Journal of Hand Therapy xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Hand Therapy

journal homepage: www.elsevier.com/locate/jht

Research Paper

Fugl-Meyer Assessment for upper extremity in stroke: A pyschometric systematic review

Pablo de Blas-Zamorano, PhD Candidate [a,b], Pedro Montagut-Martínez, PhD [c], David Pérez-Cruzado, PhD [d,*], Jose-Antonio Merchan-Baeza, PhD [e]

[a] Health Sciences Faculty, University of Málaga, Málaga, Spain
[b] Faculty of Social and Health Sciences, University of Murcia, Murcia, Spain
[c] Department of Nursing, UCAM Universidad Católica de Murcia, Murcia, Spain
[d] Department of Physiotherapy, Health Sciences Faculty, University of Málaga, Málaga, Spain
[e] Research Group on Methodology, Methods, Models and Outcomes of Health and Social Sciences (M3O), Faculty of Health Sciences and Welfare, Centre for Health and Social Care Research (CESS), University of Vic-Central University of Catalonia (UVic-UCC), Vic, Cataluña, Spain

ARTICLE INFO

Article history:
Received 13 October 2024
Revised 8 March 2025
Accepted 9 April 2025
Available online xxxx

Keywords:
Fugl-Meyer Assessment
Stroke
Sensorimotor function
Upper limb

ABSTRACT

Background: Stroke is a very common pathology, with a high prevalence of impairment of the sensorimotor function of the upper limb. There are many scales that assess this variable. Fugl-Meyer Assessment is one of the most widely used, but there are many versions of this scale to assess this construct.
Purpose: To identify, critically appraise, and summarize the different Fugl-Meyer Assessment (FMA) versions to recognize which is better to measure upper-extremity sensorimotor function in stroke patients.
Study Design: A psychometric systematic review was carried out in this research.
Methods: An exhaustive search was conducted during June and July 2023 in the following databases: Pubmed, Web of Science (WoS), Open Grey, Scielo, Cochrane, Dialnet, and LILACS. By the most updated consensus-based standards fot the selection of health status Measurement INstruments methodology (COSMIN) and preferred reporting items for systematic reviews and meta-analyses (PRISMA) statements for a systematic review of the measurement properties of existing patient-reported outcome measures to select the most appropriate outcome measurement instrument. The protocol of this systematic review was registered in Open Science Framework (https://osf.io/dy54p/).
Results: About 22 versions of the FMA were identified in 36 studies. The Upper Extremity subscale of the Fugl-Meyer Assessment (FMA-UE) was the version most frequently analyzed (n = 13) among the included articles.
Conclusions: Fugl-Meyer Assessment (Full version, 50 items, FMA) was the questionnaire version with the most evaluated psychometric characteristics and most translations, with the better score regarding the quality of evidence, which indicates that is the recommended FMA version to assess upper-extremity sensorimotor function in stroke patients.

Abbreviations: ARAT, Action Research Arm Test; BBT, Box and Block Test; CAHAI, Chedoke Arm and Hand Activity Inventory; COSMIN, Consensus-based standards for the selection of health status Measurement INstruments methodology; FMA, Fugl-Meyer Assessment; FMA-6, Shortened Fugl-Meyer Assessment Upper Extremity (6-item); FMA-12, Short form of the Fugl-Meyer Assessment (12-item); FMA-13-Kinect, Fugl-Meyer Assessment Upper Extremity (13 items) evaluating using Depth-Sensing Camera (Kinect); FMA-37, Short form of the Fugl-Meyer Assessment (37-item); FMA-Col, Colombian Spanish version of the Fugl-Meyer Assessment; FMA-Dan, Danish version of the Fugl-Meyer Assessment; FMA-Ita, Italian version of the Fugl-Meyer Assessment; FMA-Jap, Japanese version of the Fugl-Meyer Assessment; FMA-Kor, Korean version of the Fugl-Meyer Assessment; FMA-ML-10, Machine-Learning Fugl-Meyer (10-item); FMA-Remote, Remote evaluation of the upper extremities with Fugl-Meyer Assessment; FMA-Rom, Romanian version of the Fugl-Meyer Assessment; FMA-S, Sensory subscale of the Fugl-Meyer Assessment; FMA-UE-Dan, Danish version of the upper-extremity subscale of the Fugl-Meyer Assessment; FMA-UE-Ita, Italian version of the Fugl-Meyer Assessment Upper Extremity; FMA-UE-Rom, Romanian version of the upper extremity Fugl-Meyer Assessment; FMA-UE-SE, Shoulder and Elbow Fugl-Meyer Assessment; FMA-UE, Upper Extremity subscale of the Fugl-Meyer Assessment; FMA-UE-w/h, Wrist and Hand Fugl-Meyer; FMA-Urd, Urdu version of the Fugl-Meyer Assessment; ICC, Intraclass Correlation coefficient; MAL, Motor Activity Log; MAS, Motor Assessment Scale; NHPT, Nine-Hole Peg Test; NIHSS, National Institute of Health Stroke Scale; OSF, Open Science Framework; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; WMFT, Wolf Motor Function Test; WoS, Web of Science

* Corresponding author. Department of Physiotherapy, Health Sciences Faculty, University of Málaga, Málaga, Spain.
E-mail address: dpcruzado@uma.es (D. Pérez-Cruzado).

0894-1130/
https://doi.org/10.1016/j.jht.2025.04.004

Introduction

Stroke is the second leading cause of death[1] and the first leading cause of disability worldwide.[2] Fifteen million people suffer a stroke each year around the world.[3] The burden of disability after a stroke is significant and is increasing very fast.[1] In the acute phase after stroke, approximately 60-80% of survivors present upper or lower limb motor impairments.[4]

Physical limitations are common sequels after stroke, manifested in many physiological domains and motor aspects such as spasticity, muscle synergies, coordination, strength, range of motion, dexterity, skin sensation, position sense, and arm stability.[5–7] All of them cause an inability to use the upper extremity to reach, grasp, and manipulate objects, affecting the performance of daily living tasks such as dressing, feeding, or bathing[5,8] and detrimental impact on the quality of life.[8]

With these data, it is necessary to have tools available to assess the different physical deficits caused by stroke. There are many tools to evaluate stroke patients' sensorimotor and functional construct. Some of them are the Fugl-Meyer Assessment (FMA), Action Research Arm Test (ARAT), Wolf Motor Function Test (WMFT), Box and Block Test (BBT), etc.[4,9,10]

Choosing the most appropriate instrument to assess upper-extremity sensorimotor function in stroke patients is challenging, and the choice should depend on the measurement properties and feasibility of the instrument.[11] Some systematic reviews demonstrate that the FMA is the most commonly used for clinical practice and research and has seen an increase in its use in recent years.[9,12,13]

FMA is internationally known, with different versions and translations. It has been translated into different languages and cultures.[4,9,10,13–15]

FMA is the "gold standard" for assessing upper-extremity sensorimotor function, due to its ability to assess aspects such as movement within synergies, mixing synergies, reflexes, wrist, hand, grip, coordination, and speed movements, thus providing a large amount of information that is very useful for understanding the sensorimotor capacity of the affected upper limb after a stroke.[9,10,12] Despite being widely used in both clinical practice and rehabilitation, a review using the consensus-based standards for the selection of health status Measurement INstruments methodology (COSMIN) that evaluated all the versions of this scale that exist to date had never been done before, which allows us to reach a clear consensus on which of the multiple existing versions of the FMA is the most recommended for its psychometric properties, to be the version with the greatest validity, and therefore, most advisable for use in both research and clinical practice.[17,18]

For all these reasons, it has been considered pertinent and necessary to carry out a systematic review to clarify which of the different versions should be chosen as the most appropriate for the sensorimotor assessment of people with stroke. For this reason, this systematic review aimed to identify, critically evaluate, and summarize the different FMA versions to recognize which measures upper-extremity sensorimotor function in stroke patients better.

Methods

Design

A systematic review of the measurement properties of the different FMA versions that assess the upper-extremity sensorimotor function in stroke patients was carried out.

The COSMIN checklist with a four-point scale was used to evaluate the measurement properties of existing patient-reported outcome measures to select the most appropriate outcome measurement instrument.[17,18]

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statements[16–18] were consulted in Equator website and were used to improve the reporting of systematic reviews and meta-analyses.

The protocol of this process, with the information and the methodology used, was registered in the Open Science Framework (OSF) (https://osf.io/dy54p/).

Search method

An exhaustive search was carried out during June and July 2023 in different databases, Medline (Pubmed), Web of Science (WoS), Open Grey, Scielo, Cochrane, Dialnet, and LILACS, by four independent reviewers. The search strategy included a combination of the following five aspects: (1) construct search, (2) AND population search, (3) AND instrument search, (4) filter for AND measurement properties, and (5) "NOT" exclusion filter.[19] To locate the higher number of tools and their different versions, the dates were not limited. To perform this search, the following key terms were used: "Fugl-Meyer", "Fugl-Meyer Assessment", "FMA", "Stroke", "Cerebral Stroke", "Cerebrovascular Accident", "assessment", "tool", "instrument", "test", "measure", and "scale" adapting this strategy to each of the search systems of the different databases. The search was limited to articles published in English and Spanish. In addition, the reference list of the included articles was checked to identify other relevant documents (Table 1).

Eligibility criteria

The inclusion criteria for this study were: (1) people who had suffered a stroke, (2) the outcome had to be the upper-extremity sensorimotor function evaluated with any FMA version, and (3) the studies had to include a FMA version and evaluate at least one psychometric property from the COSMIN checklist.

The exclusion criteria were: (1) instrument or article not available in English or Spanish languages, (2) editorials, speeches, conferences abstracts, biographies, comments, case reports, doctoral theses, practice guides, opinion articles, or articles in which selective methodology was not used.

Four reviewers independently screened titles and abstracts to check if they met the inclusion criteria. When it was impossible to make decisions based only on the abstract and title, the full-text analysis was carried out by the four reviewers. Conflicts were resolved by sharing the different problematic aspects and discussing them between the four reviewers until an agreement was reached.

Quality appraisal

The studies included in this review were evaluated by four researchers, according to the most up-to-date COSMIN Risk of Bias checklist with a four-point scale[18] (available in https://www.cosmin.nl/). It contains nine boxes with 5-18 items, and each box provides a methodological quality score for content validity, structural validity, internal consistency, cross-cultural validity, measurement error, reliability, criterion validity, hypothesis testing, and responsiveness. The content evaluated in each box can be rated as "very good", "adequate", "doubtful", and "inadequate". The score for each psychometric property is acquired by taking the lowest score of any item for each property.[18] Since the volume of articles was not been excessive, no cutoff points were established to discard any article based on methodological quality.

Table 1
Search strategy

Search strings for Pubmed	
1	("Fugl" OR "Fugl Meyer" OR "Fugl Meyer Assessment" OR "FMA")
2	("Stroke" OR "Cerebral Stroke" OR "Cerebrovascular Accident" OR "Cerebral Strokes" OR "Strokes" OR "Cerebrovascular Stroke" OR "CVA" OR "Cerebrovascular" OR "Cerebrovascular Accidents" OR "Brain Accident" OR "Brain Accidents")
3	("Instrument" OR "instruments" OR "measure" OR "measures" OR "questionnaire" OR "questionnaires" OR "scale" OR "scales" OR "tool" OR "tools" OR "survey" OR "test")
4	("Instrumentation" OR "methods" OR "Validation Studies" OR "Comparative Study" OR "psychometrics" OR "psychometr*" OR "clinimetr*" OR "clinometr*" OR "outcome assessment" OR "outcome assessment" OR "outcome measure*" OR "observer variation" OR "observer variation" OR "Health Status Indicators" OR "reproducibility of results" OR "reproducib*" OR "discriminant analysis" OR "reliab*" OR "unreliab*" OR "valid*" OR "coefficient" OR "homogeneity" OR "homogeneous" OR "internal consistency" OR "cronbach*" OR "correlation*" OR "selection*" OR "reduction*" OR "agreement" OR "precision" OR "imprecision" OR "precise values" OR "test-retest" OR "reliab*" OR "stability" OR "interrater" OR "inter-rater" OR "intrarater" OR "intra-rater" OR "intertester" OR "inter-tester" OR "intratester" OR "intra-tester" OR "interobserver" OR "inter-observer" OR "intraobserver" OR "intraobserver" OR "intertechnician" OR "inter-technician" OR "intratechnician" OR "intra-technician" OR "interexaminer" OR "inter-examiner" OR "intraexaminer" OR "intra-examiner" OR "interassay" OR "inter-assay" OR "intraassay" OR "intra-assay" OR "interindividual" OR "inter-individual" OR "intraindividual" OR "intra-individual" OR "interparticipant" OR "inter-participant" OR "intraparticipant" OR "intra-participant" OR "kappa" OR "kappa's" OR "kappas" OR "repeatab*" OR "generaliza*" OR "generalisa*" OR "concordance" OR "intraclass correlation*" OR "discriminative" OR "known group" OR "factor analysis" OR "factor analyses" OR "dimension*" OR "subscale*" OR "item discriminant" OR "interscale correlation*" OR "error" OR "errors" OR "individual variability" OR "variability" OR "uncertainty" OR "standard error of measurement" OR "sensitiv*" OR "responsive*" OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR "IRT" OR "Rasch" OR "Differential item functioning" OR "DIF" OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
5	#1 AND #2 AND #3 AND #4
6	("Protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "editorial" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
7	#5 NOT #6
8	FILTER: Language (English and Spanish)
Search strings for Web of Science (WoS)	
1	TI=("Fugl" OR "Fugl Meyer" OR "Fugl Meyer Assessment" OR "FMA")
2	TS=("Stroke" OR "Cerebral Stroke" OR "Cerebrovascular Accident" OR "Cerebral Strokes" OR "Strokes" OR "Cerebrovascular Stroke" OR "CVA" OR "Cerebrovascular" OR "Cerebrovascular Accidents" OR "Brain Accident" OR "Brain Accidents")
3	TS=("Instrument" OR "instruments" OR "measure" OR "measures" OR "questionnaire" OR "questionnaires" OR "scale" OR "scales" OR "tool" OR "tools" OR "survey" OR "test")
4	TS=("Instrumentation" OR "methods" OR "Validation Studies" OR "Comparative Study" OR "psychometrics" OR "psychometr*" OR "clinimetr*" OR "clinometr*" OR "outcome assessment (health care)" OR "outcome assessment" OR "outcome measure*" OR "observer variation" OR "observer variation" OR "Health Status Indicators" OR "reproducibility of results" OR "reproducib*" OR "discriminant analysis" OR "reliab*" OR "unreliab*" OR "valid*" OR "coefficient" OR "homogeneity" OR "homogeneous" OR "internal consistency" OR ("cronbach*" AND ("alpha" OR "alphas")) OR ("item" AND ("correlation*" OR "selection*" OR "reduction*")) OR "agreement" OR "precision" OR "imprecision" OR "precise values" OR "test-retest" OR ("test" AND "retest") OR ("reliab*" AND ("test" OR "retest")) OR "stability" OR "interrater" OR "inter-rater" OR "intrarater" OR "intra-rater" OR "intertester" OR "inter-tester" OR "intratester" OR "intra-tester" OR "interobserver" OR "inter-observer" OR "intraobserver" OR "intraobserver" OR "intertechnician" OR "inter-technician" OR "intratechnician" OR "intra-technician" OR "interexaminer" OR "inter-examiner" OR "intraexaminer" OR "intra-examiner" OR "interassay" OR "inter-assay" OR "intraassay" OR "intra-assay" OR "interindividual" OR "inter-individual" OR "intraindividual" OR "intra-individual" OR "interparticipant" OR "inter-participant" OR "intraparticipant" OR "intra-participant" OR "kappa" OR "kappa's" OR "kappas" OR "repeatab*" OR (("replicab*" OR "repeated") AND ("measure" OR "measures" OR "findings" OR "result" OR "results" OR "test" OR "tests")) OR "generaliza*" OR "generalisa*" OR "concordance" OR ("intraclass" AND "correlation*") OR "discriminative" OR "known group" OR "factor analysis" OR "factor analyses" OR "dimension*" OR "subscale*" OR ("multitrait" AND "scaling" AND ("analysis" OR "analyses")) OR "item discriminant" OR "interscale correlation" OR "error" OR "errors" OR "individual variability" OR ("variability" AND ("analysis" OR "values")) OR ("uncertainty" AND ("measurement" OR "measuring")) OR "standard error of measurement" OR "sensitiv*" OR "responsive*" OR (("minimal" OR "minimally" OR "clinical" OR "clinically") AND ("important" OR "significant" OR "detectable") AND ("change" OR "difference")) OR ("small*" AND ("real" OR "detectable") AND ("change" OR "difference")) OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR "IRT" OR "Rasch" OR "Differential item functioning" OR "DIF" OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
5	#1 AND #2 AND #3 AND #4
6	("Protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "editorial" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
7	#5 NOT #6
8	FILTER: Language (English and Spanish)
Search strings for Open Grey	
1	("Fugl" OR "Fugl Meyer" OR "Fugl Meyer Assessment" OR "FMA")
2	("Stroke" OR "Cerebral Stroke" OR "Cerebrovascular Accident" OR "Cerebral Strokes" OR "Strokes" OR "Cerebrovascular Stroke" OR "CVA" OR "Cerebrovascular" OR "Cerebrovascular Accidents" OR "Brain Accident" OR "Brain Accidents")
3	("Instrument" OR "instruments" OR "measure" OR "measures" OR "questionnaire" OR "questionnaires" OR "scale" OR "scales" OR "tool" OR "tools" OR "survey" OR "test")

(*continued on next page*)

Table 1 (*continued*)

4	(“Instrumentation” OR "methods” OR “Validation Studies” OR “Comparative Study” OR “psychometrics” OR "psychometr*" OR "clinimetr*" OR "clinometr*" OR “outcome assessment (health care)” OR “outcome assessment” OR “outcome measure*" OR “observer variation” OR “observer variation" OR “Health Status Indicators” OR “reproducibility of results" OR "reproducib*" OR “discriminant analysis” OR "reliab*" OR "unreliab*" OR "valid*" OR "coefficient" OR "homogeneity" OR "homogeneous" OR “internal consistency” OR ("cronbach*" AND ("alpha" OR "alphas")) OR ("item" AND ("correlation*" OR "selection*" OR "reduction*")) OR "agreement" OR "precision" OR "imprecision" OR “precise values” OR "test-retest" OR ("test" AND "retest") OR ("reliab*" AND ("test" OR "retest")) OR "stability" OR "interrater" OR "inter-rater" OR "intrarater" OR "intra-rater" OR "intertester" OR "inter-tester" OR "intratester" OR "intra-tester" OR "interobserver" OR "inter-observer" OR "intraobserver" OR "intraobserver" OR "intertechnician" OR "inter-technician" OR "intratechnician" OR "intra-technician" OR "interexaminer" OR "inter-examiner" OR "intraexaminer" OR "intra-examiner" OR "interassay" OR "inter-assay" OR "intraassay" OR "intra-assay" OR "interindividual" OR "inter-individual" OR "intraindividual" OR "intra-individual" OR "interparticipant" OR "inter-participant" OR "intraparticipant" OR "intra-participant" OR "kappa" OR "kappa's" OR "kappas" OR "repeatab*" OR (("replicab*" OR "repeated") AND ("measure" OR "measures" OR "findings" OR "result" OR "results" OR "test" OR "tests")) OR "generaliza*" OR "generalisa*" OR "concordance" OR ("intraclass" AND "correlation*") OR "discriminative" OR “known group” OR “factor analysis” OR “factor analyses” OR "dimension*" OR "subscale*" OR ("multitrait" AND "scaling" AND ("analysis" OR "analyses")) OR “item discriminant” OR “interscale correlation” OR "error" OR "errors" OR “individual variability” OR ("variability" AND ("analysis" OR "values")) OR ("uncertainty" AND ("measurement" OR "measuring")) OR “standard error of measurement” OR "sensitiv*" OR "responsive*" OR (("minimal" OR "minimally" OR "clinical" OR "clinically") AND ("important" OR "significant" OR "detectable") AND ("change" OR "difference")) OR ("small*" AND ("real" OR "detectable") AND ("change" OR "difference")) OR “meaningful change” OR “ceiling effect” OR “floor effect” OR “Item response model" OR "IRT" OR "Rasch" OR “Differential item functioning” OR "DIF" OR “computer adaptive testing” OR “item bank” OR “cross-cultural equivalence”)
5	#1 AND #2 AND #3 AND #4
6	(“Protocol” OR “addresses” OR “biography” OR “case reports” OR “comment” OR “editorial” OR “congresses” OR “consensus development conference” OR “consensus development conference” OR “practice guideline”)
7	#5 NOT #6
8	FILTER: Language (English and Spanish)
Search strings for Scielo	
1	(Fugl Meyer)
2	(Stroke)
3	#1 AND #2
4	FILTER: Language (English and Spanish)
Search strings for Cochrane	
1	("Fugl*":ti,ab OR "Fugl Meyer*:ti,ab OR "Fugl Meyer Assessment*":ti,ab OR "FMA*":ti,ab)
2	("Stroke*":ti,ab OR "Cerebral Stroke*":ti,ab OR "Cerebrovascular Accident*":ti,ab OR "CVA*":ti,ab OR "Cerebrovascular*":ti,ab OR "Brain Accident*":ti,ab OR "Cerebrovascular Stroke*":ti,ab)
3	("Instrument":ti,ab OR "instruments":ti,ab OR "measure":ti,ab OR "measures":ti,ab OR "questionnaire":ti,ab OR "questionnaires":ti,ab OR "scale":ti,ab OR "scales":ti,ab OR "tool":ti,ab OR "tools":ti,ab OR "survey":ti,ab OR "test":ti,ab)
4	AB (“Instrumentation” OR methods” OR “Validation Studies” OR “Comparative Study” OR “psychometrics” OR "psychometr*" OR "clinimetr*" OR "clinometr*" OR “outcome assessment (health care)” OR “outcome assessment” OR “outcome measure*" OR “observer variation” OR “observer variation” OR “Health Status Indicators” OR “reproducibility of results” OR "reproducib*" OR “discriminant analysis” OR "reliab*" OR "unreliab*" OR "valid*" OR "coefficient" OR "homogeneity" OR "homogeneous" OR “internal consistency” OR ("cronbach*" AND ("alpha" OR "alphas")) OR ("item" AND ("correlation*" OR "selection*" OR "reduction*")) OR "agreement" OR "precision" OR "imprecision" OR “precise values” OR "test-retest" OR ("test" AND "retest") OR ("reliab*" AND ("test" OR "retest")) OR "stability" OR "interrater" OR "inter-rater" OR "intrarater" OR "intra-rater" OR "intertester" OR "inter-tester" OR "intratester" OR "intra-tester" OR "interobserver" OR "inter-observer" OR "intraobserver" OR "intraobserver" OR "intertechnician" OR "inter-technician" OR "intratechnician" OR "intra-technician" OR "interexaminer" OR "inter-examiner" OR "intraexaminer" OR "intra-examiner" OR "interassay" OR "inter-assay" OR "intraassay" OR "intra-assay" OR "interindividual" OR "inter-individual" OR "intraindividual" OR "intra-individual" OR "interparticipant" OR "inter-participant" OR "intraparticipant" OR "intra-participant" OR "kappa" OR "kappa's" OR "kappas" OR "repeatab*" OR (("replicab*" OR "repeated") AND ("measure" OR "measures" OR "findings" OR "result" OR "results" OR "test" OR "tests")) OR "generaliza*" OR "generalisa*" OR concordance OR (intraclass AND correlation*) OR discriminative OR “known group” OR “factor analysis” OR “factor analyses” OR dimension* OR subscale* OR (multitrait AND scaling AND (analysis OR analyses)) OR “item discriminant” OR “interscale correlation" OR error OR errors OR “individual variability” OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR “standard error of measurement” OR sensitiv* OR responsive* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small*AND (real OR detectable) AND (change OR difference)) OR “meaningful change” OR “ceiling effect” OR “floor effect” OR “Item response model” OR IRT OR Rasch OR “Differential item functioning” OR DIF OR “computer adaptive testing” OR “item bank” OR “cross-cultural equivalence”)
5	#1 AND #2 AND #3 AND #4
6	(“Protocol” OR “addresses” OR “biography” OR “case reports” OR “comment” OR “editorial” OR “congresses” OR “consensus development conference” OR “consensus development conference” OR “practice guideline”)
7	#5 NOT #6
8	FILTER: Language (English and Spanish)
Search strings for Dialnet	
1	(Fugl Meyer Assessment)
2	(Stroke)
3	#1 AND #2
4	(Tesis)
5	#3 NOT #4
6	FILTER: Language (English and Spanish)

(*continued on next page*)

Table 1 (*continued*)

Search strings for LILACS	
1	((Fugl) OR (Fugl Meyer) OR (Fugl Meyer Assessment) OR (FMA))
2	(Stroke)
3	((Psychometr*) OR (clinimetr*))
4	#1 AND #2 AND #3
5	FILTER: Language (English and Spanish)

Source: self-made.

Data extraction

Titles and abstracts of all articles identified in this review were reviewed. Four experts independently reviewed full-text articles, and conflicts were resolved by consensus. The information from each article was extracted based on the following aspects: name of the tool and version used, study population, tool characteristics, analysis of psychometric properties, COSMIN score, and statistical results of the research.

Synthesis

Data on the psychometric properties of the articles that met the inclusion criteria were scored according to the COSMIN checklist.[18] Four independent reviewers established the scores of each article's psychometric properties, and differences were resolved through discussion and consensus. Subsequently, the result of each of the measurement properties of each study was qualified according to the Terwee criteria,[11] classifying each of the results as "positive" (+), "negative" (-), or "indeterminate" (?). Finally, after determining the methodological quality for each of the individual studies and evaluating the measurement properties, the results were qualitatively summarized to determine both the quality of the evidence and the strength of the recommendations by rating them as "high", "moderate", "low", or "very low" according to the risk of bias, inconsistency, imprecision, and indirectness, using the system GRADE.[20] Four independent reviewers assigned scores for each article's psychometric properties and the differences were resolved through discussion and consensus.

Results

Search outcomes

In total, 8.178 articles were identified, of which, 36 were finally included in this systematic review, up to date on July 2023. The selection process is reflected in the flowchart, shown in Figure 1.

Characteristics of the included Fugl-Meyer Assessment versions

Twenty-two versions of FMA were identified that were designed to assess upper extremity sensorimotor function in stroke patients. The characteristics of the included FMA versions for measuring upper-extremity sensorimotor function and studies are presented in Table 2. In all cases, the original language used was English. However, several of these questionnaires were found to have undergone cross-cultural validation: Colombian,[27] Danish,[28,49] Italian,[29,50] Japanese,[30] Korean,[31] Urdu,[32] Romanian,[33] and Brazilian Portuguese.[37]

Most of the studies were conducted in hospitals and neurological rehabilitation centers around some places in Asia (n = 16, 40%), America (n = 13, 32.5%), Europe (n = 10, 25%), and Middle East (n = 1, 2.5%). The most common locations from where the studies were developed were the United States (n = 7),[25,40,41,44,53–55] Taiwan (n = 7),[14,23,26,34,35,42,43] and Japan (n = 4).[15,30,36,52]

Regarding the sample size of the studies, these ranged from 10 patients included in some translations and cross-cultural adaptations of the FMA[27–29,32,33] to 512 patients in the Upper-Extremity subscale of the Fugl-Meyer Assessment (FMA-UE),[41] with wide heterogeneity in terms of sample size.

Regarding the number of items, the versions ranged from 6 items in the FMA-6[36] to 50 items in the full version FMA.[14,21–26]

For the type of item used in each version, all of them kept the original scoring format on a three-point ordinal scale (0-2).

Fugl-Meyer Assessment (full version, 50 items, FMA)

The initial full version FMA version was present in seven studies,[14,21–26] and all psychometric properties were evaluated. Cross-cultural validity/measurement invariance were evaluated in the seven translated versions (FMA-Col,[27] FMA-Dan,[28] FMA-Ita,[29] FMA-Jap,[30] FMA-Kor,[31] FMA-Urd,[32] and FM-Rom[33]), being the FMA (full version, 50 items, FMA) the most translated version in this study.

Short versions of FMA

In this review, three short versions of the FMA scale have been detected, with the exception of the subscale versions, such as FMA-UE, FMA-UE-SE, FMA-UE-w/h, or FMA-S. These versions are:

FMA-37. The Short Form of the FMA with 37 items was evaluated in three studies[14,26,34] for all psychometric properties, except content validity, internal consistency, and cross-cultural validity. Structural validity revealed unidimensional structure.[14,26] The test-retest reliability values were ≥0.92. The criterion validity were between r = 0.01(26) and r = 0.093.[14] Hypothesis testing ranged from r = 0.73 to r = 0.75.[14]

FMA-12. This instrument was included in five studies[14,24,26,34,35] evaluating the same psychometric properties as the FMA-37 version (all psychometric properties, except content validity, internal consistency, and cross-cultural validity). Structural validity shows unidimensional construct.[14,26] Correlation coefficient ranged from r = 0.01[26] to r = 0.99.[24] The hypothesis testing showed a correlation between r = 0.39[35] and r = 0.75.[14]

FMA-UE-6. The Shortened FMA Upper Extremity of six items was included in a single study,[36] evaluating five psychometric properties: the internal consistency (Cronbach's alpha = 0.924), reliability (ICC = 0.994), criterion validity (r = 0.97), and hypothesis testing (r = 0.91-0.94).

Technological versions

FMA-ML-10. The Machine-Learning Fugl-Meyer of 10 items was evaluated in one study,[14] evaluating the following psychometric properties: structural validity (one-dimensional structure), measurement error, reliability (ICC = 0.92), criterion validity (r = 0.99), and hypothesis testing (r = 0.67-0.68).

FMA-13-Kinect. The FMA Upper Extremity (13 items) evaluating using Depth-Sensing Camera (Kinect) was included in a single study,[27] with a criterion validity correlation coefficient of r = 0.873 and a hypothesis testing r value of 0.387.

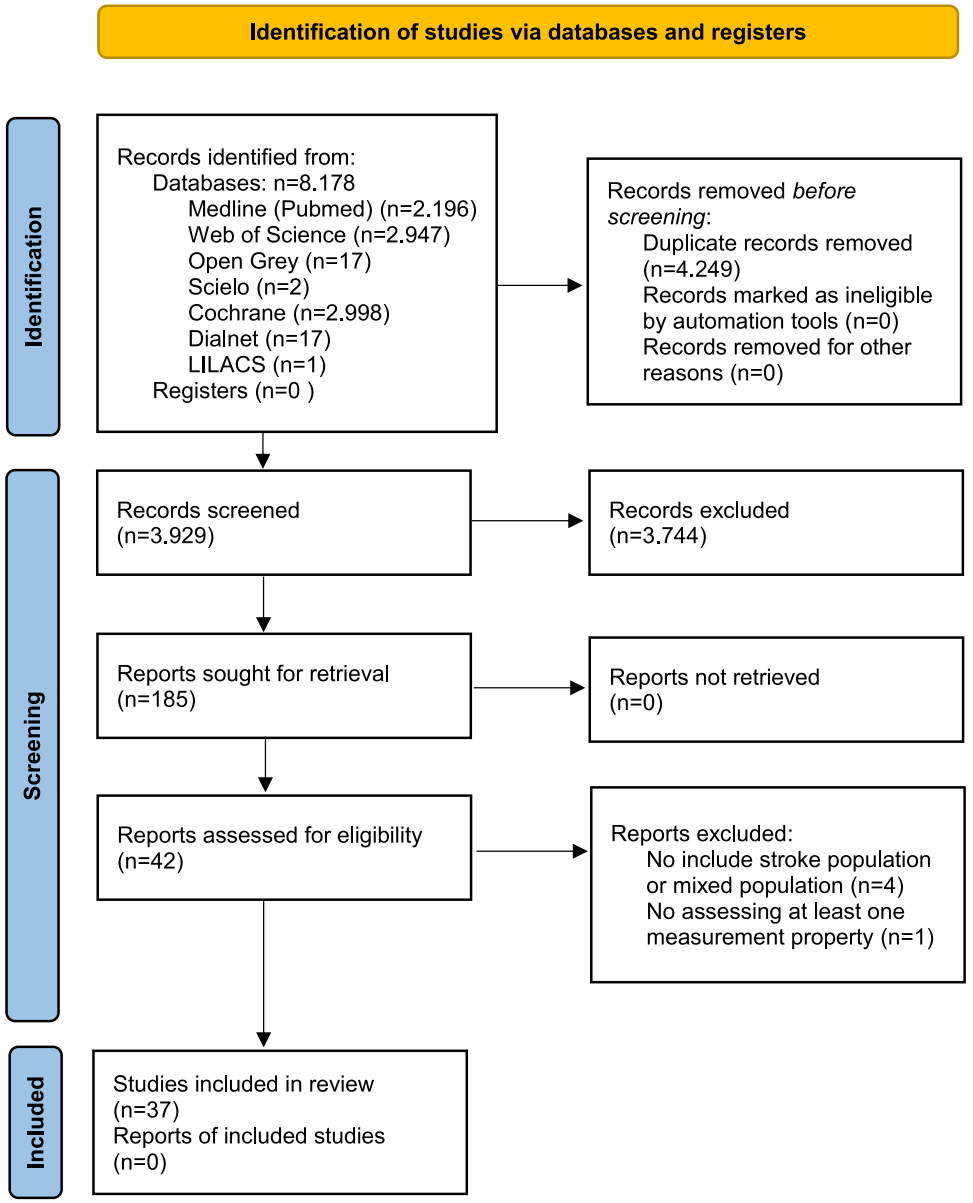

Figure 1. PRISMA 2020 flow diagram for new systematic reviews that included searches of databases and registers only. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

FMA-Remote. The Remote evaluation of the Upper Extremities with FMA was only included in one study,[52] measuring only one psychometric property, the test-retest reliability, with ICC = 0.998.

Upper extremity subscale of the Fugl-Meyer Assessment (FMA-UE)

FMA-UE is the most evaluated version in this systematic review, being evaluated in 13 studies,[15,36,38–48] and practically all psychometric properties except content validity were evaluated. This version has been translated into three languages: Danish (FMA-UE-Dan)[49], Italian (FMA-UE-Ita),[50] and Romanian (FMA-UE-Rom).[51]

Psychometric properties

The main properties assessed were responsiveness (n = 39)[14,22–26,30–36,38–40,42,43,45–47,49,51,53,54], hypothesis testing (n = 34),[14,22–25,30–33,35,36,38–40,42,43,45–47,49,51,53,54] reliability (n = 33),[14,21,23–25,30–34,36,37,39,43,44,46,48–54] and criterion validity (n = 20).[14,23,24,26,34,36,45–47]

Methodological quality of the studies and quality of results

The results for the four-point COSMIN checklist of versions of the FMA for measuring upper extremity sensorimotor function in stroke patients are shown in Table 3, where it is seen that the two versions with the greatest number of psychometric properties evaluated and the greatest number of studies are FMA Full version and FMA-UE. Among the results of methodological quality, it can be seen that the version with the highest number of ratings as "very good" is FMA-UE (n = 26), followed by FMA Full version (n = 15); while FMA-UE is also the version with the highest number of "inadequate" scores (n = 6).

The methodological quality of the studies for each of the psychometric properties assessed was not as expected, resulting in "inadequate" or "doubtful" in practically all the versions due to the

Table 2
Studies characteristics and description of Fugl-Meyer Assessment in stroke patients

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
Fugl-Meyer Assessment (Full version, 50 items) (FMA)						
Sanford et al. [21]	12 stroke patients.	Ontario, Canada. Chedoke-McMaster Rehabilitation Center. Patients were assessed by 3 therapists on separate occasions.	Fugl-Meyer Assessment is an impairment measure that consists of 155 items, rated on a 3-point ordinal scale (0-2). In total, 33 items for upper extremity, 17 items for lower extremity, 7 items for balance, 12 items for sensation, and 22 items for passive joint motion and joint pain. FMA includes three categories: (1) motor function, (2) balance, and (3) sensation. The maximal score value for each of these categories and subcategories was as follows: the motor function for the upper (66 points) and lower (34 points) extremities; balance in sitting (6 points) and in standing (8 points); touch (24 points)l and position sense (16 points). The motor part of the scale includes 50 items of upper and lower extremity motor function, with a possible score that ranges from 0- 100 points (66 from upper extremity subscale and 34 from lower extremity subscale).	1. Reliability 2. Measurement error	1. Doubtful 2. Doubtful	1. ICC (95% CI) was 0.96 for total Fugl-Meyer Assessment score. The subsections varied from 0.61 for pain measurements to 0.97 for upper extremity.
Malouin et al. [22]	32 stroke patients.	Quebec, Canada. Francois-Charron Rehabilitation Center.	Fugl-Meyer Assessment as above.	1. Content validity 2. Hypothesis testing 3. Responsiveness	1. Doubtful 2. Very good 3. Very good	2. The correlation between the total MAS and total FMA scores (including motor function, balance, and sensation) was $r = 0.96$; this correlation decreased to 0.91 when the sensation score was removed. The strongest relationships were found for the upper extremity motor function ($r = 0.93$). All correlations were positive, strong, and significant ($p < 0.005$), except for the FMA sitting balance score ($r = -0.10$).

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
Lin et al. [23]	176 stroke patients.	Taiwan. The study sample was recruited from the registry of the Quality of Life after Stroke Study in Taiwan for the period from 1 December 1999-7 December 2001. The Quality of Life after Stroke Study is a prospective cohort study of stroke patients admitted to National Taiwan University Hospital.	Fugl-Meyer Assessment as above.	1. Internal consistency 2. Reliability 3. Criterion validity 4. Hypothesis testing 5. Responsiveness	1. Very good 2. Inadequate 3. Very good 4. Very good 5. Very good	1. The Cronbach's alphas of the FMA-S at four time points after stroke ranged from 0.94 to 0.98. Indicating excellent internal consistency. 2. The inter-rater agreement of the total score of the FMA-S was excellent, with ICC (95% CI) of 0.93 (0.85-0.96). 3. The FMA-S scores were weakly to moderately correlated FMA-M scores (Spearman's rho between 0.31 and 0.44). 4. The FMA-S scores were weakly to moderately correlated with the BI (Spearman's rho ranging from 0.38 to 0.53).
Hsueh et al. [24]	50 stroke patients for the first part of the study (validity and responsive-ness) and 60 stroke patients for the second part of the study (reliability).	China. Rehabilitation Department of Kaohsiung Medical University Hospital.	Fugl-Meyer Assessment as above.	1. Reliability 2. Measurement error 3. Criterion validity 4. Hypothesis testing 5. Responsiveness	1. Doubtful 2. Doubtful 3. Very good 4. Very good 5. Very good	1. Test-retest reliabilities were high for the UE and LE subscales and total motor scores of the 4 motor scales (all ICCs ≥ 0.93 and lower limits of 95% CI ≥ 0.89). 2. The correlations were high (ρ = 0.91-0.99) among the 4 clinical motor scales with their subscales at admission and discharge, indicating high concurrent validity. 3. The scores of the 4 motor scales with their subscales at admission were moderately correlated with the BI scores at discharge (ρ ≥ 0.66), which indicates acceptable predictive validity.
See et al. [25]	31 stroke patients.	Patient's data are from 2 robotic clinical stroke trials of arm motor therapy from the United States, Canada, and Germany.	Fugl-Meyer Assessment as above.	1. Reliability 2. Hypothesis testing 3. Responsiveness	1. Doubtful 2. Very good 3. Very good	1. For intra-rater reliability, ICC was 0.99 for FMA-Total score, for FMA-Proximal subscore, and for FMA-w/h subscore. 2. For inter-rater reliability, ICC was 0.99 in FMA-Total score, and 0.98 for both FMA-proximal and FMA-w/h subscore. 3. The correlation of the FMA total score with ARAT was 0.93; with pinch force was 0.88; with grip force was 0.74; with BBT was 0.86; with NHPT was 0.75; and with SIS was 0.86.

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
Chen et al. [26]	301 stroke patients at 14 d after stroke assessment; 276 subjects at 30 d post stroke; 226 at 90 d; and 213 participants evaluated at 180 d post stroke.	Taiwan. Data were obtained from a previous study. Patients were assessed with the 50-item Fugl-Meyer motor scale four times, at 14, 30, 90, and 180 d after stroke onset. All the assessments were administered by a trained occupational therapist.	Fugl-Meyer Assessment as above.	1. Structural validity 2. Criterion validity 3. Responsiveness	1. Doubtful 2. Very good 3. Very good	1. Latent parallel analysis and ordinal factor analysis indicated that all W/H-UE-FM items represent a single unidimensional construct, wrist, and hand motor ability. 2. FM-50 vs FM-37 was 0.61-0.80. FM-50 vs FM-12 was 0.80-0.00. FM-37 vs FM-12 was 0.01-0.03.
Lin et al. [14]	267 stroke patients.	Taiwan. Patients from secondary data from other studies.	Fugl-Meyer Assessment as above.	1. Structural validity 2. Reliability 3. Measurement error 4. Hypotheses testing 5. Responsiveness	1. Doubtful 2. Inadequate 3. Inadequate 4. Very good 5. Doubtful 6. Doubtful	1. Unidimensional structure. 2. ICC for FM was 0.98 3. Pearson correlation with BI for FMA was 0.67. Pearson correlation with PASS for FMA was 0.69.
Colombian Spanish version of the Fugl-Meyer Assessment (FMA-Col)						
Barbosa et al. [27]	10 stroke patients.	Bogota, Colombia. Central Military Hospital of Colombia.	Authors did a translation from the full version Fugl-Meyer Assessment as above to a Colombian version using a standardized forward and backward translation protocol.	1. Content validity 2. Cross-cultural validity	1. Doubtful 2. Inadequate	
Danish version of the Fugl-Meyer Assessment (FMA-Dan)						
Busk et al. [28]	10 subacute stroke patients.	Denmark. Regional Center for NeuroRehabilitation at Naestved Hospital.	Authors did a translation from the full version Fugl-Meyer Assessment as above to a Danish version using an eight-step procedure.	1. Cross-cultural validity	1. Inadequate	
Italian version of Fugl-Meyer Asessment (FMA-Ita)						
Cecchi et al. [29]	10 stroke patients.	Florence, Italy. Rehabilitation unit of Fondazione Don Carlo Gnocchi, Scientific Institute.	A multidisciplinary expert group did a translation from the full version of Fugl-Meyer Assessment from the Gothenburg university as above into an Italian version of the Fugl-Meyer Assessment, following a standardized process of forward and backward translations, reviewed	1. Content validity 2. Cross-cultural validity	1. Doubtful 2. Inadequate	

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
			by an expert committee. After that, they did the clinical validation with a pilot study with 10 stroke subjects.			
Japanese version of the Fugl-Meyer Assessment (FMA-Jap)						
Amano et al. [30]	30 stroke patients.	Japan. Hyogo College of Medicine Hospital. In this study, they did a translation of the manual for the FMA and a psychometric properties analysis.	Authors did a translation from the full version Fugl-Meyer Assessment as above to a Japanese version.	1. Internal consistency 2. Cross-cultural validity 3. Reliability 4. Measurement error 5. Hypothesis testing 6. Responsiveness	1. Very good 2. Inadequate 3. Inadequate 4. Inadequate 5. Very good 6. Very good	1. Cronbach's alpha of the FMA all domains (number of items, 63) was 0.973. 2. Regarding the nonsimultaneous and simultaneous inter-rater reliability, ICC ranged from 0.809 to 0.983 (95% confidence interval [CI], 0.639-0.994; $p < 0.001$) and 0.991-0.999 (95% CI, 0.981-0.999; $p < 0.001$), respectively. Regarding the simultaneous intra-rater reliability, ICC ranged from 0.994 to 0.999 (95% CI, 0.988-1.000; $p < 0.001$). 3. The motor section of FMA had the highest values (ranging from 0.926 to 0.950), and the sensation section had the lowest values (ranging from 0.268 to 0.370). The highest correlation is with ARAT (0.950 for FMA-motor section and 0.947 for FMA total).
Korean version of the Fugl-Meyer Assessment (FMA-Kor)						
Kim et al. [31]	31 stroke patients.	Korea. Korean National Rehabilitation Research Institute. National Rehabilitation Center.	First, 3 translators and a translation committee did the translation from the full FMA version as above, to the Korean version. Second, they analyze the psychometric properties of the FMA-Kor.	1. Cross-cultural validity 2. Reliability 3. Hypothesis testing 4. Responsiveness	1. Inadequate 2. Inadequate 3. Very good 4. Very good	1. For intra-rater reliability, all domains had ICC > 0.72, except passive ROM for LE and joint pain for LE. The motor function of the UE had the highest ICC (0.977) and the passive ROM of the LE had the lowest ICC (0.512). For inter-rater reliability, all domains had ICC > 0.72, except passive ROM for LE. The tactile sensitivity for UE had the highest ICC (0.995) and the passive ROM for LE had the lowest ICC (0.498).

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
						2. FMA-Kor was significantly correlated with the MI, which had the highest correlation (r = 0.788). It was also significantly correlated with the MAS, with a lower correlation coefficient (r = 0.496). It was not significantly correlated with the BBS score (r = 0.22).
Urdu version of the Fugl-Meyer Assessment (FMA-Urd) Ikram et al. [32]	10 stroke patients for pilot testing and 50 chronic stroke patients to analyze psychometric properties.	Pakistan. A standardized stepwise forward-backward translation of FMA into Urdu was conducted with the help of bilingual translators, experts from Riphah International University, and experts familiarized with the full version FMA scale to ensure conceptual equivalency.	Authors did a translation from the full version Fugl-Meyer Assessment as above to an Urdu version.	1. Content validity 2. Structural validity 3. Internal consistency 4. Cross-cultural validity 5. Reliability 6. Hypothesis testing 7. Responsiveness	1. Doubtful 2. Inadequate 3. Very good 4. Inadequate 5. Inadequate 6. Very good 7. Very good	1. The subsequent factor analysis and scree plot analysis showed that total variance was mainly described by two components. Component 1 was associated with the upper extremity. Component 2 was associated with the lower extremity. 2. Cronbach alpha of the upper and lower extremity assessments varied from 0.80 to 0.87.5. 3. Weighted kappa coefficients for intra- and inter-rater reliability ranged from 0.75 to 0.99. 4. Urdu FMA showed a moderate correlation with FIM (r = 0.70) and a negative moderate correlation with mRS (r = −0.69) and NIHSS (r = −0.79).
Romanian version of the Fugl-Meyer Assessment (FMA-Rom) Onose et al. [33]	10 stroke patients for the pilot study.	Bucharest, Romania, Neuromuscular Clinic Division of the Teaching Emergency Hospital "Bagdasar-Arseni," in National Institute of Neurology and Neuro-Vascular Diseases.	Authors did a translation from the full version Fugl-Meyer Assessment as above to a Romanian version.	1. Content validity 2. Cross-cultural validity 3. Reliability 4. Hypothesis testing 5. Responsiveness	1. Doubtful 2. Inadequate 3. Inadequate 4. Inadequate 5. Inadequate	1. PA ≥ 40%. 2. The correlation values obtained were very good for BI and close to the limit for the mRS. FMA-BI r = 0.917. FMA-mRS r = −0.59.
Short Form of the Fugl-Meyer (37-item) (FMA-37)						

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
Hsieh et al. [34]	279 stroke patients.	Netherlands and Taiwan. National Taiwan University Hospital.	37-item Fugl-Meyer motor scale (26 UE items and 11 LE items), scored with a 3-point scale, ranging from 0-2, with a maximum score of 74 points.	1. Reliability 2. Criterion validity 3. Responsiveness	1. Inadequate 2. Very good 3. Very good	1. Rasch reliabilities for both S-FM were good (coefficients ≥ 0.92). 2. Both subscales of the S-FM and FM at 3 time points had moderate predictive validity with the comprehensive activities of daily living function (r = 0.49-0.59 for the S-FM, and 0.48 -0.53 for the FM).
Chen et al. [26]	301 stroke patients at 14 d after stroke assessment; 276 subjects at 30 d post stroke; 226 at 90 d; and 213 participants evaluated at 180 d post stroke.	Taiwan. Data were obtained from a previous study. Patients were assessed with the 50-item Fugl-Meyer motor scale four times, at 14, 30, 90, and 180 days after stroke onset. All the assessments were administered by a trained occupational therapist.	Fugl-Meyer 37 items as above.	1. Structural validity 2. Criterion validity 3. Responsiveness	1. Doubtful 2. Very good 3. Very good	1. Latent parallel analysis and ordinal factor analysis indicated that all W/H UE-FM items represent a single unidimensional construct, wrist, and hand motor ability.2. FM-50 vs FM-37 was 0.61-0.80. FM-50 vs FM-12 was 0.80-0.00. FM-37 vs FM-12 was 0.01-0.03.
Lin et al. [14]	267 stroke patients.	Taiwan. Patients from secondary data from other studies.	Fugl-Meyer 37 items as above.	1. Structural validity 2. Reliability 3. Measurement error 4. Criterion validity 5. Hypotheses testing 6. Responsiveness	1. Doubtful 2. Inadequate 3. Inadequate 4. Very good 5. Doubtful 6. Doubtful	1. Unidimensional structure. 2. ICC for UE-FM-37 was 0.94. 3. Pearson r for FM-37 was 0,93. 4. Pearson r correlation with BI for FM-37 was 0.73. Pearson r correlation with PASS for FM-37 was 0.75.
Short Form of the Fugl-Meyer (12-item) (FMA-12)						
Hsieh et al. [34]	279 stroke patients.	Netherlands and Taiwan. National Taiwan University Hospital.	FMA-Short 12 items version has 12 items (shoulder flexion 90-180°, grasp [adduct thumb], elbow 90° pronation/supination, elbow 90° wrist flexion/extension, elbow extension, and shoulder elevation) scored on a 3-point scale from 0-2 points with a maximum score of 24 points.	1. Reliability 2. Criterion validity 3. Responsiveness	1. Inadequate 2. Very good 3. Very good	1. Rasch reliabilities for both S-FM were good (coefficients ≥ 0.92). 2. Both subscales of the S-FM and FM at 3 time points had moderate predictive validity with the comprehensive activities of daily living function (r = 0.49-0.59 for the S-FM, and 0.48 -0.53 for the FM).
Hsueh et al. [24]	50 stroke patients.	China. Rehabilitation Department of Kaohsiung Medical University Hospital.	Fugl-Meyer 12 items as above.	1. Reliability 2. Measurement error 3. Criterion validity 4. Hypothesis testing 5. Responsiveness	1. Doubtful 2. Doubtful 3. Very good 4. Very good 5. Very good	1. Test-retest reliabilities were high for the UE and LE subscales and total motor scores of the 4 motor scales (all ICCs ≥ 0.93 and lower limits of 95% CI ≥ 0.89). 2. The correlations were high (r = 0.91-0.99) among the 4

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
						clinical motor scales with their subscales at admission and discharge, indicating high concurrent validity. 3. The scores of the 4 motor scales with their subscales at admission were moderately correlated with the BI scores at discharge ($r \geq 0.66$), which indicates acceptable predictive validity.
Fu et al. [35]	51 stroke patients.	Taiwan. Three different Taiwanese hospitals.	Fugl-Meyer 12 items as above.	1. Hypothesis testing 2. Responsiveness	1. Very good 2. Very good	1. The correlations ($r \geq 0.57$) between the streamlined Fugl-Meyer Assessment and the Stroke Impact Scale hand function subscales indicate good concurrent validity both at baseline and post treatment. The streamlined Wolf Motor Function Test had fair-to-good concurrent validity ($r \geq 0.39$). The streamlined Fugl-Meyer Assessment had good predictive validity ($r = 0.68$).
Chen et al. [26]	301 stroke patients at 14 after stroke assessment; 276 subjects at 30 d post stroke; 226 at 90 d; and 213 participants evaluated at 180 d post stroke.	Taiwan. Data were obtained from a previous study. Patients were assessed with the 50-item Fugl-Meyer motor scale four times, at 14, 30, 90, and 180 days after stroke onset. All the assessments were administered by a trained occupational therapist.	Fugl-Meyer 12 items as above.	1. Structural validity 2. Criterion validity 3. Responsiveness	1. Doubtful 2. Very good 3. Very good	1. Latent parallel analysis and ordinal factor analysis indicated that all W/H UE-FM items represent a single unidimensional construct, wrist, and hand motor ability. 2. FM-50 vs FM-37 was 0.61-0.80. FM-50 vs FM-12 was 0.80-0.00. FM-37 vs FM-12 was 0.01-0.03.
Lin et al. [14]	267 stroke patients.	Taiwan. Patients from secondary data from other studies.	Fugl-Meyer 12 items as above.	1. Structural validity 2. Reliability 3. Measurement error 4. Criterion validity 5. Hypotheses testing 6. Responsiveness	1. Doubtful 2. Inadequate 3. Inadequate 4. Very good 5. Doubtful 6. Doubtful	1. Unidimensional structure. 2. ICC for UE-FM-12 was 0.93. 4. 3. Pearson r for FM-12 was 0935. 4. Pearson r correlation with BI for FM-12 was 0.73. Pearson *r* correlation with PASS for FM-12 was 0.76.

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
Shortened Fugl-Meyer Assessment Upper Extremity (6 items) (FMA-6)						
Amano et al. [36]	30 stroke patients.	Nishinomiya, Japan. Hyogo College of Medicine College Hospital. Sample video evaluations recorded during a previous study.	FMA shortened upper-extremity version has 6 items scored on a 3-point scale from 0-2 points with a maximum score of 12 points.	1. Internal consistency 2. Reliability 3. Criterion validity 4. Hypothesis testing 5. Responsiveness	1. Very good 2. Inadequate 3. Very good 4. Very good 5. Very good	1. Cronbach's alpha for the shortened (6-item) FMA was 0.924 for each rater. 2. Inter-rater reliability of the shortened FMA total score, the ICC was 0.994 (95% confidence interval [CI], 0.988-0.997; $p < 0.001$). 3. Correlation between FMA-Shortened Version and FMA-UE was $r = 0.97$. 4. Correlation of the FMA-Shortened Version with ARAT was 0.94; with BBT was 0.92; with MAL was 0.91 for both subscales (amount of use and quality of movement).
Machine-Learning Fugl-Meyer (10 items) (FMA-ML-10)						
Lin et al. [14]	267 stroke patients.	Taiwan. Patients from secondary data from other studies.	In this article, the authors develop a version of Fugl-Meyer Assessment based on machine learning, (FM-ML) that consists of 10 items from the original scale, scored with a 3-point scale, ranging from 0-2.	1. Structural validity 2. Reliability 3. Measurement error 4. Criterion validity 5. Hypotheses testing 6. Responsiveness	1. Doubtful 2. Inadequate 3. Inadequate 4. Very good 5. Doubtful 6. Doubtful	1. Unidimensional structure. 2. ICC for UE-FM-ML was 0.92.3. Pearson r for FM-ML was 0.99. 4. Pearson r correlation with BI for FM-ML was 0.67. Pearson r correlation with PASS for FM-ML was 0.68.
Tanslation and adaptation of the FMA administration manual into Brazilian Portuguese						
Michaelsen et al. [37]	18 stroke patients.	Brazil. In total, 10 patients took part in the first part of the study (translation and adaptation of the manual and assess the inter-rater reliability), and 8 patients took part in the second part of the study (assess the inter-rater of the test).	The French-Canadian version of the manual of the FMA was translated into Brazilian by a physical therapist.	1. Reliability	1. Inadequate	1. Part 1: The ICCs between the two raters for the total score and upper limb score were 0.99 ($p < 0.001$) for both and 0.88 ($p = 0.006$) for the lower limbs. Part 2: ICC between two examiners for the total motor function score was 0.98 ($p < 0.001$). Comparison between the ICCs from parts

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
						1 and 2 for the items of tactile sensitivity, movement sense, and ROM shows that the reliability became excellent after the adaptation of the manual. The reliability of the joint pain assessment for the upper limbs was moderate, and for the lower limbs was low.
Upper Extremity subscale of the Fugl-Meyer Assessment (FMA-UE)						
De Weerdt et al. [38]	53 stroke patients.	Nottingham, United Kingdom.	Upper-extremity Fugl-Meyer Assessment (FMA-UE) examines reflex activity, voluntary movements within partially out and independent of synergies. Includes 33 items divided into 4 subscales: shoulder/elbow (A, 18 items), wrist (B, 5 items), hand (C, 7 items), and coordination/speed (D, 3 items). Each item is scored on an ordinal 3-point scale ranging from 0-2. The total score ranges from 0-66.	1. Hypothesis testing 2. Responsiveness	1. Very good 2. Very good	1. The Spearman rank correlation between FM and ARAT was 0.91 at 2 weeks and 0.94 at 8 weeks. These correlation coefficients are significant at the 0.01 level of significance.
Van der Lee et al. [39]	22 chronic stroke patients.	Netherlands. People with chronic stroke were admitted in this randomized clinical trial if they met the criteria. An informed consent was given and two baseline measurements were performed before the intervention.	Upper-extremity Fugl-Meyer Assessment as above.	1. Reliability 2. Hypothesis testing 3. Responsiveness	1. Inadequate 2. Very good 3. Very good	1. The difference between two measurements has to be 10.0% for the FMA. 2. Taking the mean improvement (1.7 points on the ARA test) during the reference intervention into account, the use of the baseline period (with a mean improvement of 0.3 points on the ARA test) to define stability was considered to be more valid.
Rabadi et al. [40]	104 acute stroke patients.	New York, United States. Burke Rehabilitation Hospital.	Upper-extremity Fugl-Meyer Assessment as above.	1. Hypothesis testing 2. Responsiveness	1. Very good 2. Very good	1. The 2 upper-limb motor scales (ARAT, FMA-motor score) correlated highly with one another, both on admission ($r = 0.77$, $p < 0.001$) and on discharge ($r = 0.87$, $p < 0.001$).
Woodbury et al. [41]	512 subacute stroke patients.	Approved by the University of Florida Institutional Review Board. In total, 459 subjects enrolled in the Kansas City Stroke Study. Researchers from this study, the authors	Upper-extremity Fugl-Meyer Assessment as above.	1. Structural validity 2. Internal consistency	1. Very good 2. Very good	1. The PCA retained 4 factors having eigenvalues greater than 1.0. A single component explained 68% of the variance in the data. Three other factors explained an

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
		used secondary analysis of pooled data from 2 existing datasets: a randomized therapeutic exercise clinical trial and a cohort longitudinal study of stroke recovery.				additional 5%, 4%, and 3% of the variance. 2. The Pearson reliability index, analogous to the coefficient, was 0.96.
Hsieh et al. [42]	57 stroke patients.	Taiwan. Randomized controlled trial to investigate the effectiveness of distributed constraint-induced movement therapy and bilateral arm motor training. Subjects were recruited in 3 medical centers if they met the criteria.	Upper-extremity Fugl-Meyer Assessment as above.	1. Hypothesis testing 2. Responsiveness	1. Very good 2. Very good	1. Correlations between the FMA and other outcome measures at pre treatment were relatively high (Spearman 0.71-0.76). At post treatment, FMA had moderate-to-good correlations with the other outcome measures (Spearman 0.51-0.74).
Lin et al. [43]	53 stroke patients started the study. About 11 of them had recurrent strokes, and 6 patients were lost at follow-up. A total of 35 stroke patients completed all the assessments.	Taiwan. People with stroke who were admitted consecutively to the Department of Neurology at Kaohsiung Medical University Hospital from September 1, 2006-August 31, 2007; and who met the inclusion and exclusion criteria.	Upper-extremity Fugl-Meyer Assessment as above.	1. Reliability 2. Measurement error 3. Hypothesis testing 4. Responsiveness	1. Inadequate 2. Inadequate 3. Very good 4. Very good	1. Inter-rater reliability for the 4 measures was high (ICC ≥ 0.92, 95% CI; FM = 0.96). The test-retest reliability of the 4 measures was high (ICC ≥ 0.97, 95% CI; FM = 0.99). 2. The correlation of FM with the other 3 measures was high (Spearman p = 0.85-0.86).
Sullivan et al. [44]	15 stroke patients.	Los Angeles, California, United States. A total of 17 assessments were done, because 2 subjects were measured by 2 different therapists. In this article, there is a comparison between an expert rater and trained therapist raters.	Upper-extremity Fugl-Meyer Assessment as above.	1. Reliability	1. Inadequate	1. Intra-rater reliability for the ER was high for the motor (total motor, 0.99; UE, 0.95; LE, 0.99) and sensory assessments (total sensory, 0.96). Inter-rater agreement between expert and therapist raters was high for the total motor score (ICC, 0.98; 95% CI, 0.93-0.99) and UE motor subscore (ICC, 0.99; 95% CI, 0.97-1.0) and within the moderate-to-high range for the LE motor subscore (ICC, 0.91; 95% CI, 0.69-0.97). Inter-rater agreement was high for the total sensory score (ICC, 0.93; 95% CI, 0.83- 0.98).
Wei et al. [45]	27 stroke patients.	Hong Kong. Hong Kong Polytechnic University.	Upper-extremity Fugl-Meyer Assessment as above.	1. Criterion validity 2. Hypothesis testing 3. Responsiveness	1. Very good 2. Very good 3. Very good	1. Correlation between FMA and FMA (SE) was 0.91 and 0.94; and with FMA (WH) was 0.92 and 0.94; scores at pre training and post training respectively. Correlation between FMA (SE) and FMA (WH) was 0.69 and 0.74.

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
						2. Correlation of FMA was high with MSS and ARAT (0.93 for both at pre treatment and 0.91 and 0.92 for post treatment respectively. With MSS, (SE) was 0.89 at pre treatment and 0.86 at post treatment; with MMSS (WH), the correlation was 0.9 for both pre- and post training. For MAS (E), the correlation was nonsignificant at pre training, but 0.62 at post training. With MAS (W), the correlation was 0.42 at pre training and nonsignificant for post training. With FIM, the correlation was 0.49 at pre treatment, and 0.56 at post treatment.
Kim et al. [46]	50 stroke patients for reliability and validity study, and 16 stroke patients for responsiveness study.	Three rehabilitation hospitals in South Korea.	Upper-extremity Fugl-Meyer Assessment as above.	1. Reliability 2. Measurement error 3. Criterion validity 4. Hypothesis testing 5. Responsiveness	1. Doubtful 2. Doubtful 3. Very good 4. Very good 5. Very good	1. For inter-rater reliability of the motor function, the ICC ranged from 0.930 to 1.000. For test-retest reliability with an interval of 2 weeks of the motor function, the ICC was 0.834- 0.972. 2. FMA-UE compared with the FMA-LE was 0.723 ($p < 0.05$). 3. The relationship between the upper-extremity motor function of the FMA and Jebsen-Taylor hand function was 0.757 ($p < 0.05$), and the relationship between the lower extremity motor function of the FMA and MAS was 0.725 ($p < 0.05$).
Kim et al. [47]	41 stroke patients.	Seoul, South Korea. Seoul National University Bundang Hospital institutional.	Upper-extremity Fugl-Meyer Assessment as above.	1. Criterion validity 2. Hypothesis testing 3. Responsiveness	1. Very good 2. Very good 3. Very good	1. Correlation between FMA-13 items using Kinect and real FMA was 0.873 ($p < 0.0001$).2. Correlation between log (jerky score) and Brunnstrom arm stage (3 to 6) in the hemiplegic upper extremity ($n = 27$) (Spearman correlation coefficient = −0.387 ($p = 0.046$)).

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
Amano et al. [36]	30 stroke patients.	Nishinomiya, Japan. Hyogo College of Medicine College Hospital. Sample video evaluations recorded during a previous study.	Upper-extremity Fugl-Meyer Assessment as above.	1. Internal consistency 2. Reliability 3. Criterion validity 4. Hypothesis testing 5. Responsiveness	1. Very good 2. Inadequate 3. Very good 4. Very good 5. Very good	1. Cronbach's alpha for the shortened (6-item) FMA was 0.924 for each rater. 2. Inter-rater reliability of the shortened FMA total score, the ICC was 0.994 (95% CI, 0.988-0.997; $p < 0.001$).3. Correlation between FMA-Shortened Version and FMA-UE was $r = 0.97$. 4. Correlation of the FMA-Shortened Version with ARAT was 0.94; with BBT was 0.92; with MAL was 0.91 for both subscales (amount of use and quality of movement).
Hernández et al. [48]	60 stroke patients.	Bogota, Colombia. In total, 60 consecutive patients including during 17 months at Central Military hospital of Colombia.	Upper-extremity Fugl-Meyer Assessment as above.	1. Reliability	1. Inadequate	1. For intra-rater reliability, the PA between test occasion 1 and 2 within each rater was above 79% for all tested items. For inter-rater reliability, The PA was above 90% for all items between the raters.
Tauchi et al. [15] Danish version of the upper-extremity subscale of the Fugl-Meyer Assessment (FMA-UE-Dan)	268 stroke patients.	Japan. Patients from 22 Japanese hospitals.	Upper-extremity Fugl-Meyer Assessment as above.	1. Structural validity	1. Very good	1. As a result of the one-factor analysis using CFA.
Lundquist et al. [49] Italian version of Fugl-Meyer Asessment-Upper Extremity (FMA-UE-Ita)	50 stroke patients.	Denmark. Skive Regional Hospital. Cross-cultural translation of the upper-extremity subscale of Fugl-Meyer Assessment into Danish, and evaluation of its psychometric properties, using a sample of 50 patients.	Authors did a translation from the Upper-Extremity Fugl-Meyer Assessment as above to a Danish version.	1. Cross-cultural validity 2. Reliability 3. Hypothesis testing 4. Responsiveness	1. Inadequate 2. Doubtful 3. Doubtful 4. Doubtful	1. Inter-rater reliability was high (ICC 0.95, 95% CI, 0.93–0.98). 3. Spearman's rank correlation between the FMA-UE and the MAS-UE was 0.95 ($p < 0.001$) at baseline and 0.94 ($p < 0.001$) at follow-up.
Hochleitner et al. [50]	60 subacute stroke patients.	Florence, Italy. Neurological Rehabilitation Unit of Fondazione Don Carlo Gnocchi, Scientific Institute.	Authors did a translation from the Upper-Extremity Fugl-Meyer Assessment as above to an Italian version.	1. Reliability	1. Inadequate	1. Inter-rater reliability at the item level, the percentage of agreement was satisfactory or excellent for all items (PA >70). Intra-rater reliability at the item level, the percentage of agreement was satisfactory or excellent (PA >70%).

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
Romanian version of the Upper-Extremity Fugl-Meyer Assessment (FMA-UE-Rom)						
Roman et al. [51]	64 stroke patients.	Romania. Clinical Hospital of Psychiatry and Neurology in Brasov, Romania.	Authors did a translation from the Upper-Extremity Fugl-Meyer Assessment as above to a Romanian version.	1. Structural validity 2. Internal consistency 3. Cross-cultural validity 4. Reliability 5. Hypothesis testing 6. Responsiveness	1. Inadequate 2. Very good 3. Inadequate 4. Inadequate 5. Very good 6. Very good	1. After performing the exploratory factor analysis, a single factor was extracted. 2. Cronbach Alpha = 0.981. 3. ICC of average measures = 0.992. 4. Correlation with FIM was 0.789 ($p < 0.001$), and with MRS was −0.787 ($p < 0.001$), and SRM was 1.117.
Fugl-Meyer Assessment Upper-Extremity (13 items) evaluation using Depth-Sensing Camera (Kinect) (FMA-13-Kinect)						
Kim et al. [47]	41 stroke patients.	Seoul, South Korea. Seoul National University Bundang Hospital institutional.	In this study, researchers compare the traditional observation way to evaluate FMA and the Kinect depth-camera motion data recording to evaluate FMA. - FMA-UE 33 items - FMA-UE 13 items (flexor synergy, extensor synergy, volitional motion mixing dynamic flexor and extensor synergy, and volitional movement with little or no synergy dependence).	1. Criterion validity 2. Hypothesis testing 3. Responsiveness	1. Very good 2. Very good 3. Very good	1. Correlation between FMA-13 items using Kinect and real FMA-13 items was 0.873 ($p < 0.0001$). 2. Correlation between log (jerky score) and Brunnstrom arm stage (3-6) in the hemiplegic upper extremity ($n = 27$) (Spearman correlation coefficient = −0.387 ($p = 0.046$)).
Remote evaluation of Upper Extremities with Fugl-Meyer Assessment (FMA-Remote)						
Amano et al. [52]	30 stroke patients.	Japan. Hyogo College of Medicine Hospital.	Upper-extremity Fugl-Meyer Assessment as above assessed through a remote evaluation.	1. Reliability	1. Doubtful	1. ICC = 0.998
Shoulder and Elbow Fugl-						

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
Meyer (FMA-UE-SE)						
Wei et al. [45]	27 stroke patients.	Hong Kong.Hong Kong Polytechnic University.	Shoulder and Elbow Upper Extremity Fugl-Meyer Assessment (wh-UE-FM) has the most proximal UE-FM items. Each item is scored on an ordinal 3-point scale ranging from 0-2.	1. Criterion validity 2. Hypothesis testing 3. Responsiveness	1. Very good 2. Very good 3. Very good	1. Correlation between FMA and FMA (SE) was 0.91 and 0.94; and with FMA (WH) was 0.92 and 0.94; scores at pre training and post training respectively. Correlation between FMA (SE) and FMA (WH) was 0.69 and 0.74. 2. Correlation of FMA with MSS was 0.79; with MSS (SE) was 0.79; with MMSS (WH), the correlation was 0.75. With ARAT was 0.85. For MAS (E), the correlation was 0.54. With MAS (W), the correlation was 0.4; and with FIM, the correlation was 0.52.
See et al. [25]	31 stroke patients.	Patient's data from 2 robotic clinical stroke trials of arm motor therapy from the United States, Canada, and Germany.	Shoulder-Elbow Upper Extremity subscale of Fugl-Meyer Assessment as above.	1. Reliability 2. Hypothesis testing 3. Responsiveness	1. Doubtful 2. Very good 3. Very good	1. For intra-rater reliability, ICC was 0.99 for FMA-Total score, for FMA-Proximal subscore, and for FMA-w/h subscore. For inter-rater reliability, ICC was 0.99 in FMA-Total score, and 0.98 for both FMA-Proximal and FMA-w/h subscore.2. The correlation of the FMA-SE score with ARAT was 0.8; with pinch force was 0.85; with grip force was 0.73; with BBT was 0.79; with NHPT was 0.64; and with SIS was 0.79.
Wrist and Hand Fugl-Meyer (FMA-UE-w/h)						
Wei et al. [45]	27 stroke patients.	Hong Kong. Hong Kong Polytechnic University.	Wrist and Hand Upper Extremity Fugl-Meyer Assessment (wh-UE-FM) has the 12 most distal UE-FM items. Each item is scored on an ordinal 3-point scale ranging from 0 -2, with a total possible score of 24.	1. Criterion validity 2. Hypothesis testing 3. Responsiveness	1. Very good 2. Very good 3. Very good	1. Correlation between FMA and FMA (SE) was 0.91 and 0.94; and with FMA (WH) was 0.92 and 0.94; scores at pre training and post training respectively. Correlation between FMA (SE) and FMA (WH) was 0.69 and 0.74. 2. Correlation of FMA with MSS was 0.91; with MSS (SE) was 0.82; with MMSS (WH), the correlation was 0.95. With ARAT was 0.88. For MAS (E), the correlation was 0.56. With MAS (W), the correlation was

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
						nonsignificant; and with FIM, the correlation was 0.49.
Page et al. [53]	29 stroke patients.	Midwestern United States.	Wrist and Hand Upper Extremity Fugl-Meyer Assessment as above.	1. Internal consistency 2. Reliability 3. Hypothesis testing 4. Responsiveness	1. Very good 2. Adequate 3. Very good 4. Very good	1. Cronbach's alpha for w/h-UE-FM was 0.90 for pretest 1 and 0.88 for pretest 2. 2. ICC for w/h-UE-FM was 0.97. 3. Correlation score between w/h-UE-FM and ARAT was 0.72 ($p < 0.001$).
See et al. [25]	31 stroke patients.	Patient's data from 2 robotic clinical stroke trials of arm motor therapy from the United States, Canada, and Germany.	Wrist and Hand Upper Extremity Fugl-Meyer Assessment as above.	1. Reliability 2. Hypothesis testing 3. Responsiveness	1. Doubtful 2. Very good 3. Very good	1. For intra-rater reliability, ICC was 0.99 for FMA-Total score, for FMA-Proximal subscore. and for FMA-w/h subscore. For inter-rater reliability, ICC was 0.99 in FMA-Total score, and 0.98 for both FMA-Proximal and FMA-w/h subscore. 2. The correlation of the FMA-w/h score with ARAT was 0.89; with pinch force was 0.85; with grip force was 0.73; with BBT was 0.88; with NHPT was 0.8; and with SIS was 0.88.
Page et al. [54]	32 stroke patients.	Midwestern United States. National Center for Complementary and Alternative Medicine.	Wrist and Hand Upper Extremity Fugl-Meyer Assessment as above.	1. Internal consistency 2. Reliability 3. Hypothesis testing 4. Responsiveness	1. Very good 2. Doubtful 3. Very good 4. Very good	1. Cronbach's alpha for w/h-UE-FM was 0.82 at pretest 1. 2. The ICCs for the w/h UE-FM were 0.95. 3. Spearman *r* for the correlation between ARAT and w/h-UE-FM was 0.74 ($p < 0.001$) at pretest 1.
Persch et al. [55]	150 stroke patients.	Patients from secondary analysis of data obtained during outpatient, randomized controlled trials approved by the Human Research Protection Program at The Ohio State University .Participants were recruited for the trials from across the Midwestern United States.	Wrist and Hand Upper Extremity Fugl-Meyer Assessment as above.	1. Structural validity 2. Measurement error	1. Adequate 2. Doubtful	1. EFA resulted in a 12-item analysis with values between 0.477 and 0.861.

(*continued on next page*)

Table 2 (*continued*)

Study (author and year)	Population	Setting	Instrument description	Measurement properties	COSMIN score	Measurement values
Sensory subscale of the Fugl-Meyer Assessment (FMA-S)						
Lin et al. [23]	176 stroke patients.	Taiwan. The study sample was recruited from the registry of the Quality of Life after Stroke Study in Taiwan for the period from 1 December 1999-7 December 2001. The Quality of Life after Stroke Study is a prospective cohort study of stroke patients admitted to National Taiwan University Hospital.	Sensory Fugl-Meyer Assessment (FMA-S) contains 12 three-point items scored from 0-2. Four items for light touch and eight for position sense. The total score ranges from 0-24.	1. Internal consistency 2. Reliability 3. Criterion validity 4. Hypothesis testing 5. Responsiveness	1. Very good 2. Inadequate 3. Very good 4. Very good 5. Very good	1. The Cronbach's alphas of the FMA-S at four time points after stroke ranged from 0.94 to 0.98, indicating excellent internal consistency. 2. The inter-rater agreement of the total score of the FMA-S was excellent, with ICC (95% CI) of 0.93 (0.85-0.96). 3. The FMA-S scores were weakly to moderately correlated FMA-M scores (Spearman's rho between 0.31 and 0.44). 4. The FMA-S scores were weakly to moderately correlated with the BI (Spearman's rho ranging from 0.38 to 0.53).
Sullivan et al. [44]	15 stroke patients.	Los Angeles, California. A total of 17 assessments were done, because 2 subjects were measured by 2 different therapists. In this article, there is a comparison between an expert rater and trained therapist raters.	Sensory Fugl-Meyer Assessment as above.	1. Reliability	1. Inadequate	1. Intra-rater reliability for the ER was high for the motor (total motor, 0.99; UE, 0.95; LE, 0.99) and sensory assessments (total sensory, 0.96). Inter-rater agreement between expert and therapist raters was high for the total motor score (ICC, 0.98; 95% CI, 0.93-0.99) and UE motor subscore (ICC, 0.99; 95% CI, 0.97-1.0) and within the moderate-to-high range for the LE motor subscore (ICC, 0.91; 95% CI, 0.69-0.97). Inter-rater agreement was high for the total sensory score (ICC, 0.93; 95% CI, 0.83- 0.98).

ARAT = Action Research Arm Test; BBS = Berg Balance Scale; BBT = Box & Block Test; BI = Barthel Index; CI = Confidence Interval; CFA = Confirmatory Factor Analysis; EFA = Exploratory Factor Analysis; ER = Expert Rater; FIM = Functional Independence Measure; FM/FMA = Fugl-Meyer Assessment; FM-ML = Fugl-Meyer Machine-Learning; FMA-Col = Fugl-Meyer Assessment Colombian; FMA-Dan = Fugl-Meyer Assessment Danish; FMA-Ita = Fugl-Meyer Assessment Italian; FMA-Jap = Fugl-Meyer Assessment Japanese; FMA-Kor = Fugl-Meyer Assessment Korean; FMA-LE = Fugl-Meyer Assessment Lower Extremity; FMA-LE-Ita = Fugl-Meyer Assessment Lower Extremity Italian; FMA-M = Fugl-Meyer Assessment Motor; FMA-S = Fugl-Meyer Assessment Sensory; FMA-SE = Fugl-Meyer Assessment Shoulder/Elbow; FMA-UE = Fugl-Meyer Assessment Upper Extremity; FMA-UE-Dan = Fugl-Meyer Assessment Upper Extremity Danish; FMA-UE-Ita = Fugl-Meyer Assessment Upper Extremity Italian; FMA-UE-Rom = Fugl-Meyer Assessment Upper Extremity Romanian; FMA-UE-SE/FMA-SE = Fugl-Meyer Assessment Shoulder-Elbow; FMA-UE-w/h/FMA-W/H = Fugl-Meyer Assessment Upper Extremity Wrist/Hand; FMA-Urd = Fugl-Meyer Assessment Urdu; ICC = Intraclass correlation coefficient; LE = Lower Extremity; LE-SFM = Lower Extremity Short-Fugl-Meyer; MAL = Motor Activity Log; MAS = Motor Assessment Scale; MAS-UE = Motor Assessment Scale-Upper Extremity; MRS = Modified Rankin Scale; MSS = Motor Status Scale; NHPT = Nine-Hole Peg Test; NIHSS = National Institutes of Health Stroke Scale; PA = Percentage of Agreement; PASS = Postural Assessment Scale for Stroke Patients; PCA = Principal Components Analysis; ROM = Range of Movement; S-FM = Short-Fugl-Meyer; SIS = Stroke Impact Scale; UE = Upper Extremity; UE-FM-ML = Upper-Extremity Fugl-Meyer Machine-Learning; UE-SFM = Upper-Extremity Short-Fugl-Meyer.
Source: Self-made.

Table 3
Methodological quality of the studies and quality of results reported per measurement property, instrument, and study using COSMIN and GRADE methodology

		Content validity		Structural validity		Internal consistency		Cross-cultural validity/ measurement invariance		Measurement error		Reliability		Criterion validity		Hypotheses testing		Responsiveness	
Instrument	**Articles**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**
Fugl-Meyer Assessment (Full version, 50 items) (FMA)	Sanford et al. [21]									D	?	D	+						
	Malouin et al. [22]	D	-													V	?	V	?
	Lin et al. [23]					V	+					I	+	V	-	V	?	V	?
	Hsueh et al. [24]									D	-	D	+	V	+	V	?	V	?
	See et al. [25]											D	+			V	+	V	+
	Chen et al. [26]			D	?									V	-			V	-
	Lin et al. [14]			D	?					I	?	I	+			D	?	D	?
Colombian Spanish version of the Fugl-Meyer Assessment (FMA-Col)	Barbosa et al. [27]	D	-					I	-										
Danish version of the Fugl-Meyer Assessment (FMA-Dan)	Busk et al. [28]							I	-										
Italian version of Fugl-Meyer Asesment (FMA-Ita)	Cecchi et al. [29]	D	-					I	-										
Japanese version of the Fugl-Meyer Assessment (FMA-Jap)	Amano et al. [30]					V	+	I	-	I	?	I	+			V	?	V	?
Korean version of the Fugl-Meyer Assessment (FMA-Kor)	Kim et al. [31]							I	-			I	+			V	?	V	?
Urdu version of the Fugl-Meyer Assessment (FMA-Urd)	Ikram et al. [32]	D	-	I	-	V	-	I	-			I	+			V	?	V	?
Romanian version of the Fugl-Meyer Assessment (FMA-Rom)	Onose et al. [33]	D	?					I	?			I	?			I	?	I	?
Short Form of the Fugl-Meyer (37-item) (FMA-37)	Hsieh et al. [34]											I	?	V	-			V	-
	Chen et al. [26]			D	?									V	-			V	-
	Lin et al. [14]			D	?					I	?	I	+	V	+	D	?	D	?
Short Form of the Fugl-Meyer (12-item) (FMA-12)	Hsieh et al. [34]											I	?	V	-			V	-
	Hsueh et al. [24]									D	-	D	+	V	+	V	?	V	?
	Fu et al. [35]															V	?	V	?
	Chen et al. [26]			D	?									V	-			V	-
	Lin et al. [14]			D	?					I	?	I	+	V	+	D	?	D	?
Shortened Fugl-Meyer Assessment Upper Extremity (6-item) (FMA-6)	Amano et al. [36]					V	+					I	+	V	+	V	?	V	?
Machine-Learning Fugl-Meyer (10 items) (FMA-ML-10)	Lin et al. [14]			D	?					I	?	I	+	V	+	D	?	D	?
Tanslation and adaptation of the FMA administration manual into Brazilian Portuguese	Michaelsen et al. [37]											I	+						

(*continued on next page*)

Table 3 (*continued*)

		Content validity		Structural validity		Internal consistency		Cross-cultural validity/ measurement invariance		Measurement error		Reliability		Criterion validity		Hypotheses testing		Responsiveness	
Instrument	**Articles**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**	**M**	**Q**
Upper Extremity subscale of the Fugl-Meyer Assessment (FMA-UE)	De Weerdt et al. [38]															V	?	V	?
	Van der Lee et al. [39]											I	?			V	?	V	?
	Rabadi et al. [40]															V	?	V	?
	Woodbury et al. [41]			V	-	V	?												
	Hsieh et al. [42]															V	?	V	?
	Lin et al. [43]									I	?	I	+			V	?	V	?
	Sullivan et al. [44]											I	+						
	Wei et al. [45]													V	+	V	?	V	?
	Kim et al. [46]									D	?	D	+	V	+	V	?	V	?
	Kim et al. [47]													V	+	V	?	V	?
	Hernández et al. [48]											I	?						
	Amano et al. [36]					V	+					I	+	V	+	V	?	V	?
	Tauchi et al. [15]			V	+														
Danish version of the upper-extremity subscale of the Fugl-Meyer Assessment (FMA-UE-Dan)	Lundquist et al. [49]							I	?			D	+			D	+	D	+
Italian version of Fugl-Meyer Asessment-Upper Extremity (FMA-UE-Ita)	Hochleitner et al. [50]											I	?						
Romanian version of the Upper-Extremity Fugl-Meyer Assessment (FMA-UE-Rom)	Roman et al. [51]			I	-	V	+	I	?			I	+			V	?	V	?
Fugl-Meyer Assessment Upper-Extremity (13 Items) evaluation using Depth-Sensing Camera (Kinect) (FMA-13-Kinect)	Kim et al. [47]													V	+	V	?	V	?
Remote evaluation of Upper Extremities with Fugl-Meyer Assessment (FMA-Remote)	Amano et al. [52]											D	+						
Shoulder and Elbow Fugl-Meyer (FMA-UE-SE)	Wei et al. [45]													V	+	V	?	V	?
	See et al. [25]											D	+			V	+	V	+
Wrist and Hand Fugl-Meyer (FMA-UE-w/h)	Wei et al. [45]													V	+	V	?	V	?
	Page et al. [53]					V	+					A	+			V	?	V	?
	See et al. [25]											D	+			V	+	V	+
	Page et al. [54]					V	+					D	+			V	?	V	?
	Persch et al. [55]			A	?					D	?								
Sensory subscale of the Fugl-Meyer Assessment (FMA-S)	Lin et al. [23]					V	+					I	+	V	-	V	?	V	?
	Sullivan et al. [44]											I	+						

A = adequate; D = doubtful; Empty boxes = not reported; I = inadequate; M = methodological quality; Q = quality of the results; V = very good; + = positive rating; ? = indeterminate rating; - = negative rating.
Source: Self-made.

lack of content evaluation, and in the case of content validity, the relevance of the item in the study population.

Methodological quality of structural validity. In the case of structural validity because it was not calculated for each dimension separately or due to the lack of description of the percentage of management of missing values.

Methodological quality of cross-cultural validity. In cross-cultural validity psychometric property, the low methodological quality was due to the small sample of patients and the lack of comparison between the different groups after carrying out the cross-cultural analysis.

Methodological quality of reliability. In the case of reliability and the validity of criteria, this low methodological quality was due to the lack of description of the percentages and management of the missing values and to the poststroke phase, most of the studies were carried out in acute patients who were not neurologically stable at the time of assessment.

Methodological quality of hypothesis testing. Finally, the low methodological quality in the hypothesis testing is due to the fact that the hypotheses were not formulated a priori in the studies, so it was not possible to know what was expected to be found in those studies.

Translation versions

Fifteen versions (FMA-Col,[27] FMA-Dan,[28] the FMA-Ita,[29] FMA-Jap,[30] FMA-Kor,[31] FMA-Urd,[32] FMA-Rom,[33] FMA-6,[36] the FMA-ML-10,[14] the Translation and adaptation of the FMA administration manual into Brazilian Portuguese,[37] FMA-UE-Dan,[49] FMA-UE-Ita,[50] FMA-UE-Rom,[51] FMA-13-Kinect,[47] and the FMA-Remote)[52] were only evaluated in one study. Therefore, additional studies are required to evaluate the psychometric properties of these already-translated versions and to be able to understand these FMA variants in greater depth, so that they can be used more rigorously in clinical and research setting and to be able to take advantage of the benefits they present compared with other versions.

There were only two studies of the same FMA version, the FMA-UE-w/h, in which the methodological quality was rated as "adequate", for structural validity,[55] due to a factor analysis developed, a Rasch analysis, and a good sample size. The other study[53] was rated as "adequate" for reliability due to the stability of the patients, the appropriate time interval, and the similar test conditions.

Quality of evidence

Table 4 shows the quality of the evidence analyzed in the 36 articles, on the 22 versions of FMA, analyzed using the COSMIN methodology. The versions with the highest level of evidence were FMA-UE and FMA-S, with four "High" scores each; while the versions with the lowest level of evidence were FMA-Rom, with five "Very Low", and the FMA-Urd and FMA-UE-Dan versions with four "Very Low" scores each.

Quality of evidence of content validity. Content validity was present only in five studies of five different versions: Full version, 50 items, FMA,[22] FMA-Col,[27] FMA-Ita,[29] FMA-Rom,[33] and FMA-Urd.[32] For the five versions, Full version, 50 items, FMA,[22] FMA-Col,[27] FMA-Ita,[29] FMA-Rom,[33] and FMA-Urd,[32] the quality of the evidence was rated as "doubtful" due to the probability that the actual measurement property is far from the estimate of the measurement property, even though there is the possibility to make it substantially different. Furthermore, for four versions that evaluate content validity, the quality of results was rated as, "-"[22,27,29,32] and the other as "?",[33] indicating that relevance of the items, understandable, and comprehensiveness of the items of the different versions has not been evaluated.

Quality of evidence of structural validity. Structural validity was assessed in nine studies: Full version, 50 items, FMA,[14,23,26] FMA-Urd,[32] FMA-37,[14,26] FMA-12,[14,26] FMA-ML-10,[14] FMA-UE,[15,41] FMA-UE-Rom,[51] FMA-UE-w/h,[55] and FMA-S.[23] Only a single version FMA-UE[15,41] obtained a quality of evidence rated as "very good". For one study was scored as "adequate" (FMA-UE-w/h).[55] The rest of the versions and studies were scored as "doubtful" (most of them) and "inadequate" (FMA-Urd[32] and FMA-UE-Rom).[51]

Quality of evidence of internal consistency. Finally, for internal consistency, eight studies evaluated this psychometric property: Full version 50-item, FMA,[23] FMA-Jap,[30] FMA-Urd,[32] FMA-6,[36] FMA-UE,[36,41] FMA-UE-Rom,[51] FMA-UE-w/h,[53,54] and FMA-S.[23] All the versions and the studies evaluating this psychometric property have obtained a "very good" rating; Full version, 50-item, FMA,[23] FMA-Jap,[30] FMA-Urd,[32] FMA-6,[36] FMA-UE,[36,41] FMA-UE-Rom,[51] FMA-UE-w/h,[53,54] and FMA-S.[23] And all of them showed a "+" criteria rating, indicating that Cronbach's alpha is ≥ 0.70, except for FMA-Urd.[32]

Discussion

This is the first systematic review to identify the versions of the FMA to measure upper-extremity sensorimotor function in stroke patients and to evaluate the psychometric properties and methodological quality of the versions, based on the most up-to-date COSMIN methodology.[17,18] Twenty-two FMA versions designed to assess upper-extremity sensorimotor function in stroke patients were identified by including 36 studies.

Clinicians and researchers must consider different factors, such as the setting and the population, where the instrument will be used, its dimensions, the number of items, and the evidence shown in the evaluation of each psychometric property, to decide which instrument is the most appropriate to measure this construct. Additionally, the language and culture of the different versions must be considered when selecting which version to use and with which population and in which country, setting, and language.

A total of 11 translations have been included in this systematic review, showing heterogeneity among them, both in the population included and in the number of psychometric properties evaluated. However, they showed similar results between them, in terms of different psychometric properties, which indicate that the different translations have been carried out with good methods and that their final version can be used in the destination country, without changing the reliability or sensitivity of the test. The most complete translation is that of FMA-Urd,[32] evaluating seven psychometric properties, and with positive psychometric scores. There was no FMA version to measure upper-extremity sensorimotor function for which all psychometric properties were assessed.

FMA-UE the version with the best methodological quality because it was studied in multiple articles, but also analyzed in studies of fewer rigors. The studies with the best methodological quality scores, that is, those with the greatest number of "very good" scores, were Lin et al.[23] and Amano et al.,[36] with four each. Regarding the quality of the results, the versions with the most "+" were FMA-UE (n = 10) and FMA Full version (n = 9); while the versions with the highest number of negative ratings ("-") were FMA Full version and FMA-12, with 15 "-" each. However, FMA-UE, despite presenting more "+" and fewer "-", has been the version with the highest number of indeterminate rating ("?") (n = 23). The studies with the best quality of results were Amano et al.,[36] Lin et al.,[14] Roman et al.,[51] Page et al.,[53] and Page et al.[54]

Table 4
Quality of evidence reported per instrument using COSMIN methodology

Instrument	Content validity	Structural validity	Internal consistency	Cross-cultural validity/ measurement invariance	Measurement error	Reliability	Criterion validity	Hypotheses testing	Responsiveness
Fugl-Meyer Assessment (Full version, 50 items) (FMA)[14,21–26]	VL	M	H		VL	L	H	M	L
Colombian Spanish version of the Fugl-Meyer Assessment (FMA-Col) (50 items)[27]	VL			VL					
Danish version of the Fugl-Meyer Assessment (FMA-Dan) (50 items)[28]				VL					
Italian version of Fugl-Meyer Asessment (FMA-Ita) (50 items)[29]	VL			VL					
Japanese version of the Fugl-Meyer Assessment (FMA-Jap) (50 items)[30]			L	VL	VL	VL		L	L
Korean version of the Fugl-Meyer Assessment (FMA-Kor) (50 items)[31]				VL		VL		L	L
Urdu version of the Fugl-Meyer Assessment (FMA-Urd) (50 items)[32]	VL	VL	L	VL		VL		L	L
Romanian version of the Fugl-Meyer Assessment (FMA-Rom) (50 items)[33]	VL			VL		VL		VL	VL
Short Form of the Fugl-Meyer (37-item) (FMA-37)[14,26,34]		M			VL	L	H	L	M
Short Form of the Fugl-Meyer (12-item) (FMA-12)[14,24,26,34,35]		M			VL	VL	H	M	M
Shortened Fugl-Meyer Assessment Upper Extremity (6 items) (FMA-6)[36]			L			VL	L	L	L
Machine-Learning Fugl-Meyer (10 items) (FMA-ML-10)[14]		L			VL	VL	H	L	L
Translation and adaptation of the FMA administration manual into Brazilian Portuguese (50 items)[37]						VL			
Upper Extremity subscale of the Fugl-Meyer Assessment (FMA-UE) (33 items)[15,36,38–48]		H	L		VL	VL	H	H	H
Danish version of the upper-extremity subscale of the Fugl-Meyer Assessment (FMA-UE-Dan) (33 items)[49]				VL		VL		VL	VL
Italian version of Fugl-Meyer Asessment-Upper Extremity (FMA-UE-Ita) (33 items)[50]						VL			
Romanian version of the Upper-Extremity Fugl-Meyer Assessment (FMA-UE-Rom) (33 items)[51]		VL	L	VL		VL		L	L
Fugl-Meyer Assessment Upper-Extremity (13 items) evaluation using Depth-Sensing Camera (Kinect) (FMA-13-Kinect)[47]							L	L	L
Remote evaluation of Upper Extremities with Fugl-Meyer Assessment (FMA-Remote) (33 items)[52]						VL			
Shoulder and Elbow Fugl-Meyer (FMA-UE-SE) (18 items)[25,45]						VL	L	M	M
Wrist and Hand Fugl-Meyer (FMA-UE-w/h) (12 items)[25,45,53–55]		L	M		L	L	L	H	H
Sensory subscale of the Fugl-Meyer Assessment (FMA-S) (12 items)[23,44]			H			VL	H	H	H

H = high; L = low; M = moderate; QE = quality of evidence; VL = very Low; COSMIN, consensus-based standards for the selection of health status Measurement INstruments methodology.
Source: Self-made.

Regarding the quality of evidence, it was "low" or "very low" for most of the measurement properties analyzed in all versions included in this systematic review. Taking into account that content validity, structural validity, and internal consistency are considered the most important measurement properties.[11,18]

Full version, 50-item FMA and FMA-UE have been the two versions with the greatest volume of studies, with the most psychometric properties analyzed, and the most translated; in addition to being the two oldest versions of FMA. This makes them the most widespread versions of FMA used by clinicians and researchers to evaluate the sensorimotor function of the upper limb in poststroke people. This makes these two FMA [Fugl-Meyer Assessment (Full version, 50 items, FMA) and Upper-Extremity subscale of the Fugl-Meyer Assessment (FMA-UE)] the most widely validated and the most suggested for use in evaluating upper-extremity sensorimotor function in stroke patients both in clinical practice and in research. Highlighting the initial full version of the Fugl-Meyer Assessment (Full version, 50 items, FMA) as the most valid and recommended because, based on the results of the present research, it is the most used version with more psychometric properties analyzed and has demonstrated better scores. This makes this FMA version the most widely validated and recommended to be the best version and the "gold standard" to assess upper-extremity sensorimotor function in people with stroke.

This does not mean that the other 20 versions should not be used, as they are very useful for certain contexts and specific environments. It can be very useful to refer to Table 2 to select the version that best suits the construct to be assessed, depending on the time available, or when there is doubt between two versions, to facilitate the choice of one of them.

Another possibility offered by the versions found in this review is to be able to use a reduced version of FMA, which allows a quick and reliable assessment, and depending on whether the evaluator wishes to carry out a general assessment of FMA; or by categories of the scale; since all of them are supported by multiple studies and evaluated psychometric properties. In the event that a therapist has very limited time, it is recommended to use the FMA-12 version, because it provides a complete assessment, and it has been evaluated in five studies that, in total, included 948 people in which they evaluated the psychometric properties of this version and compared it with other versions and with other evaluation scales, with which it presents positive correlations. In addition, this version shows good reliability and sensitivity scores.

Throughout the different studies included in this review, a large sample size has been evaluated, with very diverse aspects, such as acute/chronic stage, mild and severe sequelae, evaluation of progress with initial evaluation, and follow-ups with periodic evaluations; and FMA has proven useful for the evaluation of the construct in the different situations that can occur with the variability of these aspects. Therefore, it is a valid and reliable scale to apply in the different situations that may occur in the poststroke population; although with greater preference for the use of Full version, 50-item FMA; and FMA-EU.

The main implication that this review may have is to offer a kind of list of the different versions of FMA, with information on its different characteristics such as the number of items, the parts it evaluates, the material needed, the language in which it is, the psychometric properties that have been evaluated for each of them, and their corresponding scores. This allows clinicians to select the scale that best suits their healthcare resource and needs, and to be able to quickly refer to the original article to resolve any type of doubt. It also allows a clinical therapist who hesitates between using two versions to quickly compare them by consulting the present study. At the same time, it allows researchers to have the gold standard version of FMA, and to be able to rely on this research to justify its use in future research; as well as consult the psychometric properties of any version of FMA, which will save a lot of time, and will allow a much more valid and justified use of said tool.

After carrying out this research, it is recommended that the gold standard version of FMA be compared in a similar way to that carried out in this research, if possible using the COSMIN methodology, with other evaluation scales that evaluate the same construct, to find the tool more valid and adjusted to assess the sensorimotor function of the upper limb in the stroke population. It is also recommended that the different versions of FMA be analyzed and compared in other populations such as Parkinson's, multiple sclerosis, spinal cord injury, etc. Or extend this research to the general neurological population. It is also recommended to conduct new studies on existing versions of FMA, in order to analyze psychometric properties that have not yet been studied in many of these versions, with special emphasis on content validity, structural validity, and internal consistency, since they are considered the most important measurement properties. It would also be advisable to adhere to the COSMIN methodology regulations, in order to achieve more "very good" and "positive rating" scores in the methodological quality, and a higher quality of evidence in future studies on psychometric properties of FMA. Finally, it is recommended that translations and validations be carried out in other languages, such as Spanish or Chinese, to be able to safely use this scale in these populations; also continue studying the psychometric properties of versions that have already been translated.

Strengths and limitations

This study is the most current systematic review identifying, evaluating, and summarizing the evidence on the FMA versions to measure upper-extremity sensorimotor function in stroke patients. The review was carried out by the most up-to-date COSMIN and PRISMA standards; therefore, an adequate method was used, following expert recommendations.[11,16,17]

Although an exhaustive search was carried out, not all available FMA versions may have been identified, because only articles in English and Spanish were selected, so it is possible that articles developed and used in other languages may not have been identified.

Despite FMA being shown to be a highly valid and reliable scale, and being a "gold standard", it has limitations such as a three-point scoring range, which can limit the detection of small changes in patients who recover little; as well as functional aspects of the sensorimotor function of the upper limb, such as its use in activities of daily living.

Conclusions

Different FMA versions and transcultural translations have been developed to evaluate upper-extremity sensorimotor function in stroke patients. Assessing this construct in stroke patients will allow us to improve the physical rehabilitation in this population. FMA versions for assessing upper-extremity sensorimotor function may help to facilitate screening, goal-setting, monitoring progress, and guiding decisions. This aspect evaluation is very useful in clinical practice and research, requiring new knowledge in this area. The evaluation of this variable is crucial since it allows for identifying and evaluating the intervention focused on improving the sensorimotor function of the upper extremity in patients with stroke. Based on current evidence for all the measurement properties of the versions assessing the level of upper-extremity sensorimotor function in stroke patients, the Fugl-Meyer Assessment (Full version, 50 items, FMA) seems to be the most recommended version based on the obtained results and the COSMIN guidelines, but this requires

further evaluation. More studies of high methodological quality that evaluate the measurement properties of the FMA versions to measure upper-extremity sensorimotor function in stroke patients are needed.

CRediT authorship contribution statement

Merchan Baeza Jose Antonio: Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation. **de Blas Zamorano Pablo:** Writing – review & editing, Writing – original draft, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Montagut Martínez Pedro:** Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Pérez Cruzado David:** Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Funding for open access charge: Universidad de Málaga / CBUA.

Declaration of Competing Interest

All authors have declared no conflict of interest.

We confirm that all authors have approved the paper for submission and that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved. This work has not been published and is not under consideration for publication elsewhere, neither in whole nor in partial form. We have no conflicts of interest to disclose.

References

1. Feigin VL, Owolabi MO, Feigin VL, et al. Pragmatic solutions to reduce the global burden of stroke: a World Stroke Organization–Lancet Neurology Commission. *Lancet Neurol.* 2023;22:1160–1206. https://doi.org/10.1016/S1474-4422(23)00277-6
2. Hua X, Liu M, Wu S. Definition, prediction, prevention and management of patients with severe ischemic stroke and large infarction. *Chin Med J.* 2023;136:2912–2922. https://doi.org/10.1097/CM9.0000000000002885
3. Katan M, Luft A. Global burden of stroke. *Semin Neurol.* 2018;38:208–211. https://doi.org/10.1055/s-0038-1649503
4. Kwakkel G, Stinear C, Essers B, et al. Motor rehabilitation after stroke: European Stroke Organisation (ESO) consensus-based definition and guiding framework. *Eur Stroke J.* 2023;8:880–894. https://doi.org/10.1177/23969873231191304
5. Ingram LA, Butler AA, Lord SR, Gandevia SC. Use a physiological profile to document upper limb motor impairment in ageing and neurological conditions. *J Physiol.* 2023;601:2251–2262. https://doi.org/10.1113/JP283703
6. Li S, Ghuman J, Gonzalez-Buonomo J, et al. Does spasticity correlate with motor impairment in the upper and lower limbs in ambulatory chronic stroke survivors? *Am J Phys Med Rehabil.* 2023;102:907–912. https://doi.org/10.1097/PHM.0000000000002247
7. Zhao K, He C, Xiang W, et al. Evidence of synergy coordination patterns of upper-limb motor control in stroke patients with mild and moderate impairment. *Front Physiol.* 2023;14:1214995. https://doi.org/10.3389/fphys.2023.1214995
8. Dawson J, Engineer ND, Cramer SC, et al. Vagus nerve stimulation paired with rehabilitation for upper limb motor impairment and function after chronic ischemic stroke: subgroup analysis of the randomized, blinded, pivotal, VNS-REHAB Device Trial. *Neurorehabil Neural Repair.* 2023;37:367–373. https://doi.org/10.1177/15459683221129274
9. Bushnell C, Bettger JP, Cockroft KM, et al. Chronic stroke outcome measures for motor function intervention trials: expert panel recommendations. *Circ: Cardiovasc Qual Outcomes.* 2015;8:1–13. https://doi.org/10.1161/CIRCOUTCOMES.115.002098
10. Gunduz ME, Bucak B, Keser Z. Advances in stroke neurorehabilitation. *JCM.* 2023;12:6734. https://doi.org/10.3390/jcm12216734
11. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27:1147–1157. https://doi.org/10.1007/s11136-018-1798-3
12. Santisteban L, Térémetz M, Bleton J-P, Baron J-C, Maier MA, Lindberg PG. Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review. *PLoS ONE.* 2016;11:e0154792. https://doi.org/10.1371/journal.pone.0154792
13. Yu B, Kaku A, Liu K , et al. Quantifying impairment and disease severity using AI models trained on healthy subjects. ArXiv 2023:arXiv:2311.12781v1.
14. Lin G-H, Huang C-Y, Lee S-C, et al. A 10-item Fugl-Meyer Motor Scale based on machine learning. *Phys Ther.* 2021;101:pzab036. https://doi.org/10.1093/ptj/pzab036
15. Tauchi Y, Kyougoku M, Takahashi K, Okita Y, Takebayashi T. Dimensionality and item-difficulty hierarchy of the Fugl-Meyer assessment of the upper extremity among Japanese patients who have experienced stroke. *Top Stroke Rehabil.* 2022;29:579–587. https://doi.org/10.1080/10749357.2021.1965797
16. Moher D, Liberati A, Tetzlaff J, Altman DG. PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 2009;339:1–8. https://doi.org/10.1136/bmj.b2535
17. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19:539–549. https://doi.org/10.1007/s11136-010-9606-8
18. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21:651–657. https://doi.org/10.1007/s11136-011-9960-1
19. Terwee CB, Jansma EP, Riphagen II, De Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009;18:1115–1123. https://doi.org/10.1007/s11136-009-9528-5
20. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336:924–926. https://doi.org/10.1136/bmj.39489.470347.AD
21. Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C. Reliability of the Fugl-Meyer Assessment for testing motor performance in patients following stroke. *Phys Ther.* 1993;73:447–454. https://doi.org/10.1093/ptj/73.7.447
22. Malouin F, Pichard L, Bonneau C, Durand A, Corriveau D. Evaluating motor recovery early after stroke: comparison of the Fugl-Meyer Assessment and the Motor Assessment Scale. *Arch Phys Med Rehabil.* 1994;75:1206–1212. https://doi.org/10.1016/0003-9993(94)90006-X
23. Lin J-H, Hsueh I-P, Sheu C-F, Hsieh C-L. Psychometric properties of the sensory scale of the Fugl-Meyer Assessment in stroke patients. *Clin Rehabil.* 2004;18:391–397. https://doi.org/10.1191/0269215504cr737oa
24. Hsueh I-P, Hsu M-J, Sheu C-F, Lee S, Hsieh C-L, Lin J-H. Psychometric comparisons of 2 versions of the Fugl-Meyer Motor Scale and 2 versions of the stroke rehabilitation assessment of movement. *Neurorehabil Neural Repair.* 2008;22:737–744. https://doi.org/10.1177/1545968308315999
25. See J, Dodakian L, Chou C, et al. A standardized approach to the Fugl-Meyer Assessment and its implications for clinical trials. *Neurorehabil Neural Repair.* 2013;27:732–741. https://doi.org/10.1177/1545968313491000
26. Chen K-L, Chen C-T, Chou Y-T, Shih C-L, Koh C-L, Hsieh C-L. Is the long form of the Fugl-Meyer Motor Scale more responsive than the short form in patients with stroke? *Arch Phys Med Rehabil.* 2014;95:941–949. https://doi.org/10.1016/j.apmr.2014.01.014
27. Barbosa NE, Forero SM, Galeano CP, et al. Translation and cultural validation of clinical observational scales – the Fugl-Meyer assessment for post-stroke sensorimotor function in Colombian Spanish. *Disabil Rehabil.* 2019;41:2317–2323. https://doi.org/10.1080/09638288.2018.1464604
28. Busk H, Alt Murphy M, Korsman R, Skou ST, Wienecke T. Cross-cultural translation and adaptation of the Danish version of the Fugl-Meyer assessment for post-stroke sensorimotor function. *Disabil Rehabil.* 2022;44:4888–4895. https://doi.org/10.1080/09638288.2021.1919215
29. Cecchi F, Carrabba C, Bertolucci F, et al. Transcultural translation and validation of Fugl–Meyer assessment to Italian. *Disabil Rehabil.* 2021;43:3717–3722. https://doi.org/10.1080/09638288.2020.1746844
30. Amano S, Umeji A, Uchita A, et al. Clinimetric properties of the Fugl-Meyer assessment with adapted guidelines for the assessment of arm function in hemiparetic patients after stroke. *Top Stroke Rehabil.* 2018;25:500–508. https://doi.org/10.1080/10749357.2018.1484987
31. Kim T, Hwang SH, Lee WJ, et al. The Korean Version of the Fugl-Meyer Assessment: reliability and validity evaluation. *Ann Rehabil Med.* 2021;45:83–98. https://doi.org/10.5535/arm.20225
32. Ikram M, Rehman SSU, Sunnerhagen KS, Alt Murphy M. Urdu translation and cross-cultural validation of the Fugl-Meyer assessment in people with stroke. *Disabil Rehabil.* 2022;44:8048–8053. https://doi.org/10.1080/09638288.2021.2003449
33. Onose G, Anghelescu A, Ionescu A, et al. Translation of the Fugl-Meyer assessment into Romanian: transcultural and semantic-linguistic adaptations and clinical validation. *Front Neurol.* 2023;13:1022546. https://doi.org/10.3389/fneur.2022.1022546
34. Hsieh Y-W, Hsueh I-P, Chou Y-T, Sheu C-F, Hsieh C-L, Kwakkel G. Development and validation of a short form of the Fugl-Meyer Motor Scale in patients with stroke. *Stroke.* 2007;38:3052–3054. https://doi.org/10.1161/STROKEAHA.107.490730
35. Fu TS-T, Wu C, Lin K, et al. Psychometric comparison of the shortened Fugl-Meyer Assessment and the streamlined Wolf Motor Function Test in stroke rehabilitation. *Clin Rehabil.* 2012;26:1043–1047. https://doi.org/10.1177/0269215511431474
36. Amano S, Umeji A, Takebayashi T, Takahashi K, Uchiyama Y, Domen K. Clinimetric properties of the shortened Fugl-Meyer Assessment for the assessment of arm

motor function in hemiparetic patients after stroke. *Top Stroke Rehabil.* 2020;27:290–295. https://doi.org/10.1080/10749357.2019.1701176
37. Michaelsen SM, Rocha AS, Knabben RJ, Rodrigues LP, Fernandes CGC. Translation, adaptation and inter-rater reliability of the administration manual for the Fugl-Meyer assessment. *Rev Bras Fisioter.* 2011;15:80–88.
38. De Weerdt WJG. Measuring recovery of arm-hand function in stroke patients: a comparison of the Brunnstrom-Fugl-Meyer test and the Action Research Arm test. *Physiother Canada.* 1985;37:65–70. https://doi.org/10.3138/ptc.37.2.065
39. Van Der Lee JH, Beckerman H, Lankhorst GJ, Bouter LM. The responsiveness of the Action Research Arm Test and the Fugl-Meyer Assessment scale in chronic stroke patients. *J Rehabil Med.* 2001;33(3):110. https://doi.org/10.1080/165019701750165916
40. Rabadi MH, Rabadi FM. Comparison of the Action Research Arm Test and the Fugl-Meyer Assessment as measures of upper-extremity motor weakness after stroke. *Arch Phys Med Rehabil.* 2006;87:962–966. https://doi.org/10.1016/j.apmr.2006.02.036
41. Woodbury ML, Velozo CA, Richards LG, Duncan PW, Studenski S, Lai S-M. Dimensionality and construct validity of the Fugl-Meyer Assessment of the upper extremity. *Arch Phys Med Rehabil.* 2007;88:715–723. https://doi.org/10.1016/j.apmr.2007.02.036
42. Hsieh Y, Wu C, Lin K, Chang Y, Chen C, Liu J. Responsiveness and validity of three outcome measures of motor function after stroke rehabilitation. *Stroke.* 2009;40:1386–1391. https://doi.org/10.1161/STROKEAHA.108.530584
43. Lin J-H, Hsu M-J, Sheu C-F, et al. Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Phys Ther.* 2009;89:840–850. https://doi.org/10.2522/ptj.20080285
44. Sullivan KJ, Tilson JK, Cen SY, et al. Fugl-Meyer Assessment of sensorimotor function after stroke: standardized training procedure for clinical practice and clinical trials. *Stroke.* 2011;42:427–432. https://doi.org/10.1161/STROKEAHA.110.592766
45. Wei X-J, Tong K, Hu X. The responsiveness and correlation between Fugl-Meyer Assessment, Motor Status Scale, and the Action Research Arm Test in chronic stroke with upper-extremity rehabilitation robotic training. *Int J Rehabil Res.* 2011;34:349–356. https://doi.org/10.1097/MRR.0b013e32834d330a
46. Kim H, Her J, Ko J, et al. Reliability, concurrent validity, and responsiveness of the Fugl-Meyer Assessment (FMA) for hemiplegic patients. *JPTS.* 2012;24(9):893.
47. Kim W-S, Cho S, Baek D, Bang H, Paik N-J. Upper extremity functional evaluation by Fugl-Meyer Assessment Scoring using Depth-Sensing Camera in hemiplegic stroke patients. *PLoS ONE.* 2016;11:e0158640. https://doi.org/10.1371/journal.pone.0158640
48. Hernández, Galeano E, Barbosa C, et al. Intra- and inter-rater reliability of Fugl-Meyer Assessment of upper extremity in stroke. *J Rehabil Med.* 2019;51:652–659. https://doi.org/10.2340/16501977-2590
49. Lundquist CB, Maribo T. The Fugl–Meyer assessment of the upper extremity: reliability, responsiveness and validity of the Danish version. *Disabil Rehabil.* 2017;39:934–939. https://doi.org/10.3109/09638288.2016.1163422
50. Hochleitner I, Pellicciari L, Castagnoli C, et al. Intra- and inter-rater reliability of the Italian Fugl-Meyer assessment of upper and lower extremity. *Disabil Rehabil.* 2023;45:2989–2999. https://doi.org/10.1080/09638288.2022.2114553
51. Roman N, Miclaus R, Repanovici A, Nicolau C. Equal opportunities for stroke survivors' rehabilitation: a study on the validity of the upper extremity Fugl-Meyer Assessment Scale translated and adapted into Romanian. *Medicina.* 2020;56:409. https://doi.org/10.3390/medicina56080409
52. Amano S, Umeji A, Uchita A, et al. Reliability of remote evaluation for the Fugl–Meyer assessment and the action research arm test in hemiparetic patients after stroke. *Top Stroke Rehabil.* 2018:1–6. https://doi.org/10.1080/10749357.2018.1481569
53. Page SJ, Levine P, Hade E. Psychometric properties and administration of the wrist/hand subscales of the Fugl-Meyer Assessment in minimally impaired upper extremity hemiparesis in stroke. *Arch Phys Med Rehabil.* 2012;93:2373–2376.e5. https://doi.org/10.1016/j.apmr.2012.06.017
54. Page SJ, Hade E, Persch A. Psychometrics of the wrist stability and hand mobility subscales of the Fugl-Meyer Assessment in moderately impaired stroke. *Phys Ther.* 2015;95:103–108. https://doi.org/10.2522/ptj.20130235
55. Persch AC, Gugiu PC, Velozo CA, Page SJ. Rasch analysis of the wrist and hand Fugl-Meyer: dimensionality and item-level characteristics. *J Neurol Phys Ther.* 2015;39:185–192.