



University for the Common Good

## **Properties of pain assessment tools for use in people living with stroke: systematic review**

Edwards, Sophie A.; Ioannou, Antreas; Carin-Levy, Gail; Cowey, Eileen; Brady, Marian; Morton, Sarah; Sande, Tonje A.; Mead, Gillian; Quinn, Terry J.

*Published in:*  
Frontiers in Neurology

*Publication date:*  
2020

*Document Version*  
Peer reviewed version

[Link to publication in ResearchOnline](#)

*Citation for published version (Harvard):*

Edwards, SA, Ioannou, A, Carin-Levy, G, Cowey, E, Brady, M, Morton, S, Sande, TA, Mead, G & Quinn, TJ 2020, 'Properties of pain assessment tools for use in people living with stroke: systematic review', *Frontiers in Neurology*.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

If you believe that this document breaches copyright please view our takedown policy at <https://edshare.gcu.ac.uk/id/eprint/5179> for details of how to contact us.

# **Title: Properties of pain assessment tools for use in people living with stroke: systematic review.**

**Authors:** Edwards SA<sup>1</sup>, Ioannou A<sup>2</sup>, Carin-Levy G<sup>3</sup>, Cowey E<sup>4</sup>, Brady M<sup>5</sup>, Morton S<sup>6</sup>, Sande TA<sup>7</sup>, Mead G<sup>6</sup>, Quinn TJ<sup>1</sup>

## **Author Affiliations:**

1. Institute of Cardiovascular and Medical Sciences, University of Glasgow, UK
2. Internal Medicine Department, Nicosia General Hospital, Cyprus.
3. School of Health Sciences, Queen Margaret University, Edinburgh, UK
4. School of Medicine, University of Glasgow, UK
5. NMAHP Research Unit, Glasgow Caledonian University, UK
6. Centre for Clinical Brain Sciences, University of Edinburgh, UK
7. Centre for Medical Informatics, Usher Institute, University of Edinburgh, UK

## **Corresponding Author:**

Dr Terry Quinn

Institute of Cardiovascular and Medical Sciences, University of Glasgow

New Lister Building Campus

Glasgow Royal Infirmary

Glasgow G4 0SF

[Terry.quinn@glasgow.ac.uk](mailto:Terry.quinn@glasgow.ac.uk)

@DrTerryQuinn

**KEYWORDS:** Stroke; stroke care; pain; assessment; evaluation; psychometric

## **ABSTRACT**

**Background:** Pain is a common problem after stroke and is associated with poor outcomes. There is no consensus on the optimal method of pain assessment in stroke. A review of the properties of tools should allow an evidence based approach to assessment.

**Objectives:** We aimed to systematically review published data on pain assessment tools used in stroke, with particular focus on classical test properties of: validity, reliability, feasibility, responsiveness.

**Methods:** We searched multiple, cross-disciplinary databases for studies evaluating properties of pain assessment tools used in stroke. We assessed risk of bias using the Quality Assessment of Diagnostic Accuracy Studies tool. We used a modified harvest plot to visually represent psychometric properties across tests.

**Results:** The search yielded 12 relevant articles, describing 10 different tools (n=1106 participants). There was substantial heterogeneity and an overall high risk of bias. The most commonly assessed property was validity (eight studies) and responsiveness the least (one study). There were no studies with a neuropathic or headache focus. Included tools were either scales or questionnaires. The most commonly assessed tool was the Faces Pain Scale (FPS) (6 studies). The limited number of papers precluded meaningful meta-analysis at level of pain assessment tool or pain syndrome. Even where common data were available across papers, results were conflicting e.g. two papers described FPS as feasible and two described the scale as having feasibility issues.

**Conclusion:** Robust data on the properties of pain assessment tools for stroke are limited. Our review highlights specific areas where evidence is lacking and could guide further research to identify the best tool(s) for assessing post-stroke pain. Improving feasibility of assessment in stroke survivors should be a future research target.

**Systematic review registration number:**PROSPERO CRD42019160679

Available

from:[https://www.crd.york.ac.uk/prospero/display\\_record.php?ID=CRD42019160679](https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019160679)

## **INTRODUCTION**

Pain is a common problem after stroke [1]. Estimates of the frequency of pain vary across papers, depending on the population assessed and whether the focus is incident or prevalent pain. Large cohorts of mild to moderate stroke survivor suggest pain incidence of around 10% [2], while in smaller cohorts figures range from 30% during the first months [3], to 48% at one year [4] and 43% at ten years [5] after stroke.

Post-stroke pain is associated with disability and reduced quality of life [1]. It is independently associated with fatigue [6], depression [7] and has been strongly linked with suicidality [8,9]. Pain after stroke can have a variety of aetiologies and manifestations, including: shoulder pain, headache, neuropathic pain and exacerbation of pre-existing pain. Pain symptoms can present at any point during stroke recovery and may progress to chronic pain if not recognised and treated appropriately.

The first step in managing post-stroke pain is recognition and measurement. However, management of pain has not always been given the same priority as other aspects of stroke care such as instituting secondary prevention [10]. Pain assessment is a complicated task made more challenging in the context of stroke. Since pain is a subjective experience, self-report scales and questionnaires are the most commonly employed pain assessment tools in clinical practice and pain may be part of a more general health related quality of life assessment [11]. However, stroke impairments such as cognitive decline and communication issues may make it difficult for stroke survivors to communicate the presence and experience of pain using these tools [12,13]. Other impairments such as visual issues or loss of motor skills may further complicate the use of self-completion questionnaires or visual analogue scales.

Accepting these caveats, there is a range of pain assessment tools available that could be used with stroke survivors. Some are generic, some are specific to a certain pain syndrome and some are developed exclusively for stroke. At present there is no consensus on the best approach to assessing post-stroke pain and no standardised tool is recommended for research or practice [14]. In the absence of a gold standard pain assessment in stroke survivors and with the great variety of assessment tools available, clinicians may struggle to know the most appropriate approach for their patients. The choice of assessment tools should be guided by evidence, particularly, the psychometric properties of the pain assessment tools available.

Classical test features such as validity and responsiveness have been described for certain pain tools, however, equally important are end-user evaluations such as acceptability and feasibility within the person's healthcare setting.

A summary of psychometric properties of pain assessment tools could help clinicians and researchers choose the most appropriate measure, highlighting strengths and limitations and also showing where new evidence is needed. Thus, we conducted a systematic review to compare methods of pain assessment following stroke with a particular focus on properties of validity, reliability, feasibility and responsiveness.

## **METHODS**

We performed a systematic review, following best practice [15] and where appropriate Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guidance [16]. Two assessors (SE, TQ) performed all aspects of title selection, data extraction and analyses with disagreements resolved through discussion.

As our focus was test properties, we structured our review question using the format recommended for test accuracy evidence synthesis [17].

- Index test: Any measure of pain that gives an objective read out.
- Reference standard: Any measure that provides data on the classical test properties of interest namely validity, reliability, feasibility and responsiveness.
- Condition: Stroke of any kind and at any stage in stroke journey.
- Setting: Any healthcare setting.

**Search strategy:** We searched the following databases, chosen to represent the various disciplines that may assess post-stroke pain: Medline (Ovid), Embase (Ovid), CINAHL (EBSCO) and PsychInfo (EBSCO). All were searched from inception to 1<sup>st</sup> May 2020. Search concepts were 'stroke' and 'pain' and 'assessment'. We used validated search filters for 'stroke' and 'pain', taken from the relevant Cochrane review group (Supplementary materials). We complemented our search by contacting members of an international stroke pain research group to ensure we had not missed relevant studies.

We screened titles, abstracts and then full text to inform decisions on inclusion. Forward and backward citation searching was conducted for relevant studies using Web of Science functionality. As a test of search validity we pre-specified two papers (one original research

and one review) that should be returned on our literature search [1, 18]. As a further test we cross-checked our included papers with a systematic review of pain assessment in aphasia, recognising that the topics were distinct but were likely to have considerable overlap [14].

**Selection criteria:** The population of interest was adult stroke survivors at any stage of recovery. We did not include traumatic brain injury. If a mixed population was included, stroke had to represent more than 75% of the group. The test of interest was any form of pain assessment, including scales, questionnaires, observations and other patient reported outcome measures. Outcomes of interest were psychometric properties of the tools as defined below. We included studies of any quantitative design, conducted in any healthcare setting, noting setting as part of our data extraction. We only included studies published in peer reviewed journals but applied no other restrictions.

**Data collection process and data items:** We designed and piloted a bespoke data collection form using the paper that original research paper that informed our internal validation [18].

We collected data on the following:

Study details: publication date, country, study design (*i.e.* cross-sectional, prospective, retrospective), psychometric properties assessed (validity, feasibility, intra/inter-reliability, responsivity), sample size.

Stroke details: stroke classification (for example ischaemic or haemorrhagic), time since stroke, setting (classified as: acute stroke unit, rehabilitation, outpatient, community, using descriptions in the original paper), inclusion/exclusion criteria in original study, noting if there were specific exclusions relating to language or cognition.

Pain assessment: type of pain (see below), method(s) of pain assessment (*i.e.* pain scales, questionnaires, stroke specific or generic), pain assessor(s) (*i.e.* researcher or clinical discipline). For articles comparing multiple methods of pain assessment, we included all tools and recorded the primary pain assessment tool.

**Categorization of pain syndromes:** We categorised pain using the following pre-specified labels: neuropathic, nociceptive (noting the site *i.e.* lower limb), headache or experimental (*i.e.* investigator induced pain). We classified stroke shoulder pain as a distinct category as it can include both nociceptive and neuropathic elements. Our pain classification was based on the description in the original paper. Where the nature of the pain syndrome was not clear,

two reviewers (SE, TQ) discussed and came to consensus. For some papers, lack of detail precluded applying any label with certainty, and these were categorised as ‘non-specified’.

**Psychometric properties:** We were interested in the following psychometric properties: validity, reliability, feasibility, responsiveness. These were defined as [19, 20]:

- **Validity:** the extent to which an instrument measures what is intended, in this case, is the tool a measure of pain? The concept of ‘accuracy’ would be included as a measure of validity.
- **Reliability:** the internal consistency of an instrument, and the degree to which it is free from error on repeated. We included measures of inter-observer, intra-observer and internal reliability.
- **Feasibility:** usability, and acceptability of an instrument from the perspective of assessors and those being assessed.
- **Responsiveness:** the ability of the instrument to distinguish clinically important changes over time.

On initial scoping it became clear that a traditional quantitative meta-analysis would not be possible, due to the substantial clinical heterogeneity across studies in terms of populations assessed, methods used, nature of pain assessments and psychometric properties described. To allow cross-study comparisons, we created summary measures of the study findings at the level of the psychometric property studied. Our categorisation was based on the conclusions of the original paper and was agreed by consensus. We classified results as positive, neutral or inconclusive.

**Risk of bias:** We assessed risk of bias for included studies at the outcome level. Two (SE, TQ) investigators individually assessed papers and agreed final grading. No single quality assessment tool would be suitable for the variety of methodologies that were included in our eligible papers. We elected to use the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool [21]. QUADAS-2 is designed for assessing studies of test accuracy and uses a framework suited to our review with assessment of bias and applicability across four domains: patient selection, index tests, reference standard, flow and timing [17]. As recommended, we took the original QUADAS-2 anchoring statements and modified to suit our review (modified domain questions included in supplementary materials). We used

*robvis* R package software to create summary ‘traffic light’ plots [22]. Due to the limited number of studies and heterogeneity in summary measures we did not perform quantitative assessment for publication bias.

**Evidence Synthesis:** We created two summary tables (Table 1, Table 2): the first describes key characteristics of the included articles and the second summarises their results. Our data were heterogeneous and required representation of differing constructs across various axes. To allow a visual representation that included pain syndrome, pain assessment tool and results of psychometric testing across various constructs we created a visual plot using a modified harvest plot [23]. We created a matrix that plotted results by pain assessment tool (we created space in the plot for subcategorising by pain scales and questionnaires) against each psychometric property of interest. We colour-coded according to pain type with one unit of plot space per study/experiment and then assigned the results of the study as positive (above a horizontal line of no effect), neutral (below the line) or inconclusive (crossing the line).

## RESULTS

The initial search yielded 2851 articles, with 12 [9,18,24-33] papers (n=1106 participants) meeting the inclusion criteria (Figure 1). Our search results suggested a valid search as they included the two pre-selected papers and had all the relevant studies from the previous aphasia review. The number of participants ranged from 19 to 388. The most commonly employed design was cross-sectional (n=6) with the majority of studies (n=6) conducted in a rehabilitation setting (Table 1, Supplementary Materials).

In total, 10 different pain scales and questionnaires were assessed across the 12 studies (Table 1). These were: Visual Analogue Scale (VAS [differing scales described as VAS]), the Faces Pain Scale (including a revised version), Numerical Rating Scale, and various combinations of these; the Pain Assessment Scale for Seniors with Severe Dementia-II (PACSLAC-II), and three questionnaires: AbilityQ, ShoulderQ and the neuropathic pain diagnostic questionnaire (DN4). Of the included assessments, only the ShoulderQ was developed specifically for stroke. The Faces Pain Scale was the most commonly reported, with a version used in six of the 12 studies.



Where a pain category was described, the most commonly studied was shoulder pain. Neuropathic pain and Headache were not studied, except in those papers that did not differentiate pain type. There was heterogeneity in the tools assessed for each pain category, with no pain category having more than two studies using a common tool (Table 3).

There was a high risk of bias detected in the majority of included papers (n=8; Figure 2). Highest risk of bias and issues with generalisability was seen for the domain of patient selection (n=10; judged high risk). This was due to exclusion of patients for whom pain assessment would be expected in clinical practice, including those with pre-stroke pain (n=5 papers), aphasia (n=3) and cognitive impairment (n=3). There was poor reporting of study methods relevant to the risk of bias assessment, particularly around blinding of results when a study compared scales. Only four papers were judged to have overall low risk of bias [18,31-33]

We created a visual synthesis of the psychometric properties of the tools used to assess pain as a modified harvest plot (Figure 3). The harvest plot approach allows visual display of data across several axes in one figure. We represented each study as a single unit (square), and colour coded based on pain type. Columns delineated differing assessment tools and rows the properties of interest. A horizontal line that bisected each row was a line of uncertain effect, if a study claimed that the psychometric property of interest was 'good' i.e. acceptable for clinical use then the study was placed above the line, if the paper reported that the study was 'poor' i.e. would not be suitable it was placed below the line.

All psychometric domains of interest were reviewed by at least one article, although the statistical approach to these assessments varied. Validity was the psychometric property evaluated most frequently (n=8), and responsiveness was only considered by one article. In general pain scales assessed were judged to be valid measures by the authors of the studies, with only two studies reporting concerns around validity (Figure 3). A version of the Faces Pain Scale was the most commonly assessed pain, with evaluations of validity (n=3), reliability (n=3) and feasibility (n=2). However, results were conflicting, for example feasibility of FPS was assessed as good, neutral and poor across the studies (Figure 3).

## **DISCUSSION**

We aimed to systematically review the psychometrics of pain assessment tools when used with stroke survivors. We found a limited literature with substantial heterogeneity in the tools used, the research methods employed and the properties assessed. The available data were limited by risk of bias and modest sample sizes. Thus, we are unable to recommend a preferred tool based on published psychometric properties. However, through our evidence synthesis, we have highlighted important evidence gaps that can inform the direction of future research activity in the pain assessment space.

Our mapping of the evidence using the harvest plot demonstrates the many limitations in the evidence base. Of the four key psychometric properties, there was little information on reliability, and responsiveness. Even where there was a portfolio of papers on a single tool it was difficult to draw conclusions. There were more studies on visual scales than questionnaires, with few studies using a scale specifically developed for stroke and no studies with a neuropathic or headache pain focus.

Our findings of inconsistent and inconclusive evidence are not unique to stroke. A previous review of pain assessment in aphasia concluded that ‘a feasible, reliable and valid pain assessment instrument is not yet available’ [14]. Dementia is another clinical condition where pain is common but potentially difficult to assess. Although there is more published literature on dementia pain assessment tools [34], conclusions of reviews are similar ‘limited evidence about reliability, validity and clinical utility’ [35]. This seems a missed opportunity, as well as the clinical importance, pain assessment could be a useful research outcome [36].

Our assessment of risk of bias suggests common areas of concern particularly around reporting and generalisability. Exclusion of stroke survivors with aphasia, dementia or comorbidity threatens the external validity of study results. Similar exclusions have been demonstrated in other aspects of stroke assessment [37]. Certain scales may not be suitable for all stroke impairments, but simply excluding those people who may struggle to complete an assessment creates bias in any resulting estimates.[38]

Our review has several strengths. We performed a comprehensive search, followed best practice guidance and embedded internal validation steps. Given the disparate nature of relevant studies, we used non-traditional methods for evidence synthesis and assessment of quality. There are limitations to our approach. Despite internal and external validity steps we

may have missed relevant papers. We were not able to perform quantitative meta-analysis either at an aggregate level or at the level of differing pain types, but instead used a relatively novel method of visual data synthesis. Our modified harvest plot approach gives a summary of the totality of the data across various axes, allowing for visual comparisons across tools. This approach could be applied in other complex reviews with substantial heterogeneity in the supporting literature.

Despite the prevalence of post-stroke pain, studies describing the best way to assess for this problem are limited in number and quality. Our evidence mapping and quality assessments highlight particular pain syndromes and tests that have no empirical evidence base. No pain assessment had sufficient data to be considered definitive and further, robust research for any pain tool would be a welcome addition.

In light of this uncertainty what conclusions can be made? Patient based scales, such as faces pain scale, seem to have the most supporting evidence and are a valid means to assess pain. Our review suggests there are many evidence gaps requiring future research, but methods to improve feasibility of assessment seem an important target.

**Conflict of Interest:** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Author Contributions:** SAE contributed to all aspects of searching, data extraction and analysis, provided critical review and contributed to draft manuscripts. IA assisted with data extraction, provided critical review and contributed to draft manuscripts. GLC provided critical review, assisted with formatting and contributed to draft manuscripts. EC provided critical review, assisted with formatting and contributed to draft manuscripts. MB provided critical review, expert aphasia advice and contributed to draft manuscripts. SM provided critical review and contributed to draft manuscripts. TAS provided critical review and contributed to draft manuscripts. GM devised the study question, coordinated the team, contributed to draft manuscripts. TQ provided critical review and contributed to draft manuscripts.

**Funding:** SAE was supported by a University of Glasgow studentship; the Pain in Stroke Research Group is supported by a British Association of Physicians and National Institute of Health Research Grant.

**Acknowledgments:** We acknowledge the support of the Pain in Stroke Research Group.

**Table 1:** Key Characteristics of included papers

**Table 1:** Characteristics of included papers (shortened version)

Author/s	Study Design	Psychometric properties assessed	Number included	Age (years) (mean, SD)	Stroke setting	Exclusion Criteria	Type of pain	Pain Assessment tool	Pain Assessor
1. Benaim [9]	Cross-sectional	Validity, Reliability	127	63 ± 8	Rehabilitation	cognitive impairments, psychiatric disorders	Shoulder Pain	FPS	Unknown
2. Chuang [31]	Prospective	Reliability	50	52.6 ± 11.0	Outpatient	other acute pain conditions, major medical problems, psychological impairments, aphasia	Arm/Shoulder pain	v-NPRS-FPS	Clinical staff (Rehabilitation physicians)
3. Dogan [26]	Case control	Validity	60 including non-stroke control (n=30)	64.2 ± 9.42	Rehabilitation	Pre-existing pain conditions, cognitive impairment, aphasia	Shoulder Pain	FPS	Unknown
4. Korner-Bitensky [27]	Cross-sectional	Validity	90	Not available	Rehabilitation	cognitive impairments, central post-stroke pain syndrome	Experimental (thermal)	10-cm v-VAS	Clinical staff (SLT), Researcher
5. Price [18]	Case control	Feasibility, Validity	144 including non-stroke	72.5 mean	Acute stroke unit	reduced conscious level or dysphasic	Experimental (pressure)	v/m/h-VAS	Researcher

			controls (n=48)						
6. Smith [25]	Retrospective	Feasibility	388	77 (IQR:66–86)	Acute stroke unit	subsequent strokes	Not specified	FPS and/or NRS	Clinical staff (Nurses)
7. Roosink [24]	Cross- sectional	Validity	19	57.5 ± 7.5	Rehabilitation	other chronic pain conditions, neurological deficits	Shoulder Pain	DN4	Unknown
8. Turner- Stokes (2003) [28]	Cross- sectional	Validity, Reliability, Feasibility	49	52.6 ± 3.1	Rehabilitation	not specified	Shoulder Pain	AbilityQ, ShoulderQ	Researcher
9. Turner- Stokes (2006) [29]	Retrospective	Responsiveness	30	47.2 ± 2.2	Rehabilitation	not specified	Shoulder Pain	AbilityQ, ShoulderQ	Clinical staff (Nurses)
10. Mandysova [30]	Cross- sectional	Validity, Reliability, Feasibility	80	71.0 ± 13.7 (range 22–94)	Acute stroke unit	reduced conscious level	Not specified	VAS/NRS, NRS, FPS-R	Researcher
11. Pomeroy [32]	Prospective	Reliability	33	74 (range 57–89)	Community	reduced conscious level, other pain conditions, no irregular pain medication, no	Shoulder Pain	10-cm v- VAS	Clinical staff (Physiotherapist)

---

neurological/MSK  
disorders

---

12. Soares [33]	Cross- sectional	Reliability, Validity	36	61 median (range 46 - 71.75)	Acute stroke unit	neurological disorders	Experimental (mechanical)	PACSLAC- II	Clinical staff (Neurology nurses)
--------------------	---------------------	--------------------------	----	---------------------------------	-------------------	------------------------	------------------------------	----------------	---

---

*Study design and setting were categorised and agreed by two raters (SE, TQ)*

*FPS = Faces Pain Scale; NRS/NPRS = Numerical Rating Scale; VAS = Visual Analogue Scale, v-/m-/h- = vertical/mechanical/horizontal*

*NPRS-FPS and VAS/NRS indicate combined versions of scales*

*DN4 = neuropathic pain diagnostic questionnaire; PACSLAC-II = Pain Assessment Scale for Seniors with Severe Dementia-II*

*SLT = Speech and Language Therapy*

*N.B. more comprehensive version of table is available in Supplementary Materials*



**Table 2:**Summary of results from included articles

Author/s	Pain assessment (comparator)	Results
1. Benaim [9]	FPS (VAS, VRS)	<ul style="list-style-type: none"> <li>• <i>Validity</i>: Correlation of FPS with VAS and VRS in both left and right hemisphere stroke (<math>r=0.65-0.82</math>)</li> <li>• <i>Reliability</i>: <ul style="list-style-type: none"> <li>– <i>Inter-rater</i>:<math>K:0.64</math> (SE=0.11) and <math>K:0.44</math> (0.09) in left and right hemisphere stroke respectively.</li> <li>– <i>Intra-rater</i>:<math>K:0.74</math> (0.13) and <math>K:0.53</math> (0.10) in left and right hemisphere stroke respectively.</li> </ul> </li> <li>• <i>Feasibility</i>: FPS was preferred in left hemisphere stroke, VAS was preferred in right hemisphere stroke.</li> </ul>
2. Chuang [31]	v-NPRS-FPS	<ul style="list-style-type: none"> <li>• <i>Reliability (intra-rater)</i>:<math>ICC=0.82</math> (SE=0.81), [smallest real difference=1.87].</li> <li>• No significant systematic bias between repeated measurements for NPRS-FPS.</li> <li>• High level of stability and minimal temporal variation, range of limits of agreement (-2.50 to 1.90)</li> </ul>
3. Dogan [26]	FPS (VAS, LPS, NRS)	<ul style="list-style-type: none"> <li>• <i>Validity</i>: Correlation of FPS with other pain scales in both groups (<math>r=0.95-0.97</math> and <math>0.67-0.93</math>, respectively).</li> </ul>
4. Korner-Bitensky [27]	10-cm v-VAS	<ul style="list-style-type: none"> <li>• <i>Validity</i>: No between group difference in pain discrimination (<math>p=0.75</math>).</li> <li>• Repeated-measures ANOVA revealed no effect of group.</li> </ul>
5. Price [18]	v/m/h-VAS FPRS, NRS	<ul style="list-style-type: none"> <li>• <i>Feasibility</i>: Inability to complete scales was associated with stroke (<math>P 0.01</math>).</li> <li>• Association between completion and milder stroke (<math>P&lt;0.01</math>).</li> <li>• Associations (<math>P&lt;0.05</math>) between impairments tested and inability to complete scales.</li> </ul>
6. Smith	FPS, NRS	<ul style="list-style-type: none"> <li>• <i>Feasibility</i>: 13.4% individuals unable to provide a meaningful response to either FPS or NRS.</li> </ul>

[25]		<ul style="list-style-type: none"> <li>• Associations between inability to self-report pain and: stroke severity (P&lt;0.001), aphasia severity (P&lt;0.001) consciousness (P&lt;0.001).</li> <li>• <i>Validity</i>: Associations between inability to self-report pain and mortality (P&lt;0.001).</li> </ul>
7. Roosink [24]	DN4  (NRS)	<ul style="list-style-type: none"> <li>• <i>Validity</i>: DN4+ classified patients reported: constant pain [DN4+:n=4 (44%); DN4-:n=0 ] higher pain intensity [DN4+=4.7 (SD=2.9); DN4- =2.5 (SD=2.4)] higher impact of pain on daily living DN4+=5.9 (SD= 4.8), DN4- =2.0 (SD=2.6) more frequent loss of cold sensation [DN4+:n=7 (78%); DN4-:n=2 (20%)]</li> <li>• Signs and symptoms suggestive of neuropathic or nociceptive pain corresponded to DN4+ and DN4- respectively.</li> </ul>
8. Turner- Stokes (2003) [28]	AbilityQ, ShoulderQ  (VAS)	<ul style="list-style-type: none"> <li>• <i>Validity</i>: VAS agreement <math>\pm 1</math> on a 10-point scale was 36–59% with intraclass correlation coefficients 0.50–0.60 (<math>p &lt; 0.01</math>).</li> <li>• <i>Reliability</i>: Agreement for individual questions 55- 88%; <math>K: 0.07-0.79</math></li> <li>• Repeatability of ShoulderQ 36–72%, <math>K: 0.16–0.56</math>.</li> <li>• <i>Feasibility</i>: N=31 (63%) required help in completing AbilityQ.</li> </ul>
9. Turner- Stokes (2006) [29]	AbilityQ, ShoulderQ  (VGRS)	<ul style="list-style-type: none"> <li>• <i>Responsiveness</i>: Changes on VGRS associated with verbal reports of improvement (<math>r: 0.67</math>, <math>P &lt; 0.001</math>).</li> <li>• Responders demonstrated significant change in VGRS and verbal scores, whereas non-responder group did not. <ul style="list-style-type: none"> <li>– A change in summed VGRS score of <math>\geq 3</math> showed 77% sensitivity and 91.3% specificity for identifying responders, with a positive predictive value of 93.3%. Summed VGRS scores of <math>\leq 2</math> had a negative predictive value of 73.3%.</li> </ul> </li> </ul>
10. Mandysova [30]	VAS/NRS, NRS, FPS-R	<ul style="list-style-type: none"> <li>• <i>Validity</i>: n=19 (24%) reported pain using at least one scale.</li> <li>• Spearman correlation was 0.997 (<math>p &lt; 0.001</math>) between VAS/NRS and NRS.</li> <li>• <i>Feasibility</i>: NRS had the highest preference ranking (ranking first or second in 75% cases).</li> </ul>
11.	10-cm v-VAS	<ul style="list-style-type: none"> <li>• <i>Inter-rater reliability</i>: ICC: 0.79 for intensity, 0.75 for frequency and 0.62 for affective response.</li> </ul>

Pomeroy [32]	<ul style="list-style-type: none"> <li>• Wide limits of agreement and significant rater bias reported for 6/27 ratings.</li> <li>• <i>Intra-rater reliability</i>: ICC: 0.70 for intensity, 0.77 for frequency and 0.69 for affective response.</li> </ul>
12. Soares [33]	<ul style="list-style-type: none"> <li>• <i>Validity</i>: PACSLAC-II differentiated 4.5-lb stimulus versus 2-lb (p=0.03) or 0lb (p=0.05).</li> <li>• <i>Reliability (internal)</i>: Cronbach <math>\alpha</math>: 0.87, 0.94, and 0.96 for weights of 0, 2, and 4.5 lb, respectively.</li> </ul>

**Abbreviations:** FPS = Faces Pain Scale; NRS/NPRS = Numerical Rating Scale; VAS = Visual Analogue Scale, LPS = Likert Pain Scale, FPRS = Four-point rating scale, v-/m-/h- = vertical/mechanical/horizontal, visual graphic rating scale (VGRS), NPRS-FPS and VAS/NRS indicate combined versions of scales  
DN4 = neuropathic pain diagnostic questionnaire (DN4+ = neuropathic pain reported; DN4- = no neuropathic pain reported); PACSLAC-II = Pain Assessment Scale for Seniors with Severe Dementia-II

**Table 3:** Cross-tabulation of pain assessment tool and post stroke pain syndrome

	VAS	VAS-NRS	FPS	FPS-NRS	NRS	VRS	ShoulderQ	PACSLAC-11	DN4
<b>Shoulder/Arm Pain</b>	1	0	2	1	0	0	2	0	1
<b>Experimental</b>	2	0	0	0	0	0	0	1	0
<b>Not specified</b>	0	1	2	0	2	0	0	0	0
<b>Neuropathic</b>	0	0	0	0	0	0	0	0	0
<b>Headache</b>	0	0	0	0	0	0	0	0	0

**Figure 1:**PRISMA Flow chart for selection of studies for systematic review

The first search was performed on 31<sup>st</sup> July 2019; to ensure the review was up to date we ran a repeat search on 08/05/2020. The PRISMA contains an aggregate of both searches.’

**Figure 2:**Traffic Light plot for risk of bias in individual studies

**Figure 3:**Harvest plot of psychometric evaluation of pain scale according to the 12 included studies.

*Each unit represents a differing study.*

*Colour coding is used to represent differing pain types.*

*Position around horizontal line describes paper conclusions regarding the property of interest, where above the line indicates ‘good’, below the line indicates ‘poor’ and on the line indicates ‘uncertain’. Full description given in main manuscript.*

**Abbreviations:** *VAS=Visual Analogue Scale; NRS=Numerical Rating Scale; FPS=Faces Pain Scale; VRS=Visual Rating Scale; ShoulderQ=Shoulder pain questionnaire; PACSLAC-II=Pain Assessment Scale for Seniors with Severe Dementia-II; DN4=neuropathic pain diagnostic questionnaire;*

## REFERENCES

1. Harrison RA, Field TS. Post stroke pain: Identification, assessment, and therapy. *Cerebrovasc Dis.* 2015;39(3-4):190-201.
2. O'Donnell MJ, Diener H-C, Sacco RL, Panju AA, Vinisko R, Yusuf S. Chronic Pain Syndromes After Ischemic Stroke. *Stroke.* 2013;44:1238–1243.
3. Paolucci S, Iosa M, Toni D, Barbanti P, Bovi P, Cavallini A, Candeloro E, Mancini A, Mancuso M, Monaco S, et al. Prevalence and Time Course of Post-Stroke Pain: A Multicenter Prospective Hospital-Based Study. *Pain Med.* 2015;17:pvn019. doi:10.1093/pm/pnv019
4. Naess H, Lunde L, Brogger J. The Triad of Pain, Fatigue and Depression in Ischemic Stroke Patients: The Bergen Stroke Study. *Cerebrovasc Dis.* 2012;33:461–465.
5. Jönsson A-C, Delavaran H, Iwarsson S, Ståhl A, Norrving B, Lindgren A. Functional Status and Patient-Reported Outcome 10 Years After Stroke. *Stroke.* 2014; 45:1784–1790.
6. Hoang CLN, Salle J, Mandigout S, Hamonet J, Macian-Montoro F, Daviet J. Physical factors associated with fatigue after stroke: An exploratory study. *Top Stroke Rehabil.* 2012;19(5):369-76.
7. Lundström E, Smits A, Terént A, Borg J. Risk factors for stroke-related pain 1 year after first-ever stroke. *Eur J Neurol.* 2009;16(2):188-93.
8. Tang WK, Liang H, Mok V, Ungvari GS, Wong K. Is pain associated with suicidality in stroke? *Arch Phys Med Rehabil.* 2013;94(5):863-6.
9. Benaim C, Froger J, Cazottes C, Gueben D, Porte M, Desnuelle C, et al. Use of the faces pain scale by left and right hemispheric stroke patients. *Pain.* 2007;128(1-2):52-8.
10. KS McArthur, TJ Quinn, P Higgins, P Langhorne. Post-acute care and secondary prevention after ischaemic stroke. *BMJ.* 2011;342, d2083
11. Dansie EJ, Turk DC. Assessment of patients with chronic pain. *Br J Anaesth.* 2013;111(1):19-25.
12. Delpont B, Blanc C, Osseby GV, Hervieu-Bègue M, Giroud M, Béjot Y. Pain after stroke: A review. *Rev Neurol.* 2018;174(10):671-4.
13. TJ Quinn, E Elliott, P Langhorne. Cognitive and mood assessment tools for use in stroke. *Stroke.* 2018;49:483-490
14. de Vries NJC, Sloot PH, Achterberg WP. Pain and pain assessment in stroke patients with aphasia: A systematic review. *Aphasiology.* 2017;31(6):703-19.
15. Shenkin SD, Harrison JK, Wilkinson T, Dodds RM, Ioannidis JPA. Systematic reviews: guidance relevant for studies of older people. *Age Ageing.* 2017;46:722-728.
16. Moher D, Liberati A, Tetzlaff J, Altman D.G, Group P. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *J Clin Epidemiol.* 2009;62: 1006-12.
17. Y Takwoingi, TJ Quinn. Review of Diagnostic Test Accuracy (DTA) studies in older people. *Age and ageing.* 2018;47:349-355.

18. Price CIM, Curless RH, Rodgers H, Price CI, Curless RH, Rodgers H. Can stroke patients use visual analogue scales? *Stroke*. 1999;30(7):1357-61.
19. Milne A, Johnson JA, Tennant M, Rudnisky C, Dryden DM. 2012. Measuring health-related quality of life for patients with diabetic retinopathy. Rockville (MD): Agency for Healthcare Research and Quality (US). Available at: <https://www.ncbi.nlm.nih.gov/books/NBK248340/> (accessed Jan 2020)
20. Qi S, Diane J, Kay D. The psychometric properties, feasibility and utility of behavioural-observation methods in pain assessment of cognitively impaired elderly people in acute and long-term care: A systematic review. *JBI Libr Syst Rev*. 2012;10(17):977-1085.
21. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-36.
22. McGuinness LA. 2019. robvis: An R package and web application for visualising risk-of-bias assessments. Available at: <https://github.com/mcguinlu/robvis> (accessed Jan 2020)
23. Ogilvie D, Fayter D, Petticrew M, Sowden A, Thomas S, Whitehead M, et al. The harvest plot: A method for synthesising evidence about the differential effects of interventions. *BMC Med Res Methodol*. 2008 01;8:8.
24. Roosink M, van Dongen, Robert T. M., Renzenbrink GJ, IJzerman MJ. Classifying post-stroke shoulder pain: Can the DN4 be helpful? *Eur J Pain*. 2011;15(1):99-102.
25. Smith JH, Bottemiller KL, Flemming KD, Michael Cutrer F, Strand EA. Inability to self-report pain after a stroke: A population-based study. *Pain*. 2013;154(8):1281-6.
26. Dogan SK, Ay S, Oztuna D, Aytur YK, Evcik D. The utility of the faces pain scale in the assessment of shoulder pain in turkish stroke patients: Its relation with quality of life and psychological status. *Int J Rehabil Res*. 2010;33(4):363-7.
27. Korner-Bitensky N, Kehayia E, Tremblay N, Mazer B, Singer F, Tarasuk J, et al. Eliciting information on differential sensation of heat in those with and without poststroke aphasia using a visual analogue scale. *Stroke*. 2006;37(2):471-5.
28. Turner-Stokes L, Rusconi S. Screening for ability to complete a questionnaire: A preliminary evaluation of the AbilityQ and ShoulderQ for assessing shoulder pain in stroke patients. *Clin Rehabil*. 2003;17(2):150-7.
29. Turner-Stokes L, Jackson D. Assessment of shoulder pain in hemiplegia: Sensitivity of the ShoulderQ. *Disabil Rehabil*. 2006;28(6):389-95.
30. Mandysová P, Nedvěďová A, Ehler E. A comparison of three self-report pain scales in Czech patients with stroke. *Cent Eur J Nurs Midwifery*. 2017;8(1):572-9.
31. Chuang L, Wu C, Lin K, Hsieh C. Relative and absolute reliability of a vertical numerical pain rating scale supplemented with a faces pain scale after stroke. *Phys Ther*. 2014;94(1):129-38.
32. Pomeroy VM, Frames C, Faragher EB, Hesketh A, Hill E, Watson P, et al. Reliability of a measure of post-stroke shoulder pain in patients with and without aphasia and/or unilateral spatial neglect. *Clin Rehabil*. 2000;14(6):584-91.

33. Soares CD, Panuganti PK, Shrivastava A, Aroor S, Keinath KM, Bromagen MC, et al. Experimental pain assessment in patients with poststroke aphasia. *Neurology*. 2018;91(9):e793-9.
34. Harrison JKH, Noel-Storr AH, Demeyere N, Reynish EI, Quinn TJ. Outcome measures in a decade of dementia and mild cognitive impairment research. *Alz Res Therapy*. 2016;48:8
35. Lichtner V, Dowding D, Esterhuizen P, Closs SJ, Long AF, Corbett A, et al. Pain assessment for people with dementia: A systematic review of systematic reviews of pain assessment tools. *BMC Geriatr*. 2014 12/17;14:138.
36. Ritchie CW, Terrera GM, Quinn TJ. Dementia trials and dementia tribulations: methodological and analytical challenges in dementia research. *Alz Res Therapy*. 2015; 7:31
37. Pendlebury ST, Chen PJ, Bull L, Silver L, Mehta Z, Rothwell PM. Methodological factors in determining rates of dementia in TIA and stroke: (I) impact of baseline selection bias. *Stroke*. 2015; 46:641-646.
38. RA Lees, K Hendry BA, N Broomfield, D Stott, AJ Larner, TJ Quinn. Cognitive assessment in stroke: feasibility and test properties using differing approaches to scoring of incomplete items. *International journal of geriatric psychiatry*. 2017; 32:1072-1078.