studies using PROMs compared to those assessing only non-patient reported outcomes (U = 12,092, p = < 0.001). No difference in response rates was seen in studies assessing populations with or without SCI, nor when comparing studies in different continents.

Table 5 Heatmap showing percentage use of outcomes measured for included studies based on year of study

Outcome Measures	1980s (Total = 15)	1990s (Total = 65)	2000s (Total = 124)	2010s (Total = 256)	2020s (Total = 76)
NDI	0.00%	3.10%	7.30%	14.80%	22.70%
EQ-5D	0.00%	0.00%	0.00%	1.20%	5.30%
Pain score	13.30%	13.90%	12.90%	27.30%	28.00%
ASIA/Frankel	26.70%	32.30%	37.10%	40.20%	50.70%
Radiological Fusion	86.70%	72.30%	66.90%	68.00%	62.70%
Radiological Stability	60.00%	40.00%	41.30%	34.80%	45.30%
Radiological - Other	26.70%	21.50%	18.60%	18.00%	22.70%
Mortality	60.00%	52.30%	62.10%	60.20%	66.70%
Complications	60.00%	66.20%	75.00%	76.60%	78.70%
PROM	13.30%	18.50%	22.60%	34.40%	34.70%
Functional Assessment	13.30%	15.40%	21.00%	33.20%	36.00%
Hospital stay	0.00%	12.30%	17.70%	15.20%	32.00%
Discharge Destination	0.00%	3.10%	5.70%	4.70%	6.70%
Re-operation	6.70%	4.60%	9.70%	8.20%	42.70%

*1970s not included as only one study from this decade met inclusion criteria

Discussion

Summarizing existing outcome measures used

Our systematic review has identified heterogeneity in selection of outcome measures for cervical spine fracture research. Larger prospective studies more commonly employed PROMs as a method of collecting patient outcome data. Conversely, radiological outcomes were more commonly recorded in smaller, single unit retrospective studies. Larger prospective studies may be better resourced to collect more outcome data using validated tools, with the ability to prospectively follow-up patients and better ensure completion of questionnaires. Radiological outcomes may be more easily obtained, but the greater reporting of radiological outcomes in smaller retrospective studies may also reflect

the complexity of importing radiological source material from different centers or from larger populations. Also, the cost of mandating specific imaging in larger studies could be prohibitive.

Only a small proportion of the published literature focused on the elderly population (14.4%), but the elderly/frail population are increasingly becoming the main casualties of cervical fractures as the world's population ages [15]. Interestingly, we found studies assessing only patients over 65 years of age more commonly reported PROMs, pain scores, functional assessments, length of hospital stay, and discharge destination, but reported less frequently regarding radiological outcomes. This may reflect the growing evidence that radiological bony fusion is less common in the elderly/frail who sustain an odontoid fracture of the cervical spine, irrespective of treatment strategy [16–20]. Quality of

life outcomes may be better suited to this population. The majority of the papers included in our review reported on upper cervical spine fractures and surgical management. In comparison, conservative management strategies and their outcomes are less well defined [21]. This is an unmet need with the ageing population. A better understanding of the optimal methods of non-operative management of different types of cervical spine fractures in the elderly/frail and how best to measure clinical outcomes in these populations requires further scrutiny.

Outcome measures in patients with SCI were less likely to use PROMs and functional assessments, which was unexpected, given the importance of neurological function following recovery [22]. The use of a validated classification system (e.g., Frankel/ASIA) allows comparison of cohorts across different studies and quantification of recovery [23, 24]. Scivoletto et al. [25] sought expert opinion on use of outcome measurement tools after SCI, and identified clinicians most commonly used neurological function, pain, spasticity, gait and ability to self-care to define a patient's recovery. By contrast, in our study, the use of a validated grading classification to define neurological outcome was noted in only 57.8% of papers including SCI patients.

We identified a disparity in outcomes used in MEDCs compared to LEDCs. This has implications for the extrapolation of study outcomes between populations. It may reflect the financial cost and the logistical complexity of post-discharge follow-up of patients in LEDCs. The low proportion of intercontinental studies (0.4%) suggests that more could be done by MEDCs to include LEDC populations in future trials. Certainly the recent trend in global neurosurgery will help in this regard and specifically; studies need to address differences in measured clinical outcomes, and the patients' recovery goals between populations [26–28].

Patient involvement in reporting of outcomes

Since 1979, as the rate of reporting of radiological outcomes within studies of cervical spine fractures has declined, there has been a shift toward better understanding the functional outcomes of patients. While there are clinician reported outcome scores (e.g. JOA/FIM), PROMs recognize the importance of patients involvement/perspectives in the research process [29]. There is often discordance between what physicians and patients perceive as important outcomes [30]. We identified over twice as many studies in the 2020s reported PROMs compared to the 1980s. Pain scores were the most commonly used type of PROM, used in one in five studies. While the NDI was used in around one in ten studies, in total eleven different PROMs were identified that aim to quantify recovery following injury across numerous different domains.

Timing of outcome assessment

There was a lack of clear reporting on timing of assessments following injury. Duration of follow-up and the timing of outcome measure assessments following injury differed significantly between studies and were not always reported. The optimal duration of follow-up may be context dependent, for example, longer if assessing motor recovery following traumatic SCI, when recovery tends to plateau after a period of approximately 12–18 months [24], but shorter in frail/elderly patients with odontoid fractures who have a mortality rate as high as 34.1–37.5% at one year [31, 32]. Better reporting of the timing of outcome measurements would aid in comparisons of outcomes across study populations.

Strengths and limitations of this review

Using broad search terms, and including all treatment modalities, meant this review included a large number, and a wide range of studies. It is likely that we have identified all of the common outcome measures in use. Including non-English language studies may have provided a comparison of outcomes used in different cultures/languages. Nevertheless, with studies performed in every continent, the included outcome measures likely represent those used worldwide for cervical spine fractures.

As we aimed to understand the methods used to report outcomes, we did not assess the quality of the included studies or attempt any meta-analysis of outcome data. Including only studies meeting certain quality criteria may have systematically excluded those reporting certain outcome measures.

Development of a core outcome set

The heterogeneity of outcome measures for cervical spine fractures identified in this systematic review demonstrates the requirement for standardization in future clinical studies. Developing a COS for cervical spine fractures requires ascertainment of which outcome measures are optimal for the patient groups under study. The differences in outcome measures chosen in the existing literature show that these would need to be stratified by patient, population, and injury features. Determining which outcome measures are optimal would benefit from stakeholder input, especially from patients themselves. There is a paucity of literature assessing patients' beliefs and values regarding the definition of a 'good outcome' following a cervical fracture [33]. Patient self-assessment and expectations regarding their recovery will be context specific, and depend on the injury, their age, pre-existing frailty and comorbidities, alongside their values and beliefs. Defining the most appropriate PROMs for cervical spine fractures, as part of a COS would help standardize

outcome measures, allowing comparison across studies and modes of treatment. The involvement of patients in defining these COS is vital [34]. We have registered this requirement on the COMET database (https://www.comet-initiative.org/Studies/Details/2030) and plan to conduct a Delphi study to identify these ideal context-specific core outcomes in order to improve the quality of published outcomes for patients with cervical spine fractures.

Conclusions

Overall, the most commonly reported clinical outcomes were complications, radiological fusion and mortality. There was significant heterogeneity in the types of outcome reported across studies of differing populations. Less than a third of the included studies included PROMs, but use of PROMs was more common in larger/prospective studies, in studies assessing patients over 65 years of age and in studies assessing patients managed non-operatively. Over time, the use of radiological outcomes has declined, with a trend to greater use of PROMs. Response rates were poorer in the studies assessing patients over 65 years of age and when PROMs were used, highlighting the challenge in particular in the assessment of outcomes in the elderly cohort. Moreover, the review identified a shortage of studies from LEDCs and a lack of intercontinental studies examining outcomes following cervical spine fractures, which requires a global approach from the neurosurgical community to remedy.

Declarations

Conflict of interests The authors did not receive support from any organization for the submitted work. The authors have no relevant financial or non-financial interests to disclose.

References

1. Passias PG, Poorman GW, Segreto FA et al (2018) Traumatic fractures of the cervical spine: analysis of changes in incidence, cause, concurrent injuries, and complications among 488,262 patients from 2005 to 2013. World Neurosurg 110:e427–e437

2. Fredo HL, Rizvi SA, Lied B, Ronning P, Helseth E (2012) The epidemiology of traumatic cervical spine fractures: a prospective population study from Norway. Scand J Trauma Resusc Emerg Med 20:85

3. McDaid D, Park AL, Gall A, Purcell M, Bacon M (2019) Understanding and modelling the economic impact of spinal cord injuries in the United Kingdom. Spinal Cord 57(9):778–788

4. Merritt CH, Taylor MA, Yelton CJ, Ray SK (2019) Economic impact of traumatic spinal cord injuries in the United States. Neuroimmunol Neuroinflamm. https://doi.org/10.20517/2347-8659.2019.15

5. Baxter J, Lisk R, Osmani A et al (2021) Clinical outcomes in patients admitted to hospital with cervical spine fractures or with hip fractures. Intern Emerg Med 16(5):1207–1213

6. Joseph C, Delcarme A, Vlok I, Wahman K, Phillips J, Nilsson WL (2015) Incidence and aetiology of traumatic spinal cord injury in Cape Town, South Africa: a prospective, population-based study. Spinal Cord 53(9):692–696

7. Rahimi-Movaghar V, Sayyah MK, Akbari H et al (2013) Epidemiology of traumatic spinal cord injury in developing countries: a systematic review. Neuroepidemiology 41(2):65–85

8. McCaughey EJ, Purcell M, McLean AN et al (2016) Changing demographics of spinal cord injury over a 20-year period: a longitudinal population-based study in Scotland. Spinal Cord 54(4):270–276

9. Coster WJ (2013) Making the best match: selecting outcome measures for clinical trials and outcome studies. Am J Occup Ther 67(2):162–170

10. Williamson PR, Altman DG, Blazeby JM et al (2012) Developing core outcome sets for clinical trials: issues to consider. Trials 13:132

11. Gargon E, Gurung B, Medley N et al (2014) Choosing important health outcomes for comparative effectiveness research: a systematic review. Value Health 17(7):A435

12. Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M (2014) The COMET Initiative database: progress and activities from 2011 to 2013. Trials 15:279

13. Kirkham JJ, Gorst S, Altman DG et al (2015) COS-STAR: a reporting guideline for studies developing core outcome sets (protocol). Trials 16:373

14. Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR (2018) A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol 96:84–92

15. Bank M, Gibbs K, Sison C et al (2018) Age and other risk factors influencing long-term mortality in patients with traumatic cervical spine fracture. Geriatr Orthop Surg Rehabil 9:2151459318770882

16. Graffeo CS, Perry A, Puffer RC et al (2017) Deadly falls: operative versus nonoperative management of Type II odontoid process fracture in octogenarians. J Neurosurg Spine 26(1):4–9

17. Iyer S, Hurlbert RJ, Albert TJ (2018) Management of odontoid fractures in the elderly: a review of the literature and an evidence-based treatment algorithm. Neurosurgery 82(4):419–430

18. Lofrese G, Musio A, De Iure F et al (2019) Type II odontoid fracture in elderly patients treated conservatively: is fracture healing the goal? Eur Spine J 28(5):1064–1071

19. Pal D, Sell P, Grevitt M (2011) Type II odontoid fractures in the elderly: an evidence-based narrative review of management. Eur Spine J 20(2):195–204

20. Perry A, Graffeo CS, Carlstrom LP et al (2018) Fusion, failure, fatality: long-term outcomes after surgical versus nonoperative management of Type II odontoid fracture in octogenarians. World Neurosurg 110:e484–e489

21. Peck GE, Shipway DJH, Tsang K, Fertleman M (2018) Cervical spine immobilisation in the elderly: a literature review. Br J Neurosurg 32(3):286–290

22. Copley P, Jamjoom A, Khan S (2020) The management of traumatic spinal cord injuries in adults: a review. Orthop Trauma 34(5):255–265

23. Kirshblum S, Botticello A, Benedetto J et al (2020) A comparison of diagnostic stability of the ASIA impairment scale versus Frankel classification systems for traumatic spinal cord injury. Arch Phys Med Rehabil 101(9):1556–1562

24. Kirshblum S, Snider B, Eren F, Guest J (2021) Characterizing natural recovery after traumatic spinal cord injury. J Neurotrauma 38(9):1267–1284

25. Scivoletto G, Galli G, Torre M, Molinari M, Pazzaglia M (2019) The overlooked outcome measure for spinal cord injury: use of assistive devices. Front Neurol 10:272

26. Dewan MC, Rattani A, Fieggen G et al (2018) Global neurosurgery: the current capacity and deficit in the provision of essential neurosurgical care. Executive summary of the global neurosurgery initiative at the program in global surgery and social change. J Neurosurg 1:1–10

27. Haglund MM, Fuller AT (2019) Global neurosurgery: innovators, strategies, and the way forward. J Neurosurg 131(4):993–999

28. Lartigue JW, Dada OE, Haq M et al (2021) Emphasizing the role of neurosurgery within global health and national health systems: a call to action. Front Surg 8:690735

29. Weldring T, Smith SM (2013) Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). Health Serv Insights 6:61–68

30. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL (2008) Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care 14(8):530–539

31. Shafafy R, Valsamis EM, Luck J et al (2019) Predictors of mortality in the elderly patient with a fracture of the odontoid process. Bone Joint J 101-B(3):253–259

32. Venkatesan M, Northover JR, Wild JB et al (2014) Survival analysis of elderly patients with a fracture of the odontoid peg. Bone Joint J 96-B(1):88–93

33. Schoenfeld AJ, Bono CM (2011) Measuring spine fracture outcomes: common scales and checklists. Injury 42(3):265–270

34. Young B, Bagley H (2016) Including patients in core outcome set development: issues to consider based on three workshops with around 100 international delegates. Res Involv Engagem 2:25