BMJ Open

Validity and reliability of the Modified Tardieu Scale as a spasticity outcome measure of the upper limbs in adults with neurological conditions: a systematic review and narrative analysis

Xiaoyi Shu [1], Ciara McConaghy,[2] Alec Knight [3]

To cite: Shu X, McConaghy C, Knight A. Validity and reliability of the Modified Tardieu Scale as a spasticity outcome measure of the upper limbs in adults with neurological conditions: a systematic review and narrative analysis. *BMJ Open* 2021;**11**:e050711. doi:10.1136/bmjopen-2021-050711

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-050711).

Received 26 February 2021
Accepted 03 December 2021

[1]School of Kinesiology, Shanghai University of Sport, Shanghai, China
[2]Department of Physiotherapy, Guy's and St Thomas' NHS Foundation Trust, London, UK
[3]Department of Primary Care and Public Health Sciences, King's College London, London, UK

Correspondence to
Xiaoyi Shu;
shuxiaoyi@sus.edu.cn

ABSTRACT

Purpose To evaluate published evidence on the Modified Tardieu Scale (MTS) as a tool to assess spasticity in the upper limbs of adults with neurological conditions.

Data sources A systematic search of six electronic databases (PubMed/MEDLINE, CINAHL, EMBASE, the Cochrane Library, Web of Science and Physiotherapy Evidence Database) from inception to 31 December 2020. A search strategy was developed using key elements of the research question: population, intervention (action), outcome.

Study eligibility criteria Inclusion criteria: (1) adult participants with neurological conditions; (2) upper limb muscles/joints as tested elements; (3) studies testing the MTS and (4) reliability or validity reported. Exclusion criteria: (1) non-English articles; (2) non-empirical articles and (3) studies testing the Tardieu Scale.

Study appraisal Evidence quality was evaluated using the US National Heart, Lung, Blood Institute quality assessment tool for observational cohort and cross-sectional studies.

Results Six reliability studies met the inclusion criteria. Overall, most articles reported good-to-excellent levels of inter-rater, intrarater and test–retest reliability. However, limitations, such as study design weaknesses, statistical misuses and reporting biases, undermine confidence in reported conclusions. The validity of the MTS also remained questionable based on the results of one study.

Conclusions and implications This review did not find sufficient evidence to either support or reject the use of the MTS in assessing spasticity in the upper limbs of adults with neurological conditions. Despite the paucity of research evidence, the MTS may still remain a clinically useful tool to measure the motor aspect of spasticity. Future research would benefit from a focus on test standardisation, while the wider field would require the development of a consensual definition of spasticity.

Strengths and limitations of this study

- ► To our knowledge, this is the first systematic review of the Modified Tardieu Scale (MTS) for adults with neurological conditions, examining the features and quality of the included papers.
- ► As this review only includes English-language studies and the MTS is used internationally, it is possible that some non-English language publications including relevant data may not have been among our results.
- ► Conclusions are limited by a small number of identified papers, most of which selected elbow flexor muscles to test.

INTRODUCTION

An upper motor neuron (UMN) syndrome refers to a cluster of broad symptoms including both positive and negative features, with spasticity being one of the most common positive features.[1] The traditional definition of spasticity describes it as a motor disorder with a velocity-dependent increase in tonic stretch reflexes, with exaggerated tendon jerks that result from a loss of inhibition of the stretch reflex.[2] A high prevalence of spasticity has been found in a number of neurological conditions, such as stroke, spinal cord injury, traumatic brain injury and multiple sclerosis.[3–5] Patients with spasticity have poorer quality of life and greater burden of care compared with those without it.[6] Such effects are particularly true when spasticity is found in the upper limbs, leading to profound disabilities, low levels of independence and high cost of care.[7 8] Therefore, an accurate evaluation of upper limb spasticity is crucial in spasticity management potentially leading to better outcomes and care cost saving. The Ashworth Scale (AS) and Modified Ashworth Scale (MAS) are the most commonly used clinical tools to assess spasticity.[9] However, the validity of the AS and MAS has been questioned since they assess the resistance to the passive movement at a constant stretch velocity instead of the velocity-dependent feature of spasticity.[10]

In addition to the AS and MAS, the Tardieu Scale (TS) is another commonly used tool for spasticity assessment.[11] Evidence suggests that the TS is significantly better at identifying spasticity than the AS.[12] The main strength of the TS is to identify the presence and severity of spasticity by objectively measuring the muscle responses to passive stretches at different speeds.[12] This characteristic allows the TS to reflect the velocity-dependent feature of spasticity better than the AS and MAS.[11] The TS was first described in French,[13] and later developed and translated into English by Held and Pierrot-Deseilligny.[14] Boyd and Graham[15] standardised the testing speed and procedures, and this version was later known as the Modified Tardieu Scale (MTS). The MTS comprises four elements, including R1, R2, R2-R1 and X score. R1 represents an angle of catch, where a sudden increase of muscle resistance is felt during a fast passive stretch. R2 is an angle indicating the tested muscle length at a slow passive range of motion. R2-R1 was introduced to differentiate spasticity from contracture.[15] Contracture is indicated if the value of R2-R1 is small, while a large value indicates spasticity.[15] The X score describes types of muscle resistance when passive stretches are applied during the assessment (grade 0–5).

A systematic review was published in 2006 on the validity and reliability of the TS.[16] However, the research evidence underpinning it was limited by methodological weaknesses. These included an unclear search date range, a lack of validity studies and samples comprising mostly paediatric participants. Despite two decades passing since the development of the MTS and its common use clinically,[17] relatively little published research has examined its features. Moreover, evidence on evaluation of its psychometric properties is scarce. Therefore, this systematic review is warranted to explore existing evidence on the validity and reliability of the MTS to assess spasticity in the upper limbs of adults with neurological conditions. On the basis of the identified research, we will make conclusions on the quality of evidence to date and recommendations for future research by identifying gaps in evidence surrounding its use.

METHOD

Design

Systematic review and narrative synthesis.

Information sources

A systematic literature search was conducted using PubMed/MEDLINE, CINAHL PLUS, EMBASE, the Cochrane Library, Web of Science and Physiotherapy Evidence Database from inception to 31 December 2020. After all the duplicates were removed, a hand search of the reference lists of the remaining articles was conducted to identify additional relevant studies.

Search strategy and study selection

A search strategy was developed in PubMed/MEDLINE using key elements of the research question: population, intervention (action) and outcome. It was later adapted for the other databases searched. Part 1 (population) included common neurological conditions and the term 'disability' as well as variations of these keywords. Part 2 (intervention/action) included the search term 'Modified Tardieu Scale' within text words. The last part (outcome) focused on spasticity and its variations. Medical subject headings were included in searches where possible. Three parts were combined using Boolean operator AND (see online supplemental appendix 1). The search procedure followed the Preferred Reporting Items for Systematic Review and Meta-Analysis reporting guidelines.[18] XS and AK consulted a librarian and designed the search strategy together, while XS and CM conducted the search independently.

Eligibility criteria

The inclusion criteria were: (1) adult participants with neurological conditions; (2) upper limb muscles/joints as tested elements; (3) studies testing the MTS and (4) reliability or validity reported. The exclusion criteria were: (1) non-English articles; (2) non-empirical studies and (3) studies testing the TS.

Data extraction, synthesis and analysis

Following completion of the search, collection of articles and application of the inclusion and exclusion criteria, data on reliability and validity were extracted from the articles. These data were recorded by XS and CM independently and the results were verified. The results were summarised under the following headings: (1) Authors (year); (2) Study design; (3) Participants (n=, health conditions, source of recruitment); (4) Test characteristics (Tested muscle, testing position, testing time and interval, testing speed, measuring, rater training, MTS version, any other relevant details) and (5) Summary of results (inter-rater, intrarater and test–retest reliability; validity).

Quality assessment

The internal validity of individual studies was assessed using the US National Heart, Lung, Blood Institute (USNHLBI) quality assessment tool for observational cohort and cross-sectional studies[19] (online supplemental appendix 2). According to recommendations from Sommer *et al*,[20] studies were rated as poor, fair or good with scores of less than 50%, 50%–75% and greater than 75%, respectively.

Patient and public involvement

Patient and public involvement was a part of this study from conception to completion, with author AK being a community physiotherapy service user for spasticity management. The MTS was used by his community physiotherapist, XS, to assess his spasticity. XS and AK discussed the validity and reliability of the MTS during a physiotherapy session. XS and AK commenced a collaboration to undertake this research paper, with an aim

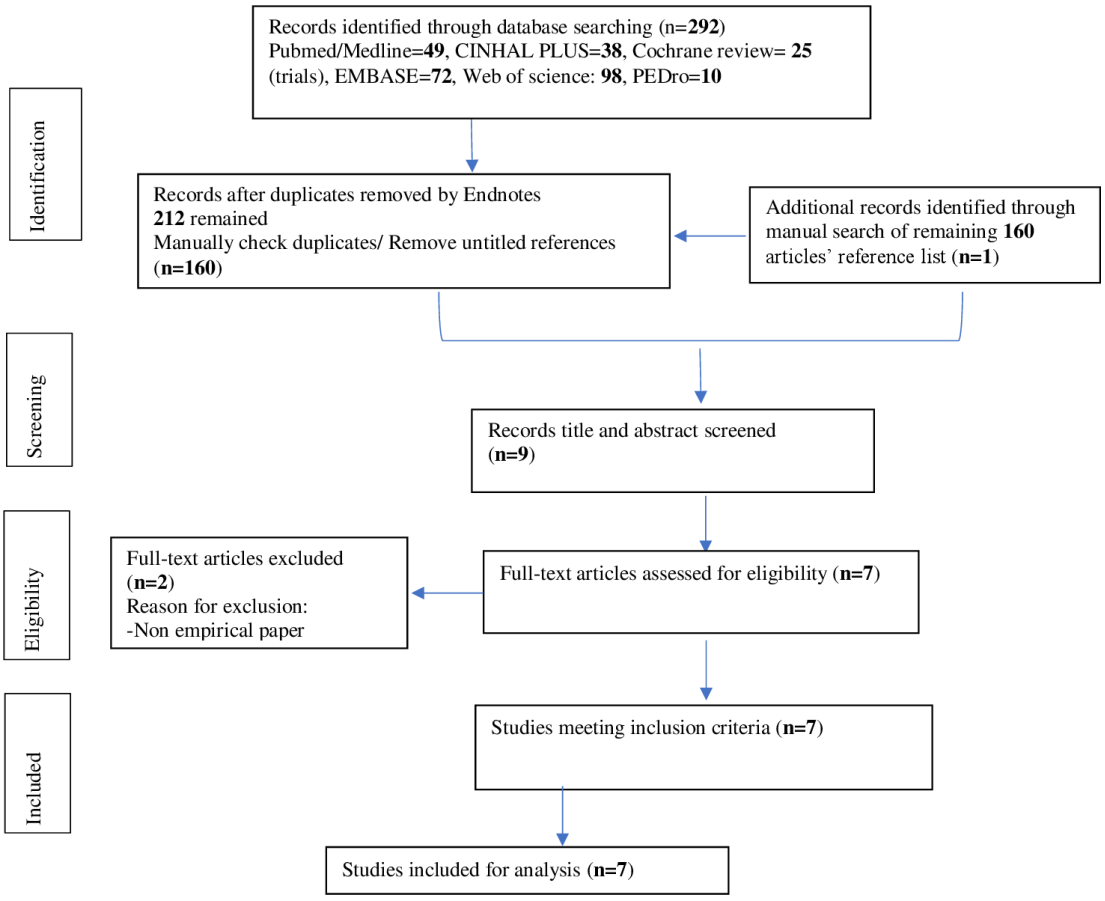

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart of the study identification and inclusion process.

to contribute to the spasticity research literature and to inform practice for clinicians.

RESULTS

Overview

A total of 292 studies were retrieved from the search of six databases. After the duplicates were removed by a software (Endnote) and researchers manually, the remaining 160 articles were retained for screening. One additional article was identified through reference list searching. Nine papers were retrieved for eligibility assessment after title and abstract screening. Finally, seven articles were included for this review after two non-empirical papers were removed. The details of the search process are displayed in figure 1.

Characteristics of included studies

Seven observational studies met all inclusion criteria. Descriptive data of the studies are listed in table 1. Six articles reported data on reliability: four of these examined inter-rater reliability,[21–24] two studied intrarater reliability[24 25] and three investigated test–retest reliability.[21 23 26] Only one article reported data on criterion validity.[27] The total number of participants across seven studies was 317, with sample sizes between 20 and 91. Most participants were stroke survivors (n=222). The studies described the test characteristics at varying levels of detail. For instance, raters' training was described in three studies.[21 23 25] Two studies specified testing speed and marked the tested joints.[25 26] Elbow flexors were the most commonly tested muscles across the studies.[21–26] The statistical methods were also described in all papers except Sonvane and Kumar.[26] The most frequently used statistical methods were intraclass correlation coefficients (ICC) (six studies) and t-tests (seven studies).

Summary of the study results

Table 1 displays reliability estimates of the MTS in the reviewed studies: ICC values range from 0.56 to 0.851 for inter-rater reliability,[21–24] from 0.70 to 0.998 for intrarater reliability[24 25] and from 0.627 to 0.815 for test–retest reliability.[21 23] Most studies listed data of all four items of the MTS (R1, R2, R2-R1 and X scores).[22 24–26] Nevertheless, Mehrholz *et al*[21] and Waninge *et al*[23] only reported ICC values of two items in their papers. X scores, a categorical item, were measured using ICC by Ansari *et al*[22] and Singh *et al*,[25] while Mehrholz *et al*[21] and Li *et al*[24] chose Kappa to assess X scores and reported a wide range between 0.33 and 0.87. Instead of ICC or Kappa, Sonvane and Kumar[26] selected Pearson's correlation coefficients and t-tests to assess the overall reliability of the MTS. They reported satisfactory correlation values (r=0.913–0.973)

Table 1 Data extraction table for MTS studies reporting reliability and validity data

Reliability studies

Author	Study design	Participants	Test characteristics	Summary of results		
				Inter-rater reliability	Intrarater reliability	Test–retest reliability
Sonvane and Kumar[26]	Cross-sectional study	▶ n=60 ▶ Stroke ▶ Source of participants not specified	▶ Tested muscle: elbow flexors ▶ Testing position: sat on a chair with shoulder in adduction for elbow flexors ▶ Testing time and interval: unspecified time with a 2-day interval ▶ Testing speed: counting 1001, 1002, 1003… (V1), count 1, 2, 3… (V3) ▶ Measurement instrument: goniometer on specific body landmark ▶ Rater training: not specified ▶ MTS version: not specified	–	–	R1: r=0.973, p<0.001 R2: r=0.913, p<0.001 R2-R1: r=0.924, p<0.001 X Scores: r=0.937, p<0.001
Singh *et al*[25]	Cross-sectional study	▶ n=91 ▶ Stroke ▶ Hospital admissions	▶ Tested muscle: elbow flexors ▶ Testing position: sat on a chair with shoulder in adduction for elbow flexors ▶ Testing time and interval: a 2-day intervals ▶ Testing speed: counting 1001, 1002, 1003… (V1) count 1, 2, 3… (V3) ▶ Measurement instrument: goniometer with specific body landmarks ▶ Rater training: trained by two experienced neurophysiotherapists, but no further details ▶ MTS version: 5-point scale	–	R1: ICC: 0.998; CI: 0.997 to 0.999 R2: ICC: 0.978; CI: 0.966 to 0.986 R2-R1: ICC: 0.991; CI: 0.986 to 0.994 X Scores: ICC: 0.847; 0.769–0.899	–
Li *et al*[24]	Cross-sectional study	▶ n=51 ▶ Stroke ▶ Consecutive admissions to a hospital ward	▶ Tested muscle: elbow flexors ▶ Testing position: supine with arm by the side and head in neutral position ▶ Testing time and interval: 7:00 and 8:30 with a 1-day interval ▶ Testing speed: as slow as possible (V1), as fast as possible (V3) ▶ Measurement instrument: not specified ▶ Rater training: not specified ▶ MTS version: 6-point scale	R1:ICC:0.78; CI: 0.64 to 0.87 R2: ICC:0.58; CI: 0.36 to 0.73 R2-R2: ICC: 0.76; CI: 0.48 to 0.80 X Scores: Kappa: 0.73; SE: 0.08; p-value<0.001	R1: ICC: 0.71; CI: 0.53 to 0.82 R2: ICC: 0.83; CI: 0.72 to 0.90 R2-R1: ICC: 0.70; CI: 0.53 to 0.82 X Scores: Kappa: 0.73; SE: 0.08; p-value<0.001	–
Waninge *et al*[23]	Cross-sectional study	▶ n=35 ▶ Profound intellectual and multiple disabilities ▶ Source of participants not specified	▶ Tested muscle: elbow (muscle group not specified) ▶ Testing position: unknown protocol specified in a Dutch language article ▶ Testing time and interval: time not specified with a 1-week interval ▶ Testing speed: slow motion within 1 s (V1), fast motion within half second (V3) ▶ Measurement instrument: goniometer ▶ Rater training: trained in using protocol but not the scale ▶ MTS version: not specified	R1: p level: 0.886; ICC: 0.806; LOA: 40; SCC: 0.813 R2: p level: 0.540; ICC: 0.851; LOA: 38; SCC: 0.825	–	R1: p level: 0.592; ICC: 0.815; LOA: 35.2; SCC: 0.792 R2: p level: 0.890; ICC: 0.627; LOA: 57.5; SCC: 0.624

Continued

Table 1 Continued

Reliability studies

Author	Study design	Participants	Test characteristics	Summary of results: Inter-rater reliability	Summary of results: Intrarater reliability	Summary of results: Test–retest reliability
Ansari *et al*[22]	Cross-sectional study	▶ n=30 ▶ Brain injury leading to hemiplegia ▶ Participants attended a rehabilitation clinic	▶ Tested muscle: elbow flexors ▶ Testing position: sat on a chair with shoulder in adduction ▶ Testing time and interval: not specified ▶ Testing speed: not specified ▶ Measurement instrument: goniometer ▶ Rater training: no formal training ▶ MTS version: 5-point scale ▶ Others: raters were blinded to the results	R1: ICC: 0.74; CI: 0.52 to 0.87 R2: ICC: 0.56; CI: 0.26 to 0.76 R2-R1: ICC: 0.72; CI: 0.50 to 0.86 X Scores: ICC: 0.74; CI: 0.53 to 0.87	–	–
Mehrholz *et al*[21]	Cross-sectional study	▶ n=30 ▶ Severe brain injury patients ▶ Attending rehabilitation department	▶ Tested muscle: Shoulder flexor and external rotator, elbow flexor and extensors, wrist flexors and extensors ▶ Testing position: supine with the arm by the body, elbow in extension, wrist in a neutral position. Then the elbow was extended from maximal flexion position ▶ Testing time and interval: 9:00–10:00 with a 10 minutes interval ▶ Testing speed: as slow as possible (V1), as fast as possible (V3) ▶ Measurement instrument: goniometer ▶ Rater training: 45 minutes training session ▶ MTS version: 6-point scale ▶ Others: raters were blinded to the results	R1 (elbow flexor): ICC: 0.72 X Scores: mean Kappa: 0.33–0.51; SE: 0.03–0.07; p value<0.05 (wrist flexion p value not significant)	–	R1 (elbow flexor): ICC: 0.73 X Scores: Kappa: 0.53–0.87; SE: 0.02–0.05, p value<0.001 (shoulder flexion and shoulder external rotation p values: not significant)

Validity study

Author	Study size	Participants	Test characteristics	Summary of results
Naghdi *et al*[27]	Cross-sectional study	▶ n=20 ▶ Stroke ▶ Consecutive admissions to local clinics	▶ Tested muscle: wrist flexors ▶ Testing position: elbow 90° in flexion ▶ Testing time and interval: 5 minutes interval between R1 and R2 ▶ Testing speed: as slow as possible (V1), as fast as possible (V3) ▶ Measurement instrument: goniometer with specific body landmark ▶ Rater training: not specified ▶ MTS version: 5-point scale	Difference between group A (MTS=0) and group B (MTS=2) p=0.008 (Hslp/Mslp) p=0.08 (Hmax/Mmax) p=0.53 (Hslp) p=0.22 (latency)

CI, 95% Confidence Interval; Hmax/Mmax, maximum mean amplitude of the H-reflex relative to maximum mean amplitude of the M-wave; Hslp, the developmental slope of the H-reflex; Hslp/Mslp, the developmental slope of the H-reflex recruitment curve relative to the developmental slope of the M-response; ICC, intraclass correlation coefficient; LOA, limits of agreement; MAS, Modified Ashworth Scale; MTS, Modified Tardieu Scale; r, Pearson's correlation coefficient; SCC, Spearman correlation coefficient.

and statistically significant p values (p<0.001) of all four elements.[26] No articles were found in the systematic search discussing the content or construct validity of the MTS, which was a notable gap in the research literature. One article investigated the criterion validity of the MTS.[27] Naghdi *et al*[27] reported poor correlations between the MTS and electrophysiological measurements of the H-reflex in all four elements when testing the wrist flexors.

Quality of evidence

We modified the USNHLBI quality assessment tool[19] because five items of the tool were not applicable to

the reviewed studies (items 6, 7, 8, 10 and 13). Accordingly, the quality of the reviewed studies was classified as poor for total scores between 1 and 4, fair for scores between 5 and 7 and good for scores above 7. The quality of most studies included in the review was poor[22–24 26 27] with studies by Mehrholz *et al*[21] and Singh *et al*[25] studies being rated good and fair, respectively, (table 2). All the studies clearly defined the purpose of the research (item 1). None of the studies justified the sample size (item 5), with variable details of participant criteria and selection (items 2, 3 and 4). Only Singh *et al*,[25] Mehrholz *et al*[21] and Naghdi *et al*[27] correctly defined the assessment of the reliability or validity of the MTS (item 11), while assessors were blinded in three studies (item 12).[21 22 25] Two studies considered confounders in statistical analysis (item 14).[21 23]

DISCUSSION

There is insufficient evidence to either support or reject the use of the MTS to assess the spasticity in the upper limbs of adults with neurological conditions. Although the selected papers showed positive results regarding the reliability of the MTS,[21–26] further analysis revealed methodological weaknesses and low study quality affected the credibility of the results. The only identified paper assessing the validity of the MTS contained study design limitations.[27] This discussion summarises the main findings and critically appraises them. Then, possible explanations and implications for clinicians and policy-makers are explored. The strengths and limitations of this review are discussed, followed by recommendations for future research.

Summary of findings

Koo and Li[28] suggested an ICC value above 0.5 indicated moderate-to-excellent reliability. Across all the papers reviewed, even though all ICC values were found above 0.5, four studies reported moderate-to-excellent levels of inter-rater, intrarater and test–retest reliabilities.[21 24–26] This is due to a lack of consensus on acceptable levels of agreement among these studies. For instance, in spite of high ICC values for most tested items, Waninge *et al*[23] reported insufficient inter-rater and test–retest reliabilities because the results of the reliability tests failed to meet their own statistical criteria (acceptable ICC>0.75). Similarly, using different assessment criteria, Ansari *et al*[22] concluded that the inter-rater reliability was in fact unacceptable due to great variabilities of the R2-R1 values and the X scores between raters.

Evidence on the validity of the MTS is lacking, and further research in this area is required. The only validity study returned by our search suggested low confidence in using the MTS clinically as a validated tool.[27] However, the authors acknowledged there was no gold standard to assess spasticity against the MTS,[27] thereby the internal validity of this study is questionable.

Furthermore, the quality of study assessment suggests five out of seven reviewed studies were poor quality.[22–24 26 27] The only study rated as good quality,[21] testing several upper limb muscle groups, but only R1 and X scores data for the elbow flexors were presented in the paper. Therefore, the available data may not be sufficient to lend credibility to the findings. In summary, the seven studies included in this review do not provide adequate evidence to support the use of the MTS to assess spasticity in the upper limbs of adults with neurological conditions. However, this may be due to the low overall quality of the reviewed studies.

Critical discussion of review findings

Closer examination of the seven studies shows all studies contained methodological limitations, such as non-standardised use of the MTS and small samples without power calculations.[21–27] The largest sample was 91 participants,[25] while 3 studies had sample sizes below 35.[21–23] Small sample sizes are known to cause a wide 95% Confidenc Interval (CI) range.[29] For instance, the 95% CI values of R2 reported by Ansari *et al*[22] and Li *et al*[24] were 0.26 to 0.76 and 0.36 to 0.73, respectively, making it difficult to have confidence in the reported means. Moreover, none of the studies acknowledged the fact that there were two versions of X scores. Boyd and Graham[15] added an extra item on the X score category of the TS, extending it from five to six scores. Mehrholz *et al*[21] used categories 0–5, while Ansari *et al*[22] and Singh *et al*[25] selected the TS version (0–4). Neither Waninge *et al*[23] nor Sonvane and Kumar[26] specified which scale was investigated. Li *et al*[24] claimed to study a 5-score scale, but their results suggested they used a 6-score scale instead. Non-standardised use of the scale made it hard to compare the X score results across the studies.

Misuses of statistical methods can be found in the reviewed papers, leading to misinterpretation of results and incorrect conclusions being drawn. For example, ICC and 95% CI are parametric statistical procedures that should only be used on continuous data.[30] However, X scores (discrete data) were analysed using ICC and 95% CI in two studies.[22 25] Similarly, Sonvane and Kumar[26] analysed X scores choosing Pearson's correlation coefficient, a measure of the association between two continuous variables.[31] Furthermore, reporting biases were noted in some studies. For instance, Waninge *et al*[23] only listed satisfactory ICC values of R1 and R2, but measurements all four elements were described in the method. The positive conclusion may have been overstated if based on the results of only two elements. The results of the X score and R2-R1 could affect the overall estimates of the MTS. In summary, despite reported moderate-to-excellent reliability, close examination of the research studies cited in this review suggests that many factors affect the credibility of the research findings, including reporting biases and methodological weaknesses.

Implications

The present results and analysis lead to some practical implications for clinicians, researchers and policy-makers.

Table 2 Quality assessment tool for observational cohort and cross-sectional studies

Criteria	Study authors						
	Sonvane and Kumar[26]	Singh *et al*[25]	Li *et al* [24]	Waninge *et al*[23]	Ansari *et al*[22]	Mehrholz *et al*[21]	Naghdi *et al*[27]
1. Was the research question or objective in this paper clearly stated?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Was the study population clearly specified and defined?	No	Yes	No	No	No	Yes	No
3. Was the participation rate of eligible persons at least 50%	NR	Yes	NR	Yes	NR	Yes	NR
4. Were all the subjects selected or recruited from the same or similar populations (including the same time)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?	CD	Yes	Yes	Yes	CD	Yes	CD
5. Was a sample size justification, power description, or variance and effect estimates provided?	No	No	No	No	No	No	No
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	NA	NA	NA	NA	NA	NA	NA
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	NA	NA	NA	NA	NA	NA	NA
8. For exposures that can vary in amount of level, did the study examine different levels of the exposure and outcome?	NA	NA	NA	NA	NA	NA	NA
9. Were the exposure measures (independent variables) clearly defined, valid, reliable and implemented across all study participants?	Yes	Yes	Yes	No	Yes	Yes	Yes
10. Was the exposure(s) assessed more than once over time?	NA	NA	NA	NA	NA	NA	NA
11. Were the outcome measure (dependent variables) clearly defined, valid, reliable and implemented across all study participants?	No	No	Yes	No	No	Yes	Yes
12. Were the outcome assessors blinded to the exposure status of participants?	NR	Yes	NR	NR	Yes	Yes	NR
13. Was loss to follow-up after baseline 20% or less?	NA	NA	NA	NA	NA	NA	NA
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	No	No	No	Yes	No	Yes	No
Quality rating	Poor	Fair	Poor	Poor	Poor	Good	Poor

Quality rating categories adapted from Sommer *et al*[20] : 'poor: ≤50%; fair: 50%–75%; good≥75%'.
CD, cannot determine; NR, not reported.

Reviewing the procedures employed across the studies suggests that further standardisation of the MTS may be warranted. For instance, sitting, supine and prone MTS testing positions have been reported.[32] Researchers found significantly lower tone in the upper limb muscles in the supine position than in sitting or standing positions.[33] In addition, testing speed was not standardised during the tool development. Mackey *et al*[34] found 'considerably variable' speed between participants when the biceps muscles were assessed. Moreover, the fact that there are two versions of the X score scale may cause confusion, resulting in miscommunication between researchers and clinical users.

The validity of the MTS remains unclear, partially due to a lack of consensus on a spasticity definition and a full understanding of its pathophysiology. A systematic review of 250 papers revealed inconsistency in defining spasticity.[1] Only about 30% of reviewed studies explained spasticity according to Lance's definition, a third simply described it as increased muscle tone, and the rest used their own definitions or none at all. Lacking an agreement on spasticity definition has made it difficult to decide on the most valid measurement. Furthermore, the classic definition of spasticity[2] is believed only to reflect its motor aspects.[35] The TS and MTS adhere to this definition by addressing the velocity-dependent feature of spasticity.[11] A more recent definition describes spasticity as a disorder of sensorimotor control, resulting from an UMN lesion, presenting as intermittent or sustained involuntary muscle activation.[35] This broad description includes other positive signs of an UMN syndrome, such as clonus and spasm, which are often used interchangeably with spasticity.[36] Moreover, an abnormal sensory process was also acknowledged as a factor to induce spasticity.[37] Nevertheless, no further recommendations have been made either to measure the sensory aspect of spasticity or to differentiate spasticity from other positive signs of an UMN syndrome. The MAS, which measures spasticity as resistance to passive movements,[38] may be more likely to address the 'intermittent or sustained involuntary muscle activation' feature described by Pandyan *et al*.[35] Therefore, to support research on the reliability and validity of the MTS, we recommend further work to establish a consensual definition of spasticity.

Despite all the pitfalls discussed above, the MTS may still be clinically useful to assess the velocity-dependent feature of the spasticity based on the current knowledge of spasticity pathophysiology and previous studies of the TS. First, losing central inhibition following UMN lesions, spasticity is caused by alpha motor neuron hyperexcitability leading to an exaggerated stretch reflex.[37] Power *et al*[39] reported that by applying the same stretch velocity, the stretch response threshold reduced and the amplitude was increased in people with spasticity compared with unaffected subjects. Therefore, the neuropathophysiology supports that using the MTS by performing a fast speed stretch could provoke spasticity as a result of an overactive stretch reflex.

Furthermore, the content validity of the TS was investigated by comparing the manual identification of spasticity with laboratory measurements.[40] The TS successfully detected 88.9% of spasticity ($p<0.05$). Patrick and Ada[12] reported 100% exact agreement between the TS and laboratory measurements. In both studies,[12 40] spasticity was identified by performing stretches at a fast speed. The MTS assesses spasticity in the same way, meaning the MTS may also have excellent content validity in assessing the motor features of the spasticity. So, the MTS could still be considered a useful clinical tool to assess the presence of spasticity.

Other strengths of the MTS should not be underestimated. For example, the costs associated with conducting the MTS are very low, including just a goniometer and a standard plinth. The MTS instructions and assessment forms are available online free of charge, and training to use the tool can be completed in as little as 45 min.[21 22] Moreover, rater experience does not seem to have a significant impact on the results. Studies recruiting experienced physiotherapists[21 25] and students[22] all reported consistent results. Feasibility of the MTS was also assessed by comparing the successful measurements to total measurement numbers.[23] The MTS has been reported to have good feasibility, with a 94% success rate.[23]

Strengths and limitations

To our knowledge, this study is the first to investigate the validity and reliability of the MTS to assess upper limb spasticity for the adults with neurological conditions. Moreover, a comprehensive search of multiple databases and hand searches of references ensured studies selected for the review were up to date. Furthermore, an in-depth analysis of results and characteristics has allowed us to understand the state of current evidence on the MTS. We acknowledge that our search was biased toward English publications and may have missed relevant studies written in other languages. Furthermore, this search only identified seven studies in total, six of which tested the elbow flexors. Therefore, the reported findings may not be applicable to other untested upper limb muscle groups.

The only similar study to ours in the literature was a systematic review of the TS conducted in 2006.[16] Of the research reviewed, most studies recruited children as subjects. Our review has added evidence of the MTS in adult population. Although neither Haugh *et al*[16] nor the present review found sufficient evidence to support the validity and reliability of the TS and MTS. A direct comparison of the two papers is problematic due to the heterogenous samples.

Recommendations

Based on a critical review of the study results, we suggest that future research studying psychometric properties of the MTS should focus on improving study designs, including test standardisation, statistical power and sample size calculation, and using appropriate statistical analysis. Second, the internal validity of the MTS study

could be compromised by an inconsistent definition of spasticity.[1] Therefore, further efforts could be made to develop the definition of spasticity and subsequently develop the MTS congruent to that definition. Finally, according to the International Classification of Functioning, Disability and Health (ICF) framework,[41] the MTS only reflects the motor aspect of the spasticity on the body functions and structures level. Multidimensional assessments of disabilities, including those on the activity and participation levels, were recommended to form a holistic understanding of the impacts' spasticity could make on individuals' lives.[1 17] For instance, the Action Research Arm Test, and the Nottingham Extended ADL can be applied as measurements of functions (activity level) and quality of life (participation level), respectively.[17 42]

CONCLUSIONS

Research suggests good-to-excellent reliability of the MTS assessing spasticity in the upper limbs in adults with neurological conditions. However, the results should be interpreted with caution due to study design limitations and inconsistencies, reporting bias and some inappropriate use of statistical analyses. To date, limited evidence supports the validity of the MTS. Further research is required to develop a consensual definition of spasticity and to validate the MTS in the future. In spite of the limitations, the MTS may still be considered for assessing upper limb spasticity due to its low cost, short training time and good reflection of the velocity-dependent positive feature of n UMN syndrome. When selecting the MTS for clinical and research purposes, clinicians should ensure test procedures are standardised to achieve reliable and consistent results. Furthermore, it is worth acknowledging that the MTS only assesses an individual's disability on the body functions and structures level based on the ICF framework. International guidelines recommend additional tests on the activity and participation levels to gain better understandings of abilities and disabilities.[17 42]

Contributors XS and AK conceptualised the study, planned the methods, conducted search, data extraction and analyses, quality assessment, write the paper and approval of the final manuscript. AK participated in method planning, conducting search, data analysis and paper writing. CM was involved in data search, verification and approval of the final manuscript. XS is responsible for the content as the guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study does not involve human participants.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. All data in this study were obtained in published papers. All generated and analysed data were included in this article.

ORCID iDs
Xiaoyi Shu http://orcid.org/0000-0001-7141-1382
Alec Knight http://orcid.org/0000-0002-2937-436X

REFERENCES

1 Malhotra S, Pandyan AD, Day CR, *et al*. Spasticity, an impairment that is poorly defined and poorly measured. *Clin Rehabil* 2009;23:651–8.
2 Lance JW. The control of muscle tone, reflexes, and movement: Robert Wartenberg lecture. *Neurology* 1980;30:1303–13.
3 Wissel J. Towards flexible and tailored botulinum neurotoxin dosing regimens for focal dystonia and spasticity - Insights from recent studies. *Toxicon* 2018;147:100–6.
4 Angulo-Parker FJ, Adkinson JM. Common etiologies of upper extremity spasticity. *Hand Clin* 2018;34:437–43.
5 Rizzo MA, Hadjimichael OC, Preiningerova J, *et al*. Prevalence and treatment of spasticity reported by multiple sclerosis patients. *Mult Scler* 2004;10:589–95.
6 Urban PP, Wolf T, Uebele M, *et al*. Occurence and clinical predictors of spasticity after ischemic stroke. *Stroke* 2010;41:2016–20.
7 Opheim A, Danielsson A, Alt Murphy M, *et al*. Early prediction of long-term upper limb spasticity after stroke: part of the SALGOT study. *Neurology* 2015;85:873–80.
8 Lundström E, Smits A, Borg J, *et al*. Four-fold increase in direct costs of stroke survivors with spasticity compared with stroke survivors without spasticity: the first year after the event. *Stroke* 2010;41:319–24.
9 Marciniak C. Poststroke hypertonicity: upper limb assessment and treatment. *Top Stroke Rehabil* 2011;18:179–94.
10 Pandyan AD, Johnson GR, Price CI, *et al*. A review of the properties and limitations of the Ashworth and modified Ashworth scales as measures of spasticity. *Clin Rehabil* 1999;13:373–83.
11 Cha Y, Arami A. Quantitative modeling of spasticity for clinical assessment, treatment and rehabilitation. *Sensors* 2020;20:5046.
12 Patrick E, Ada L. The Tardieu scale differentiates contracture from spasticity whereas the Ashworth scale is confounded by it. *Clin Rehabil* 2006;20:173–82.
13 Tardieu G, Shentoub S, Delarue R. [Research on a technic for measurement of spasticity]. *Rev Neurol* 1954;91:143–4.
14 Held J-P, Pierrot-Deseilligny E. *Reeducation mortrice des affections neurologiques*. Paris: J.B. Baillière et fils, 1969.
15 Boyd RN, Graham HK. Objective measurement of clinical findings in the use of botulinum toxin type A for the management of children with cerebral palsy. *Eur J Neurol* 1999;6:s23–35.
16 Haugh AB, Pandyan AD, Johnson GR. A systematic review of the Tardieu scale for the measurement of spasticity. *Disabil Rehabil* 2006;28:899–907.
17 Sheean G, Lannin NA, Turner-Stokes L, *et al*. Botulinum toxin assessment, intervention and after-care for upper limb hypertonicity in adults: international consensus statement. *Eur J Neurol* 2010;17:74–93.
18 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
19 National Institutes of Health. National Heart Lung, and Blood Institute Study quality assessment tools [Internet], 2013. Available: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools
20 Sommer AE, Golden BP, Peterson J, *et al*. Hospitalized patients' knowledge of care: a systematic review. *J Gen Intern Med* 2018;33:2210–29.

21 Mehrholz J, Wagner K, Meissner D, *et al*. Reliability of the modified Tardieu scale and the modified Ashworth scale in adult patients with severe brain injury: a comparison study. *Clin Rehabil* 2005;19:751–9.
22 Ansari NN, Naghdi S, Hasson S, *et al*. The modified Tardieu scale for the measurement of elbow flexor spasticity in adult patients with hemiplegia. *Brain Inj* 2008;22:1007–12.
23 Waninge A, Rook RA, Dijkhuizen A, *et al*. Feasibility, test-retest reliability, and interrater reliability of the modified Ashworth scale and modified Tardieu scale in persons with profound intellectual and multiple disabilities. *Res Dev Disabil* 2011;32:613–20.
24 Li F, Wu Y, Li X. Test-retest reliability and inter-rater reliability of the modified Tardieu scale and the modified Ashworth scale in hemiplegic patients with stroke. *Eur J Phys Rehabil Med* 2014;50:9–15.
25 Singh P, Joshua AM, Ganeshan S, *et al*. Intra-rater reliability of the modified Tardieu scale to quantify spasticity in elbow flexors and ankle plantar flexors in adult stroke subjects. *Ann Indian Acad Neurol* 2011;14:23–6.
26 Sonvane I, Kumar TS. Test retest reliability of modified Tardieu scale to quantify the spasticity in elbow flexors in patients with Cerebro vascular accident. *Ind Jour of Physioth and Occupat Therapy - An Inter Jour* 2019;13:38–42.
27 Naghdi S, Ansari NN, Abolhasani H, *et al*. Electrophysiological evaluation of the modified Tardieu scale (mts) in assessing poststroke wrist flexor spasticity. *NeuroRehabilitation* 2014;34:177–84.
28 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–63.
29 Slutsky DJ. Statistical errors in clinical studies. *J Wrist Surg* 2013;2:285–7.
30 Kuzon WM, Urbanchek MG, McCabe S. The seven deadly SINS of statistical analysis. *Ann Plast Surg* 1996;37:265–72.
31 Liu J, Tang W, Chen G, *et al*. Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Arch Psychiatry* 2016;28:115–20.
32 Scholtes VAB, Becher JG, Beelen A, *et al*. Clinical assessment of spasticity in children with cerebral palsy: a critical review of available instruments. *Dev Med Child Neurol* 2006;48:64–73.
33 Qin W, Yang M, Li F, *et al*. Influence of positional changes on spasticity of the upper extremity in poststroke hemiplegic patients. *Neurosci Lett* 2019;712:134479.
34 Mackey AH, Walt SE, Lobb G, *et al*. Intraobserver reliability of the modified Tardieu scale in the upper limb of children with hemiplegia. *Dev Med Child Neurol* 2004;46:267–72.
35 Pandyan AD, Gregoric M, Barnes MP, *et al*. Spasticity: clinical perceptions, neurological realities and meaningful measurement. *Disabil Rehabil* 2005;27:2–6.
36 Nielsen JB, Crone C, Hultborn H. The spinal pathophysiology of spasticity--from a basic science point of view. *Acta Physiol* 2007;189:171–80.
37 Trompetto C, Marinelli L, Mori L, *et al*. Pathophysiology of spasticity: implications for neurorehabilitation. *Biomed Res Int* 2014;2014:1–8.
38 Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther* 1987;67:206–7.
39 Powers RK, Campbell DL, Rymer WZ. Stretch reflex dynamics in spastic elbow flexor muscles. *Ann Neurol* 1989;25:32–42.
40 Alhusaini AAA, Dean CM, Crosbie J, *et al*. Evaluation of spasticity in children with cerebral palsy using Ashworth and Tardieu scales compared with laboratory measures. *J Child Neurol* 2010;25:1242–7.
41 World Health Organisation. How to use the ICF: a practical manual for using the International classification of functioning, disability and health (ICF) exposure draft for Comment, 2013. Available: http://apps.who.int/iris/bitstream/handle/10665/43737/9789241547321_eng.pdf;jsessionid=64A473FE7749C454CA [Accessed 30 May 2021].
42 Burridge J, Alt Murphy M, Buurke J, *et al*. A systematic review of international clinical guidelines for rehabilitation of people with neurological conditions: what recommendations are made for upper limb assessment? *Front Neurol* 2019;10:567.

Appendix 1 Full Search Strategy in All Searched Databases

Keywords:

Population:
(stroke or “cerebrovascular accident*” or “cerebro vascular accident*” or “cerebral vascular event*” or cva or cve) OR (stroke [MeSh] (where possible) OR

(“spinal cord injur*” or “spinal injur*” or sci) OR (spinal cord injuries [MeSH]) (where possible) OR

(“brain injur*” or “head injur*” or “traumatic brain injur*” or “acquired brain injur*” or tbi) OR (brain injuries [MeSh]) (where possible) OR

(“multiple sclerosis” or ms) OR (multiple sclerosis [MeSh]) (where possible) OR

(disability* or disabled) OR (disabilities [MeSh])(where possible) AND

Intervention (action): “Modified Tardieu Scale” AND

Outcome: (spasticity or hypertonia or spasm) OR (spasticity [MeSh]) (where possible)

Pubmed/Medline:

Number	Search
#1	“Modified Tardieu Scale” [Text Word]
#2	"Stroke"[Text Word] OR "cerebrovascular accident*"[Text Word] OR "cerebro vascular accident*"[Text Word] OR "cva"[Text Word] OR "cerebral vascular event*"[Text Word] OR "cve"[Text Word] OR "Stroke"[MeSH Terms]
#3	"Brain Injuries"[MeSH Terms] OR "brain injur*"[Text Word] OR "head injur*"[Text Word] OR "traumatic brain injur*"[Text Word] OR "acquired brain injur*"[Text Word] OR "tbi"[Text Word] OR "abi"[Text Word]
#4	"Spinal Cord Injuries"[MeSH Terms] OR "spinal cord injur*"[Text Word] OR "sci"[Text Word] OR "spinal injur*"[Text Word]
#5	"Multiple Sclerosis"[MeSH Terms] OR "Multiple Sclerosis"[Text Word] OR "ms"[Text Word]
#6	"disabilit*"[Text Word] OR "disabled"[Text Word] OR "Disability Studies"[MeSH Terms] OR "Physician Impairment"[MeSH Terms]
#7	#2 OR #3 OR #4 OR #5 OR #6
#8	"Muscle Spasticity"[MeSH Terms] OR "spasticity"[Text Word] OR "hypertonia"[Text Word] OR "spasm"[Text Word]
#9	#1 AND # 7 AND #8
#10	Limits: End date: 2020.12.31

CINAHL PLUS:

Number	Search Terms	Search options
S1	TX“Modified Tardieu Scale”	Search modes: Boolean/Phrase Expanders: Apply equivalent subjects
S2	TX stroke or "cerebrovascular accident*" or "cerebro vascular accident*" or cva or “cerebral vascular event*” or cve	
S3	(MH “Stroke+”)	
S4	S1 OR S2	
S5	TX “spinal cord injur*” or sci or “spinal injur*”	
S6	(MH “Spinal Cord Injuries+”) OR (MH “ Spinal Injuries+”)	
S7	S5 OR S6	
S8	TX “Brain injur*” or “head injur*” or “traumatic brain injur*” or “acquired brain injur*” or tbi or abi	
S9	(MH “Brain Injuries”)	
S10	S8 OR S9	
S11	TX “Multiple sclerosis” or ms	
S12	(MH “Multiple Sclerosis+”)	
S13	S11 OR S12	
S14	TX Disabilit* or disabled	
S15	(MH “Disability Management”) OR (MH “Disability Evaluation+”)	
S16	S14 OR 15	
S17	S4 OR S7 OR S10 OR S13 OR 16	
S18	TX Spasticity or hypertonia or spasm	
S19	(MH “Muscle Spasticity”)	
S20	S18 OR S19	
S21	S1 AND S17 AND S20	
S22	Limit: Publication Year: -2020	

EMBASE:

Number	Searches
1	"Modified Tardieu Scale" .tw.
2	(stroke or "cerebrovascular accident*" or "cerebro vascular accident*" or cva or "cerebral vascular event*" or cve).tw.
3	stroke.mp. or exp cerebrovascular accident/
4	2 OR 3
5	("spinal cord injur*" or "spinal injur*" or sci).tw.
6	spinal cord injury.mp. or exp spinal cord injury/
7	5 OR 6
8	("brain injur*" or "head injur*" or "traumatic brain injur*" or "acquired brain injur*" or tbi or abi).tw.
9	brain injury.mp. or exp brain injury/
10	8 OR 9
11	("multiple sclerosis" or ms).tw.
12	multiple sclerosis.mp. or exp multiple sclerosis/
13	11 OR 12
14	(disabilit* or disabled).tw.
15	exp disability/
16	14 OR 15
17	4 OR 7 OR 10 OR 13 OR 16
18	(spasticity or hypertonia or spasm).tw.
19	spasticity.mp. or exp spasticity/
20	18 OR 19
21	1 AND 17 AND 20 Limits: Specific year range: to 2020

Cochrane Library:

Number	Search	Limits/Filters
#1	All text "Modified Tardieu Scale"	
#2	MeSH descriptor: [Stroke] explode all trees	
#3	All text-Stroke or "cerebrovascular accident*" or "cerebro vascular accident*" or "cerebral vascular event*" or cva or cve	
#4	#2 or #3	
#5	MeSH descriptor: [Spinal Cord Injuries] explode all trees OR [Spinal injuries] explode all trees	
#6	All text "spinal cord injur*" or "spinal injur*" or sci	
#7	#5 OR #6	
#8	MeSH descriptor: [Brain injuries] explode all trees	
#9	All text "brain injur*" or "head injur*" or "traumatic brain injur*" or "acquired brain injur*" or tbi or abi	
#10	#8 or #9	
#11	MeSH descriptor: [Multiple Sclerosis] explore all trees	
#12	All text "multiple sclerosis" or ms	
#13	#11 OR #12	
#14	MeSH descriptor: [Disability Evaluation] explode all trees	
#15	All text Disabilit* or disabled	
#16	#14 OR #15	
#20	#4 OR #7 OR #10 OR #13 OR #16	
#21	MeSH descriptor: [Muscle spasticity] explode all trees	
#22	All text Spasticity or hypertonia or spasm	
#23	#21 or #23	
#24	#23 AND #20 AND #1	Limit: with Cochrane library publication date to Dec 2020

Web of Science:

Set	Search
#1	ALL=(“Modified Tardieu Scale”)
#2	ALL=(Stroke or "cerebrovascular accident*" or "cerebro vascular accident*" or "cerebral vascular event*" or cva or cve)
#3	ALL=("spinal cord injur*" or "spinal injur*" or sci)
#4	ALL=("brain injur*" OR "head injur*" OR "traumatic brain injur*" OR "acquired brain injur*" OR tbi OR abi)
#5	ALL=("multiple sclerosis" OR ms)
#6	ALL=(disabilit* or disabled)
#7	#2 OR #3 OR #4 OR #5 OR #6
#8	ALL=(spasticity OR hypertonia OR spasm)
#9	#1 AND #7 AND #8
#10	Exclude—publication years(2021)

PEDro:

Abstract & Title: “Modified Tardieu Scale” “Spasticity”
Subdiscipline: neurology
When Searching: Match all search terms (AND)

Appendix 2 Quality Assessment Tool for Observational Cohort and Cross-sectional Studies [19]

Criteria	Yes	No	Other (CD,NR, NA)*
1.Was the research question or objective in this paper clearly stated?			
2.Was the study population clearly specified and defined?			
3.Was the participation rate of eligible persons at least 50%?			
4.Were all the subjects selected or recruited from the same or similar population (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?			
5.Was a sample size justification, power description, or variance and effect estimates provided?			
6.For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?			
7.Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?			
8.For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome(e.g., categories of exposure, or exposure measured as continuous variable)?			

9.Were the exposure measures(independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
10.Was the exposure(s) assessed more than once over time?			
11.Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
12.Were the outcome assessors blinded to the exposure status of participants?			
13.Was loss to follow-up after baseline 20% or less?			
14.Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?			