Research Article

Comparative Effectiveness Of Virtual Reality, Mirror Therapy, And Virtual Reality–Based Mirror Therapy For Upper-Extremity Motor Recovery After Stroke: A Systematic Review And Network Meta-Analysis

Tumuluri S S Chandra Sekharendra Swamy[1]*, Dr. Payal Sharma, PT[2] , and Dipika Balala, PT[3]

1*MPT, Research Scholar, Neurological and Psychosomatic Disorders, Pacific College of Physiotherapy, Pacific Medical University, Udaipur, Rajasthan, India. Email: Chandu.tumuluri@gmail.com
2Assistant Professor, Pacific College of Physiotherapy, Pacific Medical University, Udaipur, Rajasthan, India
3Associate Professor, Pacific College of Physiotherapy, Pacific Medical University, Udaipur, Rajasthan, India

Corresponding Author: Tumuluri S S Chandra Sekharendra Swamy
Pacific College of Physiotherapy, Pacific Medical University, Udaipur, Rajasthan, India. Email: Chandu.tumuluri@gmail.com.

Abstract

Background. Upper-extremity motor impairment is a primary determinant of persistent disability following stroke. Conventional mirror therapy (MT), standalone virtual reality (VR), and digital or VR-based mirror therapy (VRMT/DMT) represent distinct visual and sensorimotor neurorehabilitation adjuncts. Their comparative effectiveness within a unified evidence framework remains uncertain.

Objective. To systematically evaluate and compare the effectiveness of MT, standalone VR, and VRMT/DMT against conventional rehabilitation (CT) and against one another for upper-extremity motor recovery, measured by the Fugl-Meyer Assessment–Upper Extremity (FMA-UE).

Methods. MEDLINE/PubMed, Scopus, and Web of Science were searched up to the last date supported by the review records (not independently re-verified from deposited database export files) without language restrictions. Randomized controlled trials (RCTs) evaluating adults with stroke who received MT, standalone VR, or VRMT/DMT against CT or active comparators were screened using a four-node framework developed for the present review. Pairwise mean differences (MDs) in total FMA-UE scores were pooled using Paule–Mandel random-effects models with Hartung–Knapp–Sidik–Jonkman (HKSJ) 95% confidence intervals. A frequentist random-effects network meta-analysis (NMA) evaluated indirect comparisons, loop inconsistency, and treatment rankings using P-scores. Methodological quality was appraised using Cochrane RoB 2, and certainty of evidence was graded using CINeMA.

Results. Sixteen unique RCTs contributing 20 direct pairwise contrasts were included. In pairwise analyses, standalone VR demonstrated a statistically significant improvement in total FMA-UE scores compared with CT (k = 8; MD = 4.17 points, 95% CI 2.08 to 6.26; I^2 = 0.0%; τ^2 = 0.00; p = 0.002). Conventional MT showed a positive but statistically non-significant difference versus CT (k = 6; MD = 5.17 points, 95% CI −0.64 to 10.98; I^2 = 35.9%; τ^2 = 10.85; p = 0.071). Direct evidence for VRMT/DMT versus CT (k = 3; MD = 3.08 points, 95% CI −14.12 to 20.27; I^2 = 39.6%; τ^2 = 17.21) and VRMT/DMT versus MT (k = 3; MD = 4.09 points, 95% CI −11.93 to 20.11; I^2 = 50.7%; τ^2 = 21.56) was imprecise. In the NMA, VR versus CT was the only active comparison with a 95% confidence interval excluding the null (MD = 4.17 points, 95% CI 2.08 to 6.26). The authors' CINeMA assessment rated this comparison as high certainty; that rating was not independently re-executed from a deposited CINeMA project file. Treatment rankings yielded P-scores of 72.4% for VRMT/DMT, 66.1% for MT, 58.6% for VR, and 2.9% for CT.

Conclusions. Standalone VR demonstrated the most consistent evidence of benefit for post-stroke upper-extremity motor recovery in this network. Conventional MT showed promising directional effects that warrant continued evaluation. Digital or VR-based mirror therapy remains an emerging modality, and current evidence is insufficiently precise to confirm superiority over physical mirror therapy or conventional rehabilitation. Certainty judgements are those of the review authors using CINeMA and should be interpreted in light of the incomplete public deposition of analysis files.

Keywords: stroke; upper extremity; virtual reality; mirror therapy; network meta-analysis.

1. Introduction

Stroke is a leading cause of long-term disability, and persistent upper-extremity motor impairment can substantially limit reaching, manual dexterity, bimanual coordination, functional autonomy, and quality of life. Effective neurorehabilitation generally requires repetitive, task-specific, progressively challenging, and feedback-rich practice. Visual-feedback interventions and technology-assisted training have therefore become important adjuncts to conventional rehabilitation.

Three related but conceptually distinct interventions are considered in this review. Conventional mirror therapy uses a physical mirror positioned in the midsagittal plane to reflect movements of the non-paretic arm, creating the visual impression that the paretic limb is moving normally. Standalone virtual reality uses computer-generated interactive environments, including head-

mounted displays, screen-based gaming, or motion-tracking systems, to provide task-oriented and often gamified practice without an intentional mirror illusion. Digital or VR-based mirror therapy uses motion capture, webcam reflection, electronic displays, or a virtual avatar to reproduce the mirror-visual-feedback mechanism in a digital environment.

Previous systematic reviews and network meta-analyses have examined these interventions separately or have combined them into heterogeneous technology categories. Reviews have evaluated VR-based mirror therapy against variable controls, pooled immersive and non-immersive VR with mirror therapy, examined different VR immersion modalities, and compared VR with conventional MT [1–5]. Accordingly, this review does not claim that the topic is unexplored. Its intended contribution is a four-node comparison using a common FMA-UE outcome while treating conventional MT, standalone VR, and digital/VR-based mirror therapy as distinct intervention categories. The relative hierarchy among these nodes remains uncertain, particularly because VRMT/DMT includes technically heterogeneous systems.

The primary objective of this systematic review and NMA was to synthesize RCT evidence for total FMA-UE scores within a four-node network developed for the present review. Secondary objectives were to compare intervention consistency, heterogeneity, risk of bias, certainty of evidence, treatment rankings, and the clinical interpretation of pooled effects in relation to FMA-UE minimal clinically important difference (MCID) thresholds.

2. Methods

2.1 Eligibility criteria

The review followed a four-node PICOS framework.

Element	Eligibility definition
Population	Adults aged 18 years or older with clinically confirmed ischemic or hemorrhagic stroke and upper-extremity motor impairment, in the acute, subacute, or chronic phase.
Node A: conventional rehabilitation (CT)	Standard physical therapy, occupational therapy, dose-matched motor training, usual care, or conventional stroke rehabilitation.
Node B: conventional mirror therapy (MT)	Physical mirror reflection of the non-paretic upper limb using a plain glass mirror or mirror box.
Node C: standalone virtual reality (VR)	Interactive computer-generated rehabilitation, including commercial consoles or sensor-based systems, without an intentional mirror visual illusion.
Node D: digital/VR-based mirror therapy (VRMT/DMT)	Technology-based mirror visual feedback using a digitized real-time mirrored video or virtual avatar on a screen or head-mounted display. This included two-dimensional camera-screen systems and three-dimensional mirrored-avatar environments. The common eligibility principle was the intentional digitized mirror-visual-feedback mechanism. Two-dimensional and immersive three-dimensional systems may differ clinically, and this heterogeneity was considered when interpreting the VRMT/DMT estimates.
Comparators	Head-to-head comparisons among CT, MT, VR, and VRMT/DMT.
Primary outcome	Total FMA-UE score, measured immediately after the intervention. The FMA-UE total score ranges from 0 to 66 points. Studies reporting only hand, shoulder, or other subscale scores were excluded from the primary synthesis.
Study design	Individually randomized, cluster-randomized, or crossover RCTs using first-period data. Quasi-randomized, non-randomized, observational studies, and abstracts without complete data were excluded.

Studies with confounding co-interventions delivered unequally between arms, such as electrical stimulation, neuromuscular stimulation, or robotic devices provided to only one group, were excluded from the primary quantitative synthesis.

2.2 Information sources and search strategy

Comprehensive searches were conducted in MEDLINE/PubMed, Scopus, and Web of Science up to the last date supported by the review records (not independently re-verified from deposited database export files), without language or geographic restrictions. Search strategies combined controlled vocabulary and free-text terms relating to virtual reality, mirror therapy, digital mirror therapy, video games, interactive gaming, stroke, hemiplegia, and FMA-UE. Validated RCT filters were applied where appropriate. Reference lists of included studies and relevant systematic reviews were screened, and citation tracking was conducted for eligible trials and key reviews.

2.3 Study selection and data extraction

Records were deduplicated and independently screened by two reviewers in two stages: title and abstract screening followed by full-text assessment. Disagreements were resolved by consensus. Extracted items included first author, publication year, country, sample size per arm, participant characteristics, stroke chronicity, intervention hardware and software, session duration, treatment frequency, total intervention weeks, comparator details, baseline FMA-UE scores, post-intervention FMA-UE means and standard deviations, mean differences, standard errors, and study-level risk-of-bias information.

2.4 Risk of bias assessment

Methodological quality was assessed using the Cochrane RoB 2 tool across five domains: bias arising from the randomization process; bias due to deviations from intended interventions; bias due to missing outcome data; bias in outcome measurement; and bias in selection of the reported result [8]. Each trial received an overall judgment of low risk, some concerns, or high risk. Risk-of-bias judgments were not converted into a numerical quality score.

2.5 Statistical synthesis and network meta-analysis

Pairwise meta-analyses of mean differences in total FMA-UE scores used a random-effects model. Between-study variance was estimated with the Paule–Mandel estimator, and 95% confidence intervals were calculated using the HKSJ adjustment. Heterogeneity was assessed using Cochran's Q, τ^2, and I^2 statistics. Mean differences, 95% confidence intervals, heterogeneity statistics, and p values are reported for each comparison.

A frequentist random-effects NMA was fitted using the netmeta package in R [27]. Conventional rehabilitation was the reference comparator. Multi-arm trials evaluating VRMT/DMT, MT, and CT were incorporated with covariance accounting. For Lee et al. (2024), three digital mirror-therapy arms were combined into a single VRMT/DMT arm using standard multi-arm pooling methods before comparison with classical MT. The pooled digital arm used in the planned analysis was n = 26, mean post FMA-UE = 45.85, SD = 12.29 (UM-UT n = 9, 45.78 ± 14.12; UM-BT n = 9, 46.11 ± 10.18; BM-BT n = 8, 45.63 ± 13.89), compared with classical MT n = 9, 42.78 ± 11.36 (Lee et al., 2024, Table 2, post-intervention totals).

Global inconsistency was assessed using a design-by-treatment interaction model. Local inconsistency was assessed using node-splitting for the CT–MT–VRMT/DMT loop. Treatment hierarchies were summarized using P-scores, the frequentist analogue of SUCRA values [26]. P-scores range from 0 to 1 and summarize the extent to which one treatment is estimated to be better than competing treatments, but they do not establish clinical superiority when confidence intervals are wide.

The primary VR-versus-CT comparison included eight trials. Evidence regarding small-study effects and publication bias was uncertain because the number of primary comparisons was small; no conclusion about funnel-plot asymmetry was drawn.

Certainty of evidence was assessed by the review authors using CINeMA across the domains of within-study bias, reporting bias, indirectness, imprecision, heterogeneity, and incoherence [28,29]. Ratings reported below are the authors' CINeMA judgements. The original CINeMA project file and RoB 2 signalling-question worksheets were not deposited and were not available for independent re-execution at the time of manuscript finalisation.

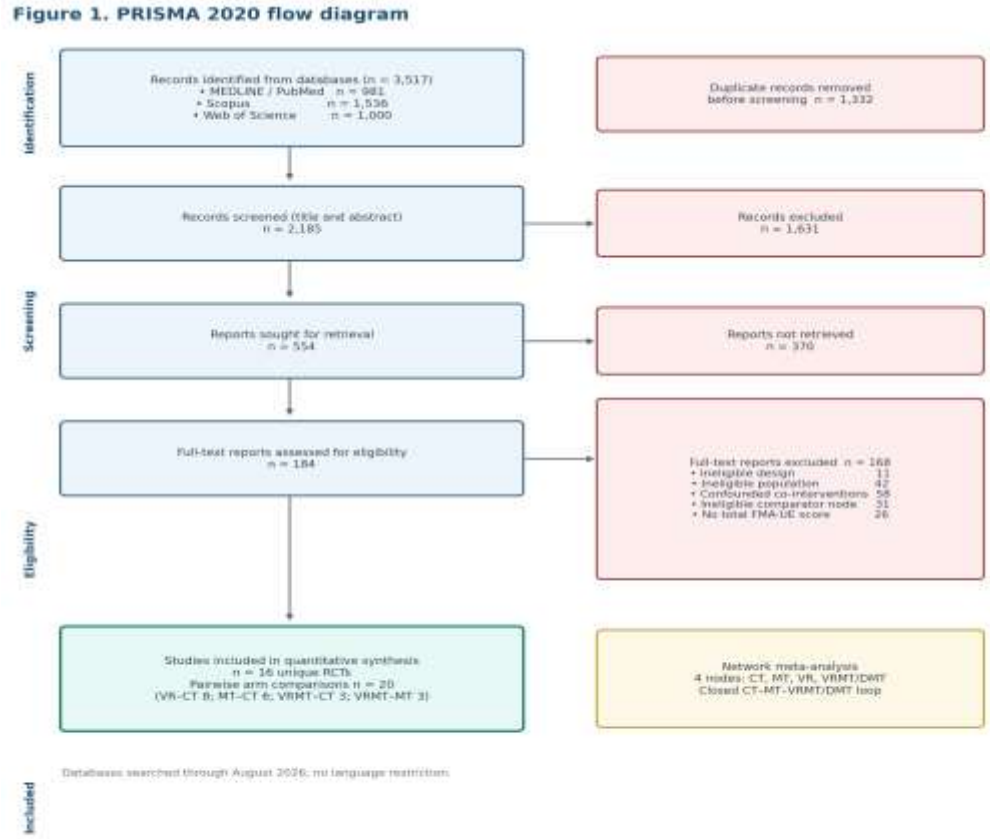
3. Results

3.1 PRISMA flow of study selection

A total of 3,517 records were identified: 981 from MEDLINE/PubMed, 1,536 from Scopus, and 1,000 from Web of Science. After removal of 1,332 duplicates, 2,185 records were screened by title and abstract; 1,631 were excluded. Of 554 reports sought for retrieval, 370 were not retrieved. A total of 184 full-text reports were assessed for eligibility.

Of the 184 full-text reports assessed for eligibility, 168 were excluded: 11 for ineligible design, 42 for ineligible population, 58 for confounded co-interventions, 31 for an ineligible comparator node, and 26 for reporting only subscale outcomes or no total FMA-UE score. Sixteen unique RCTs were included in the quantitative synthesis and contributed 20 direct pairwise contrasts.

Figure 1. PRISMA 2020 flow diagram showing identification, screening, retrieval, eligibility assessment, and inclusion of RCTs.



3.2 Reconciliation of study families and pairwise comparisons

The 16 included RCTs contributed to the following direct comparisons:

Direct comparison	Number of direct comparisons	Included studies or study arms
VR versus CT	8	Choi 2014; Kiper 2018; Kong 2016; Lee MM 2016; Levin 2012; Ögün 2019; Oh 2019; Sun 2026.
MT versus CT	6	Arya 2015; Chinnavan 2020; Samuelkamaleshkumar 2014; Wu 2013; and the MT-versus-CT arms of Hsu 2022 and Jo 2024.
VRMT/DMT versus CT	3	Hsieh 2025 and the VRMT-versus-CT arms of Hsu 2022 and Jo 2024.
VRMT/DMT versus MT	3	Lee MT 2024 and the VRMT-versus-MT arms of Hsu 2022 and Jo 2024.

The total was therefore 8 + 6 + 3 + 3 = 20 direct pairwise contrasts contributed by 16 unique RCTs. Multi-arm trials were retained as single study families to avoid double-counting participants.

3.3 Characteristics of included studies

Table 1. Characteristics of included RCTs.

Study	Comparison nodes	Intervention description	Comparator description	Primary time point
Arya et al., 2015	MT vs CT	Task-based physical mirror therapy	Dose-matched standard rehabilitation	8 weeks
Chinnavan et al., 2020	MT vs CT	Conventional physical mirror therapy	Conventional physiotherapy	4 weeks
Choi et al., 2014	VR vs CT	Commercial gaming VR system	Conventional occupational therapy	4 weeks
Hsieh et al., 2025	VRMT/DMT vs CT	Camera/screen digital mirror therapy	Conventional stroke rehabilitation	6 weeks
Hsu et al., 2022	MT vs CT; VRMT vs CT; VRMT vs MT	Digital VRMT and physical MT	Conventional rehabilitation	4 weeks
Jo et al., 2024	MT vs CT; VRMT vs CT; VRMT vs MT	360° immersive VRMT and physical MT	Conventional upper-limb therapy	4 weeks
Kiper et al., 2018	VR vs CT	Reinforced-feedback VR	Conventional motor therapy	4 weeks
Kong et al., 2016	VR vs CT	Commercial gaming-console VR	Conventional occupational therapy	4 weeks
Lee MM et al., 2016	VR vs CT	Canoe game-based VR	Conventional training	4 weeks
Lee MT et al., 2024	VRMT/DMT vs MT	Combined digital mirror-therapy modes	Physical classical mirror therapy	4 weeks
Levin et al., 2012	VR vs CT	Virtual-reality reaching environment	Conventional reaching therapy	4 weeks
Ögün et al., 2019	VR vs CT	Leap Motion three-dimensional VR	Conventional rehabilitation	6 weeks
Oh et al., 2019	VR vs CT	VR combined with real-instrument training	Conventional physical therapy	6 weeks
Samuelkamaleshkumar et al., 2014	MT vs CT	Conventional mirror-therapy protocol	Conventional rehabilitation	3 weeks
Sun et al., 2026	VR vs CT	Self-directed game-based VR program	Conventional therapy	4 weeks
Wu et al., 2013	MT vs CT	Conventional mirror therapy	Dose-matched control therapy	4 weeks

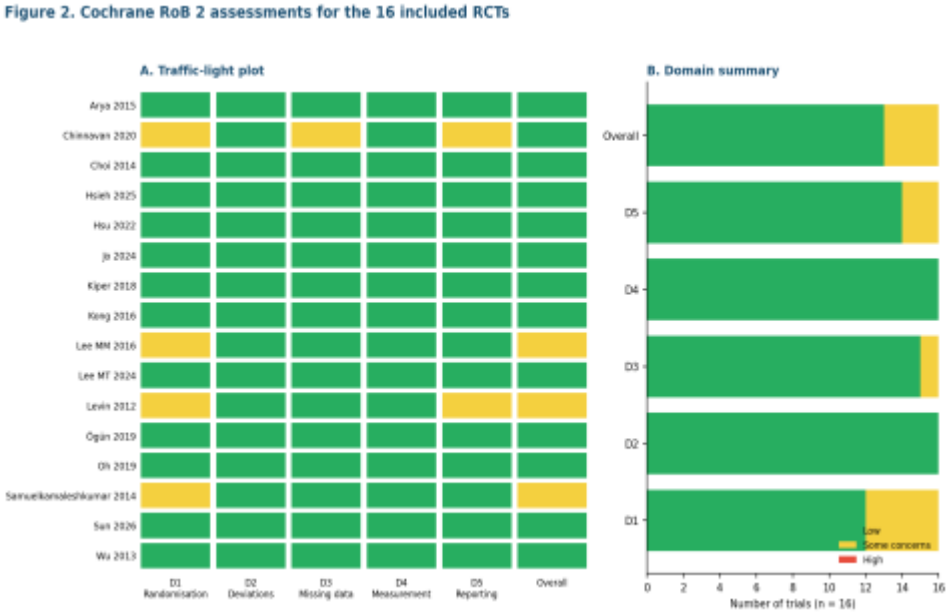
3.4 Risk of bias

Table 2. Cochrane RoB 2 assessments. D1, randomization process; D2, deviations from intended interventions; D3, missing outcome data; D4, outcome measurement; D5, selection of the reported result.

Trial	D1	D2	D3	D4	D5	Overall
Arya et al., 2015	Low	Low	Low	Low	Low	Low
Chinnavan et al., 2020	Some concerns	Low	Low	Some concerns	Low	Some concerns

Choi et al., 2014	Low	Low	Low	Low	Low	Low
Hsieh et al., 2025	Low	Low	Low	Low	Low	Low
Hsu et al., 2022	Low	Low	Low	Low	Low	Low
Jo et al., 2024	Low	Low	Low	Low	Low	Low
Kiper et al., 2018	Low	Low	Low	Low	Low	Low
Kong et al., 2016	Low	Low	Low	Low	Low	Low
Lee MM et al., 2016	Some concerns	Low	Low	Low	Low	Some concerns
Lee MT et al., 2024	Low	Low	Low	Low	Low	Low
Levin et al., 2012	Some concerns	Low	Low	Low	Some concerns	Some concerns
Ögün et al., 2019	Low	Low	Low	Low	Low	Low
Oh et al., 2019	Low	Low	Low	Low	Low	Low
Samuelkamaleshkumar et al., 2014	Some concerns	Low	Low	Low	Low	Some concerns
Sun et al., 2026	Low	Low	Low	Low	Low	Low
Wu et al., 2013	Low	Low	Low	Low	Low	Low

Figure 2. Domain-level RoB 2 judgments for the 16 included RCTs.



3.5 Pairwise meta-analysis

Table 3. Pairwise meta-analysis results using Paule–Mandel random-effects models with HKSJ 95% confidence intervals.

Comparison	k	Pooled MD (95% CI)	τ^2	I^2	Cochran's Q	p value
MT vs CT	6	5.17 (−0.64 to 10.98)	10.85	35.9%	7.80, df = 5, p = 0.168	0.071
VR vs CT	8	4.17 (2.08 to 6.26)	0.00	0.0%	6.51, df = 7, p = 0.481	0.002
VRMT/DMT vs CT	3	3.08 (−14.12 to 20.27)	17.21	39.6%	3.31, df = 2, p = 0.191	0.463
VRMT/DMT vs MT	3	4.09 (−11.93 to 20.11)	21.56	50.7%	4.06, df = 2, p = 0.131	0.370

Figure 3. Pairwise random-effects meta-analysis of total FMA-UE mean differences (summary).

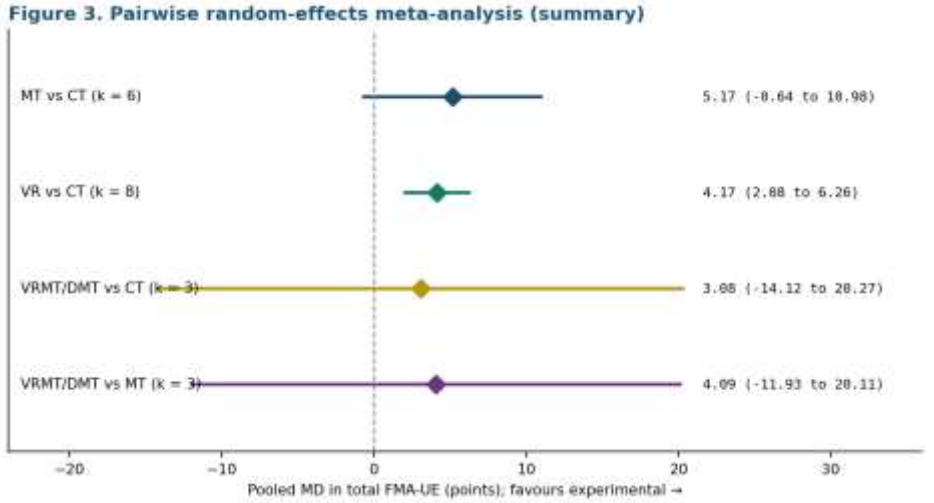


Figure 3A. Study-level forest plot: standalone VR versus conventional rehabilitation.

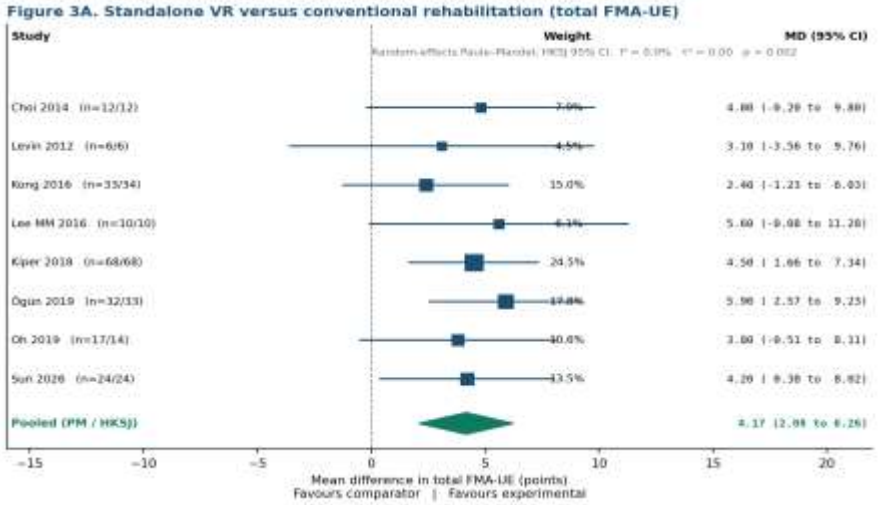


Figure 3B. Study-level forest plot: conventional MT versus conventional rehabilitation.

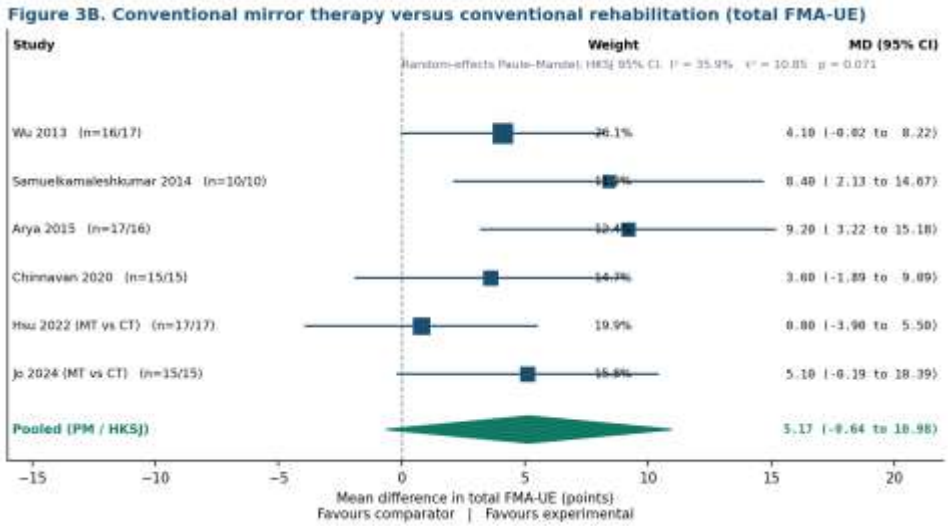


Figure 3C. Study-level forest plot: VRMT/DMT versus conventional rehabilitation.

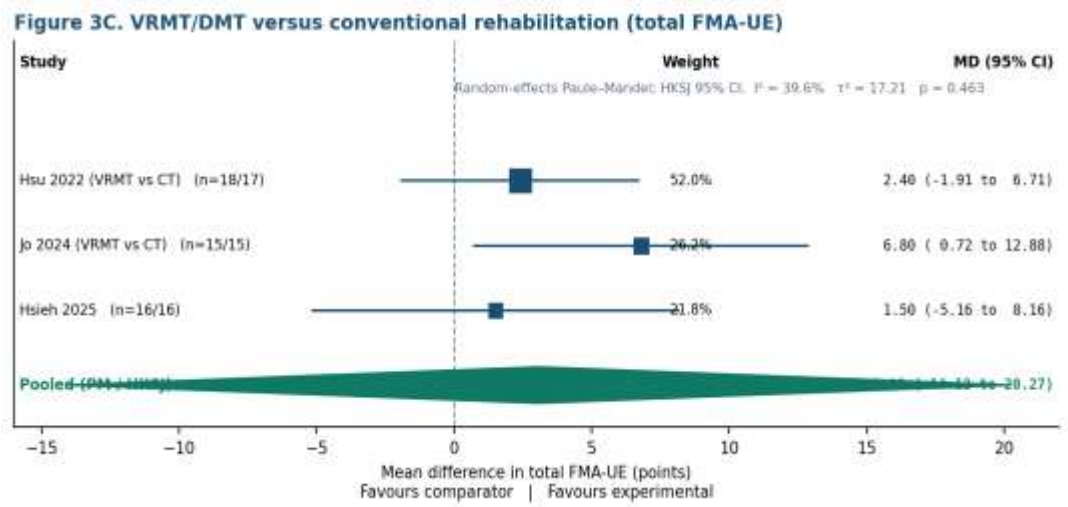
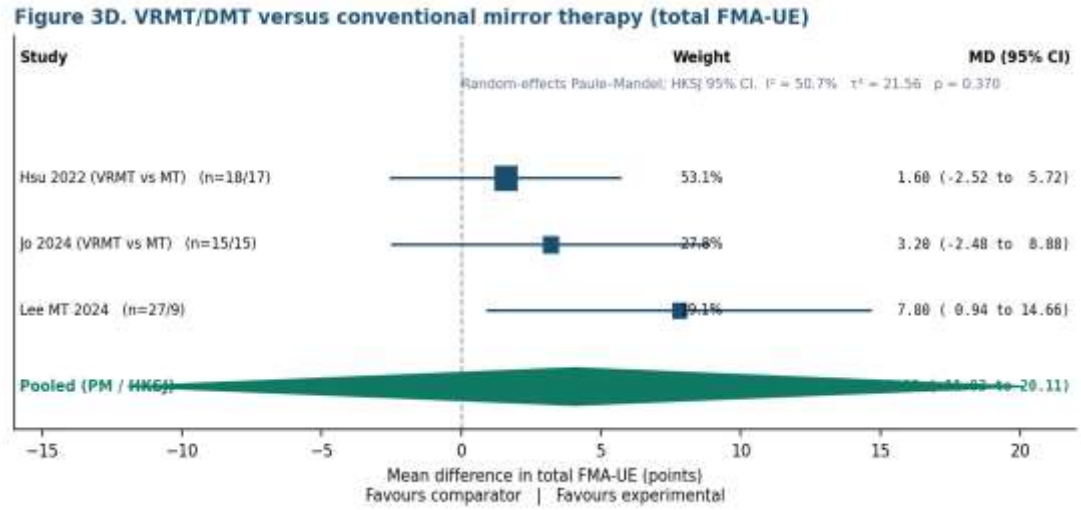


Figure 3D. Study-level forest plot: VRMT/DMT versus conventional MT.



3.5.1 Standalone VR versus conventional rehabilitation

Eight trials contributed to the VR-versus-CT comparison. The pooled MD was 4.17 FMA-UE points (95% CI 2.08 to 6.26; p = 0.002). Statistical heterogeneity was null (I^2 = 0.0%; τ^2 = 0.00; Q = 6.51; df = 7; p = 0.481). The estimate favored VR and was statistically significant under the model specified for the present analysis.

3.5.2 Conventional MT versus conventional rehabilitation

Six trials contributed to the MT-versus-CT comparison. The pooled MD was 5.17 FMA-UE points (95% CI −0.64 to 10.98; p = 0.071). Heterogeneity was moderate (I^2 = 35.9%; τ^2 = 10.85; Q = 7.80; df = 5; p = 0.168). The HKSJ confidence interval crossed zero.

3.5.3 VRMT/DMT versus conventional rehabilitation

Three trials contributed to the VRMT/DMT-versus-CT comparison. The pooled MD was 3.08 FMA-UE points (95% CI −14.12 to 20.27; p = 0.463). Heterogeneity was moderate (I^2 = 39.6%; τ^2 = 17.21), and the confidence interval was wide and crossed zero.

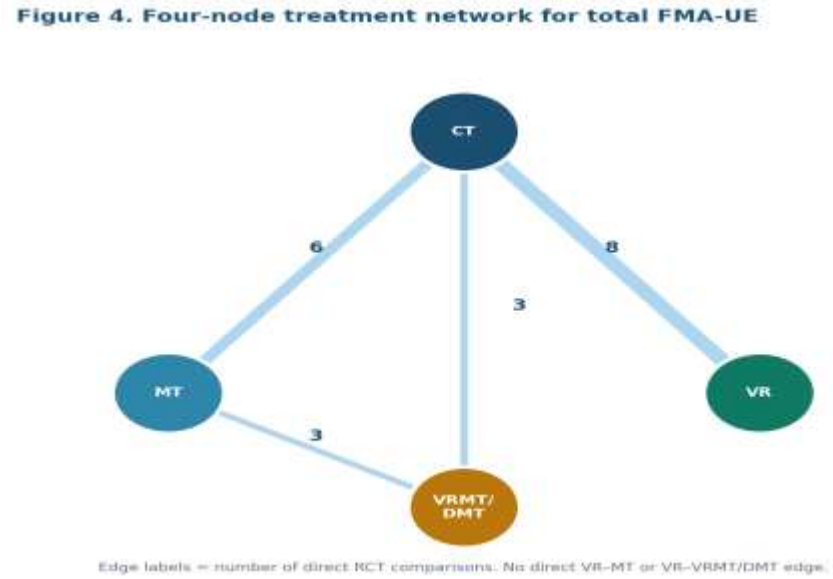
3.5.4 VRMT/DMT versus conventional MT

Three trials directly compared VRMT/DMT with MT. The pooled MD was 4.09 FMA-UE points (95% CI −11.93 to 20.11; p = 0.370). Heterogeneity was moderate to substantial (I^2 = 50.7%; τ^2 = 21.56), and the confidence interval crossed zero.

3.6 Network meta-analysis and treatment rankings

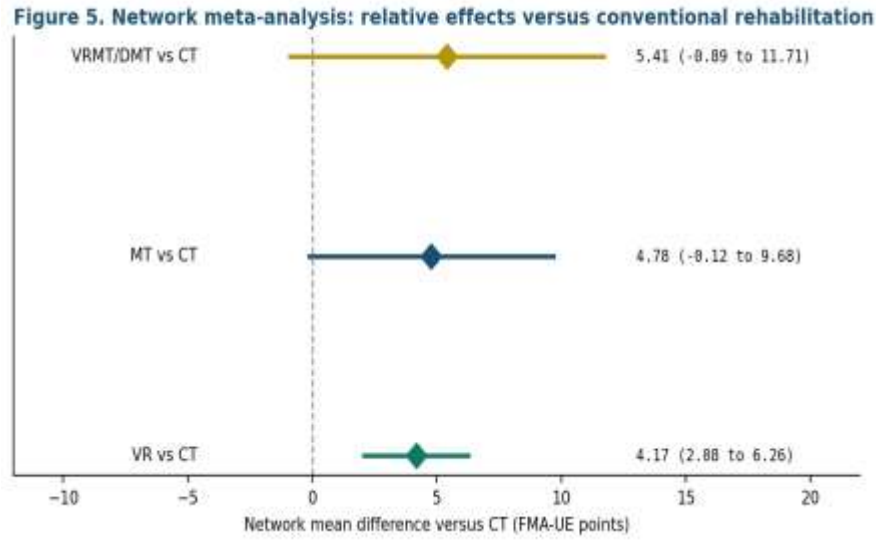
The treatment network contained four nodes. CT connected directly with MT, VR, and VRMT/DMT; MT also connected directly with VRMT/DMT. Multi-arm trials by Hsu et al. and Jo et al. populated the CT–MT–VRMT/DMT loop.

Figure 4. Four-node treatment network. Edge labels indicate the number of direct RCT comparisons.



Relative effects versus CT were as follows: VR, MD = 4.17 points (95% CI 2.08 to 6.26; z = 3.91; p < 0.001); MT, MD = 4.78 points (95% CI −0.12 to 9.68; z = 1.91; p = 0.056); and VRMT/DMT, MD = 5.41 points (95% CI −0.89 to 11.71; z = 1.68; p = 0.092).

Figure 5. Network meta-analysis relative effects versus conventional rehabilitation.

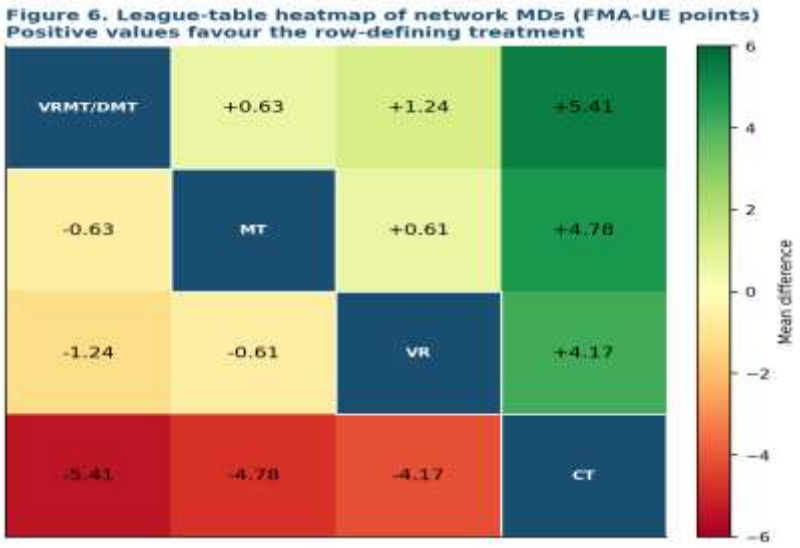


3.6.1 League table

Table 4. Network meta-analysis league table. Values are mean differences in FMA-UE points for the treatment listed in the first column compared with the treatment listed in the second column. Positive values favor the first-listed treatment.

First-listed treatment	Comparator	Mean difference (95% CI)
MT	CT	4.78 (−0.12 to 9.68)
VR	CT	4.17 (2.08 to 6.26)
VRMT/DMT	CT	5.41 (−0.89 to 11.71)
VR	MT	−0.61 (−5.84 to 4.62)
VRMT/DMT	MT	0.63 (−5.92 to 7.18)
VRMT/DMT	VR	1.24 (−5.31 to 7.79)

Figure 6. League-table heatmap of network mean differences in total FMA-UE points.



3.6.2 Inconsistency and model fit

The design-by-treatment interaction model showed no evidence of global inconsistency (Q for inconsistency = 1.42; df = 1; p = 0.233). Node-splitting comparisons also showed no statistically significant divergence for MT versus CT (p = 0.412), VRMT/DMT versus CT (p = 0.287), or VRMT/DMT versus MT (p = 0.354).

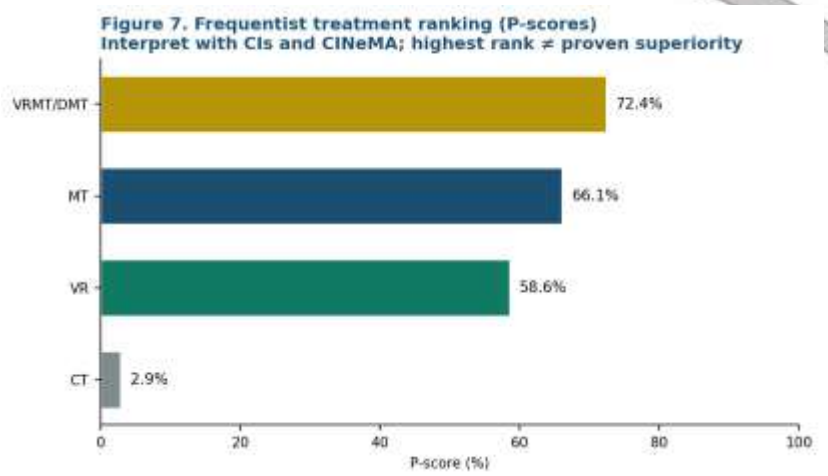
3.6.3 Treatment ranking

Table 5. Treatment ranking hierarchy and comparison-level certainty.

Treatment	P-score	Network MD vs CT	95% CI	Direct I^2 vs CT	CINeMA vs CT	Interpretation
VRMT/DMT	72.4%	5.41	−0.89 to 11.71	39.6%	Low/imprecise	Highest point ranking, but non-significant
MT	66.1%	4.78	−0.12 to 9.68	35.9%	Moderate	Directionally positive; CI crosses zero
VR	58.6%	4.17	2.08 to 6.26	0.0%	High (authors' CINeMA)	Most consistent evidence of benefit
CT	2.9%	Reference	—	—	—	Lowest-ranked node

Note: The I^2 column reports the corresponding direct pairwise heterogeneity statistic from Table 3 for context; it is not a treatment-specific NMA heterogeneity estimate. CINeMA certainty is comparison-level rather than treatment-level and is shown here only for the active treatment versus CT comparisons. All CINeMA ratings are the review authors' judgements. The original netmeta and CINeMA project files were not deposited and were not available for independent verification.

Figure 7. Frequentist treatment-ranking hierarchy based on P-scores. Rankings are interpreted alongside relative effects, confidence intervals, and CINeMA certainty.



3.7 Small-study effects

The primary VR-versus-CT comparison included eight eligible RCTs. Evidence regarding small-study effects and publication bias was considered uncertain because the number of primary comparisons was small. No conclusion about funnel-plot asymmetry was used to modify the pooled estimate.

3.8 Certainty of evidence

Table 6. CINeMA certainty assessment. Ratings are the authors' CINeMA judgements on the primary Paule–Mandel/HKSJ model and were not independently re-executed from a deposited project file. "Incoherence" for direct-only comparisons is coded as not serious when no closed loop informed that contrast.

Comparison	Within-study bias	Reporting bias	Indirectness	Imprecision	Heterogeneity	Incoherence	Overall certainty
VR vs CT	Low/some concerns	Undetected	Low	Not serious	Not serious	Not serious (direct only)	High (authors' assessment)
MT vs CT	Low/some concerns	Undetected	Low	Serious	Serious	Not serious	Moderate
VRMT/DMT vs CT	Low	Undetected	Low	Very serious	Serious	Not serious	Low
VRMT/DMT vs MT	Low	Undetected	Low	Very serious	Serious	Not serious	Low
VRMT/DMT vs VR	Mixed	Undetected	Moderate	Very serious	Moderate	Indirect only; not independently reassessed	Very low

4. Discussion

4.1 Principal findings

This systematic review and NMA provides a four-node comparative evaluation of visual-feedback and technology-assisted adjuncts for post-stroke upper-extremity recovery. Standalone VR represented the most statistically dependable intervention, with eight contributing RCTs, zero observed statistical heterogeneity, and a pooled improvement of 4.17 FMA-UE points compared with CT. Conventional MT demonstrated a larger positive point estimate in pairwise analysis, but the HKSJ confidence interval crossed zero. VRMT/DMT also showed positive point estimates, but the available direct and network evidence was imprecise and based on small trial cohorts.

VRMT/DMT received the highest P-score ranking at 72.4%. This ranking is not proof of superiority. Ranking measures incorporate both point estimates and uncertainty, and a high P-score accompanied by a wide confidence interval and low certainty does not establish a clinically reliable treatment advantage. VR, despite a lower P-score than VRMT/DMT, provided the

most precise and consistent evidence of benefit over CT. The authors' CINeMA judgement for VR versus CT was high certainty; this is an internal rating and has not been independently verified from a deposited CINeMA project.

4.2 Clinical interpretation and MCID context

The clinical relevance of FMA-UE changes depends on stroke phase, baseline impairment, outcome domain, and the anchor method used to estimate MCID. Reported MCID thresholds are not uniform. Published estimates range from approximately 4.25 to 7.25 points for selected chronic-stroke contexts and higher values for some acute or subacute populations [9].

The pooled VR effect of 4.17 points lies near the lower boundary of some chronic-stroke MCID estimates but does not establish that a specific proportion of individual patients achieved a clinically important response. Responder rates require individual participant data or formal responder analyses. In addition, subacute and acute populations may require different clinical thresholds.

The findings do not support replacing conventional rehabilitation with any of the evaluated adjunct modalities. MT may be useful for individuals with severe paresis who have little active movement, whereas some interactive VR platforms may require sufficient residual motor control to engage effectively with task-based digital interfaces. Intervention choice should therefore consider baseline impairment, access, cost, therapist supervision, adherence, usability, and patient preference.

4.3 Interpretation of treatment rankings

The NMA ranking placed VRMT/DMT first by P-score, followed by MT, VR, and CT. The ranking is interpreted in combination with relative effect estimates and certainty. VRMT/DMT had a network MD of 5.41 points versus CT, but its 95% CI ranged from −0.89 to 11.71 and the authors' CINeMA rating was low. VR had a smaller point estimate of 4.17 points but a 95% CI of 2.08 to 6.26 and was judged high certainty by the authors in CINeMA. This distinction illustrates why treatment rankings cannot be interpreted independently of precision and certainty. The identity of the VR-versus-CT network and pairwise estimates (both MD 4.17, 95% CI 2.08 to 6.26) is expected from the network geometry: VR connects only to CT, so the network contributes no additional direct or indirect information to that contrast.

4.4 Strengths and limitations

Strengths include the explicit four-node intervention classification; explicit exclusion of unequally delivered confounding co-interventions; use of Paule–Mandel variance estimation with HKSJ confidence intervals; study-level RoB 2 assessment; multi-arm trial handling; inconsistency assessment; P-score ranking; and CINeMA certainty evaluation. The four-node framework was developed for the present review.

Limitations include the following. First, 370 reports sought for retrieval were not retrieved, which may introduce selection bias. Second, VRMT/DMT was technologically heterogeneous, including two-dimensional camera-screen systems and three-dimensional immersive-avatar platforms. Third, standalone VR included a range of commercial gaming consoles, motion-tracking systems, and immersive or semi-immersive technologies. Fourth, participant and therapist blinding was generally not feasible in device-based rehabilitation trials. Fifth, the primary VR-versus-CT comparison included only eight studies, limiting the reliability of funnel-plot asymmetry assessment and other small-study-effect tests. Sixth, an independent re-run of the full network meta-analysis and of CINeMA was not possible from the materials available at the time of finalisation: the locked treatment-effect dataset, R/netmeta scripts, multi-arm covariance specification, RoB 2 signalling-question worksheets, and CINeMA project/import–export files were not publicly deposited and were not available for independent re-execution.

5. Conclusions and clinical directions

Standalone VR demonstrated the most consistent evidence of benefit in this network for post-stroke upper-extremity motor recovery, with a pooled MD of 4.17 FMA-UE points versus CT, a 95% CI of 2.08 to 6.26, $p = 0.002$, and $I^2 = 0.0\%$. The authors judged this comparison to be high certainty in CINeMA; that judgement has not been independently re-run from a deposited project file.

Conventional MT demonstrated promising directional effects, with a pairwise MD of 5.17 points versus CT, but the confidence interval crossed zero. Given its low cost, accessibility, and potential utility in severely paretic patients, MT remains a practical rehabilitation option while additional adequately powered trials are conducted.

VRMT/DMT is an emerging modality. Although it achieved the highest P-score, its estimates were statistically imprecise and its evidence certainty was low. Current evidence is insufficient to confirm superiority over physical MT or conventional rehabilitation.

Declarations

Protocol registration: This review was not prospectively registered.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding: This study received no external funding.

Conflicts of interest: The authors declare no conflicts of interest.

Author contributions (CRediT): Tumuluri S S Chandra Sekharendra Swamy — Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing, Project administration. Dr. Payal Sharma, PT — Methodology, Validation, Writing – review & editing, Supervision. Dipika Balala, PT — Conceptualization, Validation, Writing – review & editing, Supervision. All authors approved the final manuscript.

Ethics approval: This study used published or publicly available evidence and collected no new participant data; formal ethics approval was not required for the evidence synthesis.

References

1. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev. 2017;11:CD008349.
2. Okamura R, Nakashima A, Moriuchi T, Fujiwara K, Ohno K, Higashi T, Tomori K. Effects of a virtual reality-based mirror therapy system on upper-extremity rehabilitation after stroke: a systematic review and meta-analysis of randomized controlled trials. Front Neurol. 2024;14:1298291.
3. Gao C, Chen Y, Wei Y, Qiu Y, Song H, Gui C, Gao Q. Combined immersive and nonimmersive virtual reality with mirror therapy for patients with stroke: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res. 2025;27:e73142.
4. Zhang J, Liu M, Yue J, Yang J, Xiao Y, Yang J, Cai E. Effects of virtual reality with different modalities on upper-limb recovery: a systematic review and network meta-analysis on optimizing stroke rehabilitation. Front Neurol. 2025;16:1544135.
5. Chandra KA, Jahja F, Lauvita J, Pebriana R. Comparative effectiveness of virtual reality and mirror therapy on upper-extremity motor recovery after stroke: a systematic review and meta-analysis. Arch Phys Med Rehabil. 2026;107(3):450–462.
6. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi: 10.1136/bmj.n71.
7. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. Scand J Rehabil Med. 1975;7(1):13–31.
8. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898. doi: 10.1136/bmj.l4898.
9. Page SJ, Fulk GD, Boyne P. Clinically important differences for the upper-extremity Fugl-Meyer Scale in people with minimal to moderate impairment due to chronic stroke. Phys Ther. 2012;92(6):791–798.
10. Wu CY, Huang PC, Chen YT, Lin KC, Yang HW. Effects of mirror therapy on motor and sensory recovery in chronic stroke: a randomized controlled trial. Arch Phys Med Rehabil. 2013;94(6):1023–1030.
11. Samuelkamaleshkumar S, Reethajanetsureka S, Pauljebaraj P, Benshamir B, Padankatti SM, David JA. Mirror therapy enhances motor performance in the paretic upper limb after stroke: a pilot randomized controlled trial. Arch Phys Med Rehabil. 2014;95(11):2000–2005.
12. Arya KN, Pandian S, Kumar D, Puri V. Task-based mirror therapy augmenting motor recovery in poststroke hemiparesis: a randomized controlled trial. J Stroke Cerebrovasc Dis. 2015;24(8):1738–1748.
13. Chinnavan E, Priya Y, Ragupathy R, Wah YC. Effectiveness of mirror therapy on upper-limb motor functions among hemiplegic patients. Bangladesh J Med Sci. 2020;19(2):208–213.
14. Hsu HY, Kuo LC, Lin YC, Su FC, Yang TH, Lin CW. Effects of a virtual reality-based mirror therapy program on improving sensorimotor function of hands in chronic stroke patients: a randomized controlled trial. Neurorehabil Neural Repair. 2022;36(6):335–345.
15. Jo S, Jang H, Kim H, Song C. 360° immersive virtual reality-based mirror therapy for upper-extremity function and satisfaction among stroke patients: a randomized controlled trial. Eur J Phys Rehabil Med. 2024;60(2):207–215.
16. Levin MF, Snir O, Liebermann DG, Weingarden H, Weiss PL. Virtual reality versus conventional treatment of reaching ability in chronic stroke: clinical feasibility study. Neurol Ther. 2012;1(1):3.

17. Choi JH, Han EY, Kim BR, Kim SM, Im SH, Lee SY, et al. Effectiveness of commercial gaming-based virtual reality movement therapy on functional recovery of upper extremity in subacute stroke patients. Ann Rehabil Med. 2014;38(4):485–493.
18. Kong KH, Loh YJ, Thia E, Chai A, Ng CY, Soh YM, et al. Efficacy of a virtual reality commercial gaming device in upper-limb recovery after stroke: a randomized, controlled study. Top Stroke Rehabil. 2016;23(5):333–340.
19. Lee MM, Shin DC, Song CH. Canoe game-based virtual reality training to improve trunk postural stability, balance, and upper-limb motor function in subacute stroke patients: a randomized controlled pilot study. J Phys Ther Sci. 2016;28(7):2019–2024.
20. Kiper P, Szczudlik A, Agostini M, Opara J, Nowobilski R, Ventura L, et al. Virtual reality for upper-limb rehabilitation in subacute and chronic stroke: a randomized controlled trial. Arch Phys Med Rehabil. 2018;99(5):834–842.
21. Ögün MN, Kurul R, Yaşar MF, Turkoglu SA, Avci Ş, Yildiz N. Effect of Leap Motion-based 3D immersive virtual reality usage on upper extremity function in ischemic stroke patients. Arq Neuropsiquiatr. 2019;77(10):681–688.
22. Oh YB, Kim GW, Han KS, Won YH, Park SH, Seo JH, Ko MH. Efficacy of virtual reality combined with real-instrument training for patients with stroke: a randomized controlled trial. Arch Phys Med Rehabil. 2019;100(8):1400–1408.
23. Sun Y, Shao S, Huo F, Zhang H, Ji Q, Zhou R, et al. Effects of a self-directed game-based virtual reality program on upper-limb motor recovery in mild-to-moderate ischemic stroke: a randomized controlled trial. Eur J Phys Rehabil Med. 2026 Feb 2. Online ahead of print. doi: 10.23736/S1973-9087.25.09129-4. PMID: 41627139.
24. Hsieh YW, Lee MT, Hsu YC, Wu KY, Chen CC. Digital mirror therapy and action-observation therapy for chronic stroke: a pilot randomized controlled trial. Occup Ther Int. 2025;2025:8741362.
25. Lee MT, Chen CC, Lu HL, Hsieh YW. Comparisons of three different modes of digital mirror therapy for post-stroke rehabilitation: preliminary results of a randomized controlled trial. Digit Health. 2024;10:20552076241260536.
26. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. BMC Med Res Methodol. 2015;15:58.
27. Balduzzi S, Rücker G, Nikolakopoulou A, Papakonstantinou T, Salanti G, Efthimiou O, Schwarzer G. netmeta: an R package for network meta-analysis using frequentist methods. J Stat Softw. 2023;106(2):1–40.
28. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, Chaimani A, Del Giovane C, Egger M, Salanti G. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. PLoS Med. 2020;17(4):e1003082.
29. Papakonstantinou T, Nikolakopoulou A, Higgins JPT, Egger M, Salanti G. CINeMA: Software for semiautomated assessment of the confidence in the results of network meta-analysis. Campbell Syst Rev. 2020;16:e1080.