

Effectiveness and safety of virtual reality rehabilitation after stroke: an overview of systematic reviews



Silvia Bargerì,^{a,*} Sabrina Scalea,^b Federica Agosta,^{b,c,d} Giuseppe Banfi,^{b,e} Davide Corbetta,^{b,f,g} Massimo Filippi,^{b,c,d,h,i} Elisabetta Sarasso,^{b,c,j} Andrea Turolla,^{k,l} Greta Castellini,^{a,m} and Silvia Gianola^{a,m}



^aIRCCS Istituto Ortopedico Galeazzi, Unit of Clinical Epidemiology, Milan, Italy

^bVita-Salute San Raffaele University, Milan, Italy

^cNeuroimaging Research Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy

^dNeurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy

^eIRCCS Istituto Ortopedico Galeazzi, Milan, Italy

^fDepartment of Rehabilitation and Functional Recovery, IRCCS San Raffaele Scientific Institute, Milan, Italy

^gSchool of Health and Life Sciences, Glasgow Caledonian University, Glasgow, UK

^hNeurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy

ⁱNeurophysiology Service, IRCCS San Raffaele Scientific Institute, Milan, Italy

^jDepartment of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal Child Health, University of Genoa, Genoa, Italy

^kDepartment of Biomedical and Neuromotor Sciences – DIBINEM, Alma Mater Università di Bologna, Bologna, Italy

^lUnit of Occupational Medicine, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy

Summary

Background Virtual reality (VR) is an innovative neurorehabilitation modality that has been variously examined in systematic reviews. We assessed VR effectiveness and safety after cerebral stroke.

Methods In this overview of systematic reviews, we searched eleven databases (Cochrane Database of Systematic Reviews, EMBASE, MEDLINE, SCOPUS, ISI Web of Science, CINAHL, PsycINFO, Pedro, Otseeker, [Healthevidence.org](https://www.healthevidence.org), Epistemonikos) and grey literature from inception to January 17, 2023. Studies eligible for inclusion were systematic reviews published in English that included adult patients with a clinical diagnosis of stroke (acute to chronic phase) undergoing any kind of immersive, semi-immersive or non-immersive VR intervention with or without conventional therapy versus conventional therapy alone. The primary outcome was motor upper limb function and activity. The secondary outcomes were gait and balance, cognitive and mental function, limitation of activities, participation, and adverse events. We calculated the degree of overlap between reviews based on the corrected covered area (CCA). Methodological quality was assessed using the A MeaSurement Tool to Assess systematic Reviews (AMSTAR 2) and the Certainty of Evidence (CoE) using the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) approach. Discordances between results were examined using a conceptual framework based on the Jadad algorithm. This overview is registered with PROSPERO, CRD42022329263.

Findings Of the 58 reviews included ($n = 345$ unique primary studies), 42 (72.4%) had conducted meta-analysis. More than half of the reviews (58.6%) were published between 2020 and 2022 and many (77.6%) were judged critically low in quality by AMSTAR 2. Most reported the Fugl Meyer Assessment scale (FMA-UE) to measure upper limb function and activity. For the primary outcome, there was a moderate overlap of primary studies (CCA 9.0%) with discordant findings. Focusing on upper limb function (FMA-UE), VR with or without conventional therapy seems to be more effective than conventional therapy alone, with low to moderate CoE and probable to definite clinical relevance. For secondary outcomes there was uncertainty about the superiority or no difference between groups due to substantial heterogeneity of measurement scales (eg, methodological choices). A few reviews ($n = 6$) reported the occurrence of mild adverse events.

Interpretation Current evidence suggests that multiple meta-analyses agreed on the superiority of VR with or without conventional therapy over conventional therapy on FMA-UE for upper limb. Clinicians may consider embedding VR

eClinicalMedicine
2023;64: 102220

Published Online xxx
<https://doi.org/10.1016/j.eclinm.2023.102220>

*Corresponding author. IRCCS Istituto Ortopedico Galeazzi, Unit of Clinical Epidemiology, Milan, Italy.

E-mail addresses: silvia.bargerì@grupposandonato.it (S. Bargerì), s.scalea@studenti.univr.it (S. Scalea), agosta.federica@hsr.it (F. Agosta), banfi.giuseppe@univr.it (G. Banfi), corbetta.davide@hsr.it (D. Corbetta), filippi.massimo@hsr.it (M. Filippi), sarasso.elisabetta@hsr.it (E. Sarasso), andrea.turolla@unibo.it (A. Turolla), greta.castellini@grupposandonato.it (G. Castellini), silvia.gianola@grupposandonato.it (S. Gianola).

^{mm}Contributed equally.

technologies into their practice as appropriate with patient's goals, abilities, and preferences. However, caution is needed given the poor methodological quality of reviews.

Funding Italian Ministry of Health.

Copyright © 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Virtual reality; Stroke; Systematic review; Overview

Research in context

Evidence before this study

Among the available conservative interventions for improving stroke-related impairments, Virtual Reality (VR) technology has been promoted as an engaging, interactive, patient-centred and relatively inexpensive neurorehabilitation modality. Thus, we conducted a preliminary search up to February 2022 on PubMed, Embase, Cochrane Database of Systematic Reviews and others using "Virtual Reality" and "Stroke" search terms to explore literature about the effectiveness and safety of VR in people with stroke. From this preliminary search we identified several systematic reviews which resulted in discordances of findings of VR compared to conventional therapy, emphasizing the need of an overview of reviews.

Added value of this study

We provided the first comprehensive overview adopting an innovative framework to assess the discordances between systematic reviews. For upper limb function (Fugl Meyer

Assessment scale) the reviews agreed on the superiority of VR with or without conventional therapy over conventional therapy (low to moderate certainty of evidence, probable to definite clinical relevance). For secondary outcomes, there is still uncertainty about the superiority or no difference of VR due to the heterogeneity of measurement scales (eg, methodological choices).

Implications of all the available evidence

As a safe and potentially efficacious intervention, clinicians should consider learning to use and embed VR technologies into their practice to improve motor function and quality of movement according to patients' needs and preferences. Future research is needed to understand the optimal method of delivering VR (immersive, non-immersive, or semi-immersive) to achieve clinically relevant improvements. Given the current era of personalised medicine, it will be useful to define the technology and the dose of intervention appropriate for each patient.

Introduction

Worldwide, stroke is the second leading cause of death and a major cause of disability,¹ with over 12 million new strokes reported each year.² Given advancing population age, the stroke burden is expected to worsen, with a twofold increase in incidence estimated by 2050.³ Globally, the World Health Organisation (WHO) estimates that at least 62 million stroke survivors⁴ experience severe disability, with diminished ability to self-care and participate in social activities. Upper limb impairment commonly leads to difficulty in reaching, grasping, and manipulating objects, with a reduction in activities of daily living (ADLs) and health-related quality of life.^{5,6} Other common post-stroke impairments are gait and balance and cognitive dysfunction.⁷⁻⁹

With advances in health technologies, the range of interventions for stroke survivors is in continuous expansion. For example, virtual reality (VR) in neurorehabilitation has proved an engaging, interactive, patient-centred, and relatively inexpensive modality to enhance functional recovery.¹⁰⁻¹⁴ VR is defined as an advanced form of human-computer interface that enables users to explore, interact, and immerse themselves in environments that appear and feel similar to real

world objects and events, with real-time and augmented feedback about their performance through multiple sensory channels.^{15,16} VR technology can be combined with a variety of computers, mobile device screens, and head-mounted displays that engage patients in repetitive, intensive, and goal-oriented practice at multiple levels of immersion in a virtual environment.¹⁷ Furthermore, VR can be non-immersive, semi-immersive or immersive depending on the extent to which the setting creates an illusion of "presence in" and "interaction with" the virtual environment.^{18,19} Overall, VR provides task-specific scenarios that can be adapted to a patient's specific needs by targeting key elements for motor learning, such as complexity, specificity, intensity, and salience of practice, motivation, focus of attention, knowledge of performance, and results.^{20,21} As suggested,²¹ VR can also stimulate activity of the mirror neuron system, with the potential to exploit and maximize motor learning mechanisms fundamental for motor recovery after stroke.²²

Systematic reviews of VR-based post-stroke interventions have shown that VR made be superior to²³⁻²⁵ or equal to^{26,27} conventional therapy, engendering discordant conclusions about its effect on upper limb

function and activity. Therefore, we conducted an overview of systematic reviews in which we explored agreement on the effectiveness and the safety of VR technologies for clinical outcomes in stroke survivors to give a comprehensive balance of effects.

Methods

Study design

The present study was carried out in accordance with the Cochrane guidelines for overview of reviews²⁸ and followed the Preferred Reporting Items for Overviews of Reviews (PRIOR) of healthcare interventions guideline.^{29,30} The review protocol was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO) database (CRD42022329263, June 30, 2022).

Eligibility criteria

Eligible for inclusion were systematic reviews with and without meta-analysis of randomised controlled trials (RCTs) or both RCTs and non-randomised studies of interventions (NRSI) (ie, studies that do not use randomization to allocate units³¹) that involved adult patients with a clinical diagnosis of stroke from the acute to the chronic phase who presented motor impairments and were undergoing any kind of immersive, semi-immersive or non-immersive VR intervention. Immersion refers to the extent of the user's perceived "presence in" and "interaction with" the virtual environment, disconnected from the real world by a strong perceptive (but not cognitive) illusion of being in the virtual environment.^{18,19,21} Systems that utilize concave surface projection, head-mounted displays or video capture to immerse the user in a 3D virtual environment are considered immersive, while single-screen projection or desktop displays where users can interact with a computer-generated avatar of themselves in a fictitious environment (while seeing the real world around them) are referred to as non-immersive. A semi-immersive VR falls between non-immersive and fully immersive VR (eg, curved computer screens or VR glasses that enable users to navigate by a visual stimulus within a virtual environment but without other physical sensations).

The VR intervention could be provided with or without conventional therapy compared to any kind of conventional therapy. No restrictions were set regarding time since the stroke event, stroke type (ischemic, haemorrhagic), site (cortical, subcortical) or stroke severity.³²

The primary outcome was upper limb function (arm and hand) (eg, assessed with the Fugl-Meyer Assessment scale for Upper Extremity [FMA-UE] and activity (eg, assessed with the Action Research Arm Test [ARAT], the Wolf Motor Function Test [WMFT], and the Box and Block Test [BBT]).

Secondary outcomes were measures of 1) gait, including walking distance and speed (eg, assessed with the 6 Minutes Walking Test [6MWT], 10 Meters Walk Test [10MWT]), mobility (eg, assessed with the Timed Up and Go Test [TUG]), and balance (eg, assessed with the Berg Balance Scale [BBS], the Functional Reaching Test [FRT]); 2) activities of daily living (ADL) (eg, assessed with the Functional Independence Measure [FIM], the Barthel Index [BI], the Motor Activity Log [MAL], Activities-specific balance confidence [ABC]); 3) participation restriction and quality of life (eg, assessed with the Short Form Health Survey 36 [SF36], EuroQol 5 domains [EQ5D], Stroke Impact Scale [SIS]); 4) cognitive and mental function (eg, assessed with the Mini Mental State Examination [MMSE], Trail Making Test [TMT]); and 5) any adverse events (eg, motion sickness, pain, injury, falls).

Search strategy

A comprehensive search for records was carried out from study inception to 17 January 2023 in the electronic databases: Cochrane Database of Systematic Reviews, EMBASE, MEDLINE, SCOPUS, ISI Web of Science, CINAHL, PsycINFO, Pedro, Otseeker, Healthevidence.org, and Epistemonikos. Grey literature was identified by searching for sites listed on the Grey Matter Checklist (CADTH 2015) in Open Grey (www.opengrey.eu), analysing the references of published reviews, theses, conference proceedings, and other papers, besides contacting the authors of abstracts and review protocols for manuscripts or unpublished data. The PROSPERO database for systematic review protocols was also consulted. The search was restricted to English-language publications and is detailed in [Supplementary File 1](#).

Data collection

Review selection

Two reviewers independently consulted sources by screening for title and abstract. EndNote (The EndNote Team, version 20. Clarivate, Philadelphia, PA, USA; www.endnote.com) and Rayyan (Qatar Computing Research Institute, Qatar; www.rayyan.ai) software were used to manage review selection. Eligible studies were downloaded in full text and assessed for inclusion based on criteria. Conflict was resolved through discussion with a third reviewer. The selection process is illustrated in a PRISMA flow chart.³³

Data extraction and management

Two authors (SS, SB) independently extracted data from the reviews using a predefined data extraction form. The data extraction form was piloted from a small random sample of ten systematic reviews by both reviewers and then re-defined as suggested by all authors.

The following characteristics were extracted: general features (eg, country of corresponding author, contact

author, year of review publication, update of search strategy data, number and study design of the studies in the review, conflict of interest, and sources of funding for the authors), study population (eg, acute, subacute, chronic, mixed), total number of participants, description of interventions (immersive/non-immersive/not described), controls, and outcomes. In addition, we mapped all outcome measurements and their frequencies across the reviews. Common terminology criteria for categorizing adverse events was adopted.³⁴ We also collected information about evidence synthesis (eg, risk of bias assessment, meta-analysis conducted, number of studies and participants presenting at follow-up assessment).

Meta-analysis results were extracted if they included more than one study: mean difference (MD) or standardized mean difference (SMD) for continuous outcomes and odds ratio (OR) or relative risk (RR) with 95% confidence intervals (CI) for binary outcomes and measures of heterogeneity. Data were extracted from the overall meta-analysis without selecting for population subgroups (eg, subgroups haemorrhagic or ischaemic stroke). If combined outcomes of the upper and the lower limbs were reported, we extracted data only when separate estimates for the upper limb were given. The review authors were contacted if information was unclear or missing. A third reviewer resolved disagreement during data collection.

Review and trial quality assessment

Two review authors used the A Measurement Tool to Assess Reviews (AMSTAR) 2 to independently assess the methodological quality of the systematic reviews.³⁵ They also extracted data on the risk of bias of the primary studies included in each systematic review when reported.³⁶ Selection bias (ie, randomization list) and detection bias (outcome assessor blinding) were noted when appraised by any tool (eg, Cochrane risk of bias, PEDRO scale). The percentage of low risk of bias studies out of all primary studies included in each systematic review was collected: a cut-off of 75% was set for assessing the overall quality.³⁷

Overlap between primary studies

Lists of the primary studies included in each systematic review were collated and cross-referenced in a matrix of evidence table to determine the degree of overlap between reviews by means of the GROOVE tool.³⁸ When we examined the matrix for each outcome, we calculated the corrected covered area (CCA) for grading interpretation of overlap as “slight” (CCA 0–5%), “moderate” (CCA 6–10%), “high” (CCA 11–15%) or “very high” (CCA >15%).³⁹

Strategies for data synthesis and statistical analysis

We present the summary of evidence without re-analysing the outcome data. Data were extracted as they were reported in the systematic reviews (with and without meta-analysis) and then reformatted and

presented in text, tables, and figures without pooling data. We examined review characteristics, such as eligibility criteria, to ensure that the systematic reviews investigated similar clinical questions (framed as Population, Intervention, Comparison and Outcome [PICO]).⁴⁰ A visual map of outcome measurements is presented using RAWGraphs 2.0.⁴¹

For the systematic reviews without meta-analyses (ie, qualitative synthesis), we calculated the number of outcome-specific primary studies that found significantly favourable differences (ie, differences that the review reported as being statistically significant in favour of a VR intervention [CIs not overlapping with 0]). These totals are expressed as a percentage of the total number of primary studies included in each review.

For systematic reviews with meta-analyses (ie, quantitative synthesis), we tabulated the overall effects (effect size) as reported by the review authors in MD (if the same scale was used to measure the same conceptual outcome) and SMD (if different scales were used to measure the same conceptual outcome).

Concordance or discordance of effect size

We used a conceptual framework for examining concordance or discordance of effect size of meta-analyses in the direction of the effects (eg, effective interventions [MD on FMA-UE higher than 0 points], ineffective interventions [MD on FMA-UE lower than 0 points], no differences [CI of MD on FMA-UE crossed 0 points]). For this purpose, we created a visual map of the scientific evidence based on bubble plots to display the information of each review as a bubble according to the direction of effect and AMSTAR 2⁴² to compare the concordance or discordance of results.³⁹ A threshold of 80%⁴³ meta-analyses in the same direction was set for concordance of results. When there was discordance in direction of effects between the results of meta-analyses (eg, effective intervention versus no difference), we explored the possible causes of heterogeneity using the Jadad algorithm.⁴⁰ First, we ensured that the reviews asked the same clinical question by looking at the overlap (at least moderate CCA) of primary studies across the reviews.^{28,44,45} Second, we focussed on the same scale measurements⁴⁶ for the same International Classification of Functioning, Disability and Health (ICF) domain⁴⁷ (eg, FMA-UE⁴⁸ for upper limb function, ARAT for upper limb activity). We then applied a sensitivity analysis for each measurement scale without considering: 1) the reviews of both RCT and NRSI, since the best study design to evaluate an intervention should be a well-planned and conducted RCT⁴⁹; 2) reviews rated as critically low quality by AMSTAR 2³⁵ and published before 2020, since there exists a time span between running the search strategy and publication of reporting⁵⁰; 3) meta-analysis with low rate of participants (<200).³⁷ The robustness of evidence (the same direction of results) was discussed across the sources of heterogeneity.⁴⁴

Clinical relevance and interpretation of upper limb function

Owing to clinical heterogeneity of participants (eg, degree of stroke severity and/or stage of stroke recovery) in primary studies in the meta-analyses, we decided not to use established cut-offs of FMA-UE for minimal clinical important differences after stroke.^{4,51-53} Instead, we used a distribution-based approach (statistical method) based on an effect size of 0.5 between VR and conventional therapy groups⁵⁴ to calculate a threshold of a 2.4-point difference on the FMA-UE, which would indicate clinical relevance.⁵⁵ Clinical relevance was thus interpreted by considering the categories described in Man-Son-Hing et al., 2002 (ie, definite, probable, possible, definitely not).⁵⁶ For clinical interpretation of our primary outcome, we plotted the FMA-UE effect sizes along with the clinical relevance threshold. [Supplementary File 1](#) gives details on calculation and interpretation.

Certainty of evidence

Two review authors (SS, GC) independently assessed the Certainty of Evidence (CoE) for the primary outcome by the GRADE approach and adopted the algorithm from Pollock 2016 developed for Cochrane overviews of reviews.³⁷ In this algorithm, each review starts with a ranking of high certainty and is downgraded by one level for four serious methodological concerns: 1) imprecision (ie, sample size of the meta-analysis <200 participants); 2) risk of bias (trial level) in the randomization list in >75% primary studies included in the meta-analysis; 3) inconsistency (high heterogeneity $I^2 >75%$); and 4) methodological quality (review level) in selected critical items of AMSTAR 2.

Ethics

Ethical approval was not applicable for this study, as this was an overview of reviews of existing literature and did not involve direct contact with human subjects.

Role of the funding source

The study was supported by the Italian Ministry of Health (L2059-Revisioni sistematiche della letteratura scientifica in ortopedia, traumatologia e riabilitazione funzionale). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Systematic review and primary study selection

After removing duplicates, we screened 981 records, 866 of which were excluded after reviewing title and abstract. We retrieved the full text of 115 reviews for further detailed assessment, finally including 58 reviews. The study flow chart is illustrated in [Fig. 1](#) while the references of the reviews with the reasons for inclusion or exclusion are presented in [Table S1](#), [Supplementary File 2](#).

Description of reviews

A total of 58 reviews published between 2007 and 2022 included 345 unique primary studies. Most reviews included only RCTs (81%). The first country by number of reviews was China (25.9%). The number of studies per review ranged from 3 to 87, with a median of 18 per review, while the number of participants ranged from 60 to 3540, with a median of 516 patients involved. More than half of the reviews (58.6%) were published in the last 3 years. Overall, 6 (10.3%) declared a conflict of interest and 28 (48.3%) a non-industrial source of funding. Many (69%) included mixed (eg, both subacute and chronic) or chronic onset of stroke (17.2%). Eligible interventions were not fully described in around one third (38%). The main outcome was upper limb function and activity (67.2%). Overall, 42 reviews (72.4%) conducted meta-analyses. The general characteristics of the reviews are reported in [Table 1](#) and in [Table S1A](#) and [B](#), [Supplementary File 3](#) for each review.

Outcome measurements

We mapped all outcome measurements reported in the meta-analyses, 12 of which (30.8%) reported the FMA-UE scale score for upper limb function and 17 (43.6%) reported the BBT, WMFT, and ARAT scores for activity, BBS (n = 13; 27.1%) for gait and balance, FIM for activities of daily living (n = 5; 20.8%), and SIS for participation (n = 2; 50%). Single multiple scales were found for cognitive and mental function ([Fig. 2](#)).

Overlap of primary studies

[Supplementary File 4](#) presents the results of overlap analysis. There was a slight overlap across all reviews (CCA: 4.38%) for all outcomes, a moderate overlap (CCA: 9.0%) for upper limb function and activity outcome ([Fig. 3](#)), a high overlap for gait, mobility and balance (CCA: 10.8%), and a moderate overlap for ADL (CCA: 6.2%). We did not assess overlap for participation, cognitive and mental function due to few data and heterogeneity of outcome measurements.

Methodological quality of reviews

Of the 58 reviews, 45 (77.6%) were judged critically low (including all 11 reviews of both RCT and NRSI), 12 low (20.7%), and one high in methodological quality (1.7%). Most reviews did not report a protocol (69%), did not explain their selection of study design (93.1%), and did not report the reasons for exclusion (89.7%); however, the review authors did apply a satisfactory technique for assessing the risk of bias of primary studies (100%) and appropriate methods for statistical combination of results (69%). Assessment of the reviews by AMSTAR 2 domain and the summary plot are shown in [Figure S1](#) and [Table S1](#), [Supplementary File 5](#), respectively.

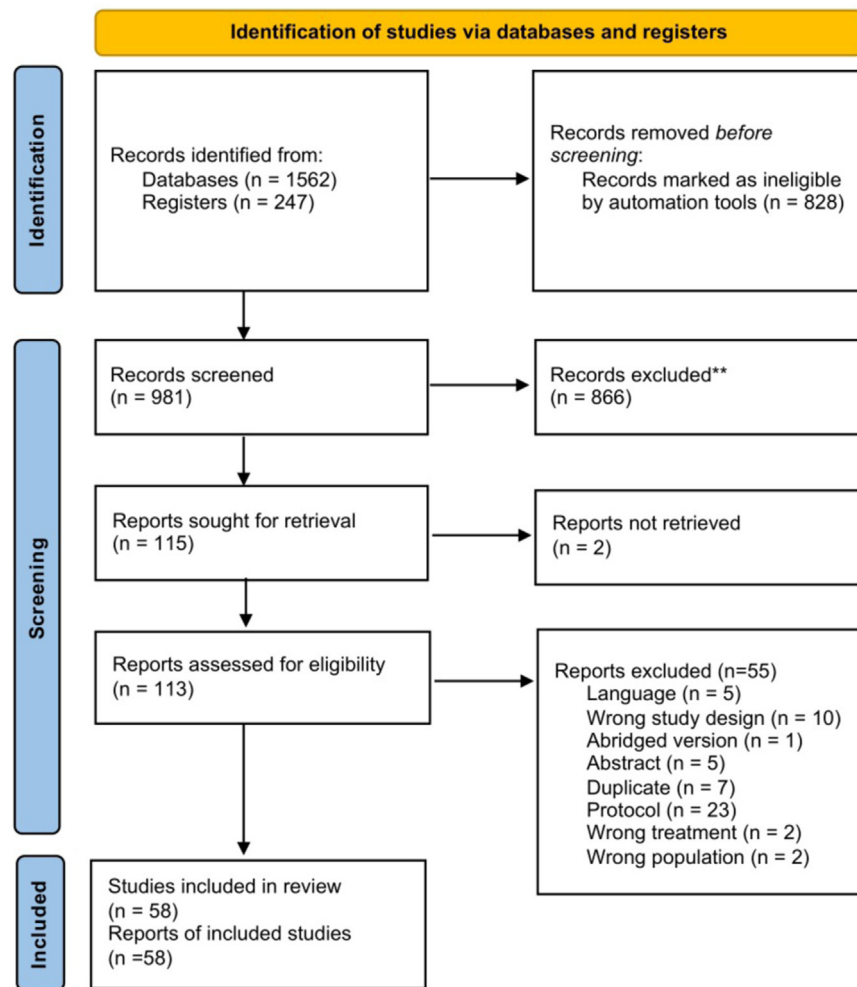


Fig. 1: PRISMA flowchart.

Risk of bias assessment in primary studies

We found that 32 out of the 58 reviews had more than 75% of trials with low risk of selection bias (randomization list described) and 16 had more than 75% of trials with low risk of detection bias (blinded outcome assessor). Assessment of risk of bias in trials is shown in Table S1, Supplementary File 6.

Summary of results

Systematic reviews without meta-analysis

We found primary studies reporting a positive effect of VR among outcomes. The most often reported outcomes were upper limb function and activity (n = 11 reviews) and gait, mobility and balance (n = 7 reviews), with a median proportion of trials reporting statistically significant results of 50% (interquartile range [IQR] 35%–67%) and 88.9% (IQR 58.3%–100%), respectively. Only one review reported findings for each specific outcome measurement (FMA-UE, WMFT, Jebsen Hand

Function Test, ARAT, BBT) for upper limb function and activity but inconsistent findings (range from 0 to 67% of statistically significant trials)⁵⁷ (Supplementary File 7).

Systematic reviews with meta-analysis

The effect size reported by the reviews is presented as bubble plots in Figures S1–S5 and in Tables S1–S5, Supplementary File 8. Strategies to resolve discordances are reported in Supplementary File 9 and the summary of results grouped by outcome measurements in Fig. 4.

Upper limb function and activity

Overall, 39 reviews (67.2%) assessed upper limb function and activity, 23 of which reported quantitative synthesis. Two reviews^{38,59} were not considered for synthesis of results due to different controls (eg, mirror therapy) and incomplete outcome data. The remaining 21 reviews (n = 39 meta-analyses) reported discordant results: 25 meta-analyses (64.1%) reported the superiority

Characteristic	N	%
Year of publication		
2005–2007	1	1.7
2008–2010	0	0.0
2011–2013	3	5.2
2014–2016	9	15.5
2017–2019	11	19.0
2020–2022	34	58.6
Country ^a		
Europe	16	27.6
America	13	22.4
Asia	24	41.4
Africa	1	1.7
Oceania	4	6.9
Population characteristics		
Acute	0	0.0
Subacute	0	0.0
Chronic	10	17.2
Mixed	40	69.0
Not reported	8	13.8
Study design		
SR of RCTs	47	81.0
SR of RCTs and NRSIs	11	19.0
Intervention		
Description reported	36	62.0
Funding		
Yes-non industry	28	48.3
No	16	27.6
Not reported	14	24.1
Outcomes ^b		
UE function and activity	39	67.2
Gait, mobility and balance	36	62.1
ADL	37	63.8
Participation	28	48.3
Cognitive and mental function	20	34.5
Adverse events	18	31.0
Meta-analysis		
Yes	42	72.4
No	16	27.6

Legend: ADL: Activity of Daily Living; NRSI: Non-Randomised Studies of Interventions; RCT: Randomised Controlled Trial; SR: Systematic Review; UE: Upper Extremity. ^aCountry belonging of SR corresponding author. ^bMore than one outcome can be included in each SR.

Table 1: General characteristics of systematic reviews.

of VR with or without conventional therapy over conventional therapy alone, whereas 14 found no differences between groups (Figure S1, Table S1, Supplementary File 8).

Upper limb function (FMA-UE)

The 12 meta-analyses that assessed upper limb function according to FMA-UE showed agreement (91.7%, n = 11) on the superiority of VR with or without conventional therapy compared to conventional therapy

alone. The median of all MDs was 3.8 (IQR 3.0–4.2), which resulted in probable to definite clinical relevance. All effect sizes were plotted without pooling the data (Fig. 5). Sensitivity analysis confirmed the direction of effects (Supplementary File 9).

Upper limb activity (ARAT, BBT, WMFT)

Seventeen meta-analyses assessed upper limb activity (ARAT n = 3, BBT n = 10, WMFT n = 4), with discordant results on each scale. According to a planned sensitivity analysis, the most updated and higher quality review of RCTs⁶⁰ showed no differences between groups on any scale of the ARAT, BBT, and WMFT (Fig. 4 and Supplementary File 9).

Composite measures of upper limb function and activity

The 10 meta-analyses that assessed upper limb function and activity using composite measures overwhelmingly agreed (80%, n = 8) on the superiority of VR with or without conventional therapy compared to conventional therapy alone (Fig. 4). Sensitivity analysis confirmed the direction of effects (Supplementary File 9).

Gait, mobility and balance

Overall, 36 reviews (62.1%) assessed gait and balance; 18 of which reported quantitative synthesis (n = 48 meta-analyses) with discordant results: 31 meta-analyses (64.6%) reported the superiority of VR with or without conventional therapy over conventional therapy alone, whereas 17 found no difference between groups (Table S2, Supplementary File 8).

Walking distance and speed (6MWT, 10MWT)

Six meta-analyses assessed gait distance and speed (n = 1 6MWT, n = 5 10MWT) and reported discordant results on the 10MWT. The meta-analysis of 6MWT found VR superior to conventional therapy (Fig. 4). A sensitivity analysis was not applied since all updated reviews of RCTs were rated as critically low and all meta-analyses involved fewer than 200 participants (Supplementary File 9).

Mobility (TUG)

Thirteen meta-analyses assessed mobility and reported discordant results for TUG. All updated reviews were rated as critically low in methodological quality. According to a planned sensitivity analysis, the meta-analysis of RCTs involving the highest number of participants in all reviews (100%, n = 4) agreed on the superiority of VR over conventional therapy (Fig. 4 and Supplementary File 9).

Balance (BBS, FRT)

Sixteen meta-analyses assessed balance (n = 13 BBS, n = 3 FRT) and reported discordant results for BBS and FRT. The updated reviews were rated as critically low in methodological quality. According to a planned

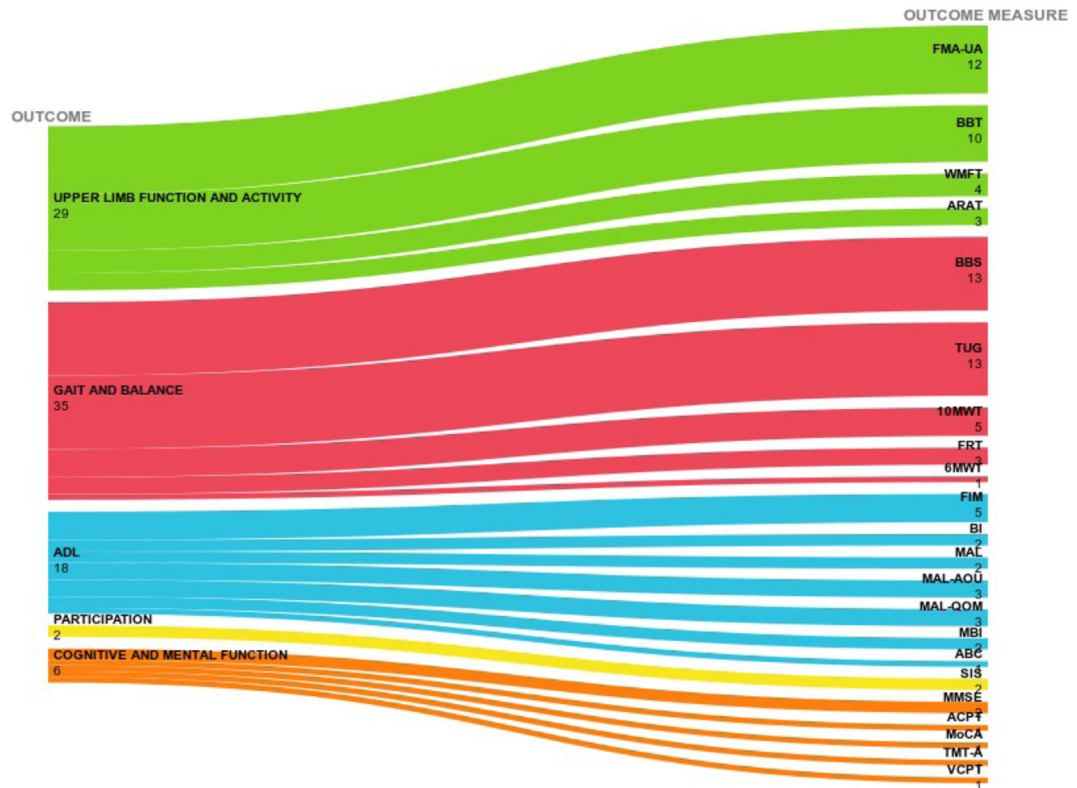


Fig. 2: Outcome measurements scale reported by the meta-analyses. **Legend:** 6MWT: Six walking test; 10MWT: 10 Meters Walk Test; ABC: Activities-specific Balance Confidence Scale; ACPT: Auditory Continuous Performance Test; ARAT: Action Research Arm Test; BBS: Berg Balance Scale; BBT: Box and Block Test; BI: Barthel Index; FIM: Functional Independent Measure; FMA-UE Fugl-Meyer Assessment-Upper Extremity; FRT: Functional Reaching Test; MAL: Motor Activity Log; MBI: Modified Barthel Index; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; SIS: Stroke Impact Scale; TMT: Trail Making Test; TUG (Timed Up and Go); VCPT: Visual continuous performance test; Wolf Motor Function Test.

sensitivity analysis, the meta-analysis of the RCTs involving the highest number of participants in all reviews (100%, n = 4) agreed on the superiority of VR over conventional therapy for BBS (Fig. 4), while no sensitivity analysis was applied for FRT (all reviews were critically low and not updated, while all meta-analyses had fewer than 200 participants) (Supplementary File 9).

Composite measures

Thirteen meta-analyses assessed gait, mobility and balance using composite measures and reported discordant results. According to a planned sensitivity analysis, the most updated and better quality review of RCTs⁶¹ found VR superior to conventional therapy in improving mobility/gait (TUG, 10MWT), whereas no difference between groups for balance measures (BBS, FRT) (Fig. 4 and Supplementary File 9).

ADL

Overall, 37 reviews (63.8%) assessed ADL; 16 of which reported a quantitative synthesis (n = 24 meta-analyses)

with discordant results: 11 (45.8%) found VR with or without conventional therapy superior to conventional therapy, whereas 13 found no differences between groups (Table S3, Supplementary File 8).

ADL (FIM, BI, MAL, ABC)

Eighteen meta-analyses assessed ADL (n = 5 FIM, n = 4 BI/modified BI, n = 8 MAL, n = 1 ABC) and agreed on the superiority of VR over conventional therapy based on FIM (80%, n = 4) while no differences between groups were found according to the BI (100%, n = 4) and MAL (87.5%, n = 7) (Fig. 4). Sensitivity analysis confirmed the direction of effects. The only meta-analysis of ABC found no differences between groups (Supplementary File 9).

Composite measures

Six meta-analyses assessed ADL using composite measures and agreed (100%, n = 6) on the superiority of VR with or without conventional therapy over conventional therapy (Fig. 4). Sensitivity analysis confirmed the direction of effects (Supplementary File 9).

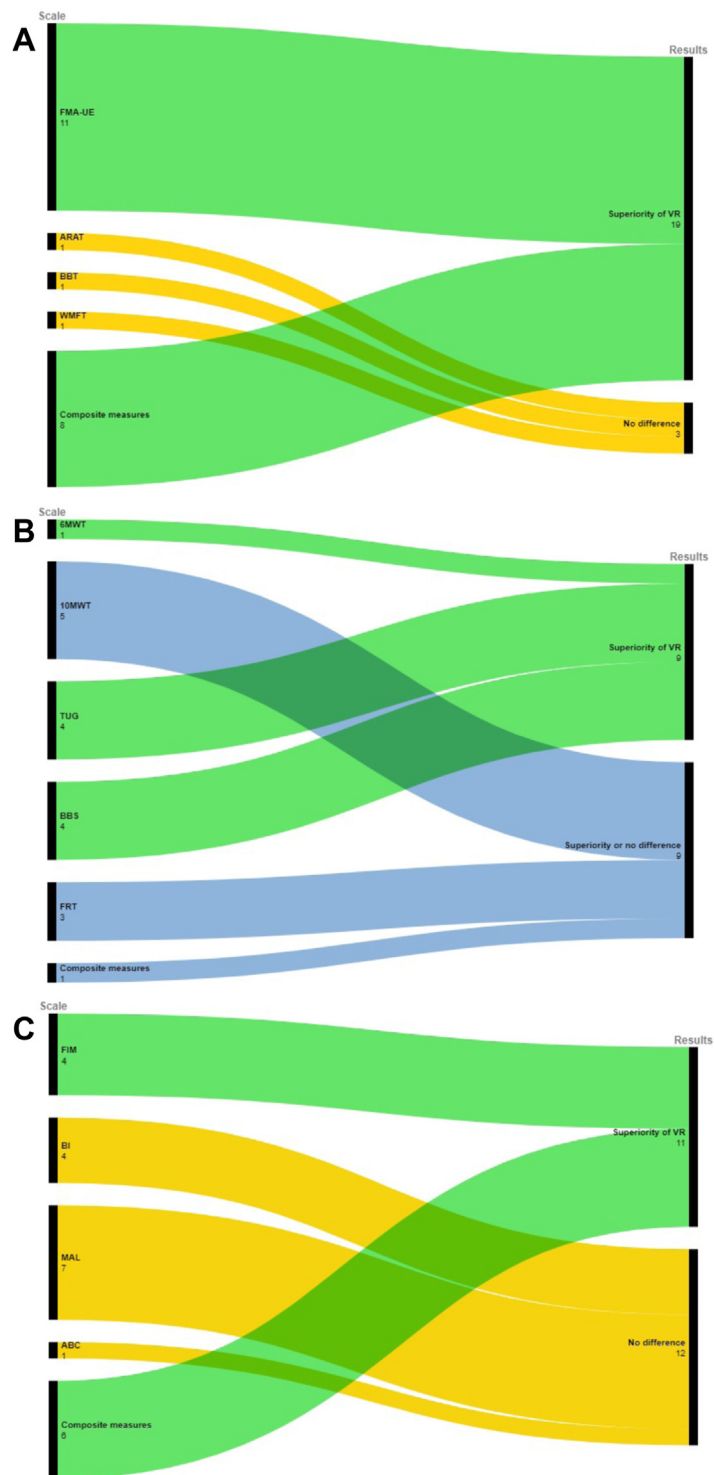


Fig. 4: Summary of results grouped by outcome measurements for A) Upper limb function and activity, B) Gait, mobility and balance, C) ADL. Node size is proportional to the number of meta-analyses assessing the corresponding outcome **Legend:** 6MWT: Six walking test; 10MWT: 10 Meters Walk Test; ABC: Activities-specific Balance Confidence Scale; ADL: Activities of Daily Living; ARAT: Action Research Arm Test; BBS: Berg Balance Scale; BBT: Box and Block Test; BI: Barthel Index; FIM: Functional Independent Measure; FMA-UE: Fugl-Meyer Assessment-Upper Extremity; FRT: Functional Reaching Test; MAL: Motor Activity Log; TUG (Timed Up and Go); Wolf Motor Function Test (WMFT).

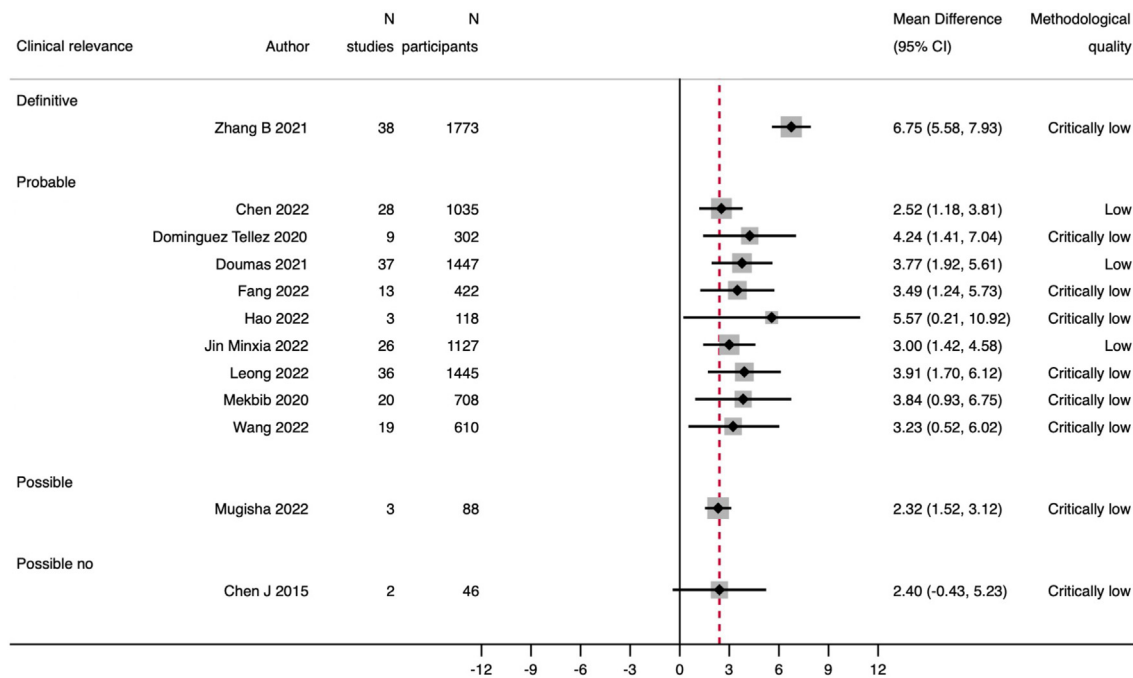


Fig. 5: Effect size expressed as mean difference of FMA-UE in the systematic reviews with meta-analyses. Red dotted lines indicate the threshold for clinical relevance. Legend: FMA-UE Fugl-Meyer Assessment-Upper Extremity.

therapy on TUG as a measure of gait mobility and BBS as measure of balance. In ADL outcome, we found superiority for FIM and no differences between groups for other scales. These can be due to (i) heterogenous delivered scales (eg, BI, modified BI), (ii) some items also covered in other outcomes (eg, MAL correlates with ARAT,⁶² ABC correlate with BBS⁶³) and (iii) imprecision of effects (less than 200 participant) for other ADL scales.

The sparse evidence for participation and cognitive and mental function resulted in either superiority or no difference between groups.

In addition, VR can be considered a safe intervention with few mild adverse events. Overall, these results warrant caution in light of the poor methodological quality of many of the reviews. Only one high-quality Cochrane systematic review²⁶ reported equal or positive effects on upper limb function and activity, respectively, of VR alone or in combination with conventional therapy compared to conventional therapy. Of note is that this review was published in 2017. In the meantime, there has been a noticeable rise in the number of SRs on VR in rehabilitation in the past three years, with nearly one-third of our sample published in 2022 alone.

Our findings for a positive effect of VR on various clinical outcomes compared to conventional therapy are shared by previous overviews on VR in neuro-rehabilitation,^{64,65} with the difference, however, that we

provide a clinical interpretation of discordant results based on an innovative framework for analysis of discordant effects across meta-analyses and evidence graded according to the GRADE approach and clinical relevance.

Our findings suggest a clinically relevant effect of VR regardless of how delivered, dose, intensity, and co-interventions on upper limb function as assessed with FMA-UE. Arm activity as assessed with the BBT, WMMF, and ARAT does not seem to be affected by VR motor rehabilitation. These findings support the hypothesis that VR may help to improve the recovery of upper limb motor function and quality of movement (FMA-UE). A possible explanation is that frequent and salient on-line feedback about performance based on objective measures of movement may be a determinant factor in improving motor control. In addition, VR may be useful for patients who need to recover not only the ability to perform a task but also refine the quality of movement during task execution. Moreover, VR might have a beneficial effect on mobility, balance and ADL. In brief, for stroke survivors who have difficulty in mobility and balance, VR training with the feedback provided by a force platform or a treadmill may be advantageous, since no serious adverse effects were observed.⁶⁶ Care should be taken, however, to minimise the risk of falls during balance training.⁶⁷

Clinicians should consider how to incorporate VR into post-stroke rehabilitation interventions for recovering

motor function and in day-to-day activity training to increase participant engagement and practice on a functional relevance task. Furthermore, given that therapists tend to overestimate the time spent in active task practice,⁶⁸ the use of an objective measurement of activity, often incorporated in VR systems, can ensure that the amount of practice is appropriate for a given patient. In this context, depending on a patient's potential for motor recovery and the aim of training, tailored exercises with feedback (visual, auditory, tactile) can be planned to improve performance (what a person does in his/her usual environment—feedback on an activity) or to improve personal capacity (what a person can do in a standard environment—feedback on a motor function, task or action).⁴⁷

Our overview disclosed an increase in the number of studies investigating the same questions, especially in the last three years. Before launching a systematic review or an RCT, authors should check registries (eg, PROSPERO, clinicaltrials.gov). There is ample room for improving study quality; review authors should consult guidelines for conducting⁶⁹ and reporting⁷⁰ studies when designing a high-quality review and journal editors should demand high methodological standards for publication. Nearly 40% of the trials were published in the last five years, adding new evidence that needs to be synthesized in systematic reviews for updating the status of available evidence.

A future area of focus is how to clearly determine the best way to administer VR (immersive, non-immersive or semi-immersive) to achieve clinically relevant improvement and to identify specific cohorts of stroke patients (according to clinical severity and/or time since stroke) that could benefit most from therapy. In the present era of gender medicine^{71,72} and individualized therapy, the type of technology and the dose of intervention appropriate for each patient need to be defined. For instance, a higher level of arousal, emotion, and/or stress may increase female susceptibility to simulator sickness (a form of motion sickness caused by interaction with a simulated environment) and discomfort during VR.⁷³ Finally, researchers should have a clear idea about VR technology acceptance, adherence, usability, and costs and whether these outcomes might change when VR is offered with or without co-interventions, additional treatments, advice or other interventions that may affect the outcome of interest and treatment compliance.

To our knowledge, this is the broadest overview of systematic reviews in stroke rehabilitation performed to date. We registered the study protocol and followed published guidelines for reporting and conduct.^{28–30} Nevertheless, some limitations should be acknowledged. Although we employed an in-depth search strategy in a large number of databases, inclusion was limited to papers published in English and five reviews are awaiting assessment for other languages. Even if the

overlap of primary studies for each outcome ranged from moderate to high, most reviews assessed heterogeneous modalities of VR interventions compared to heterogeneous control groups (eg, conventional therapy, usual care) on 22 different outcome measurements (across five outcomes). Also, since the study population included stroke patients in the sub-acute to the chronic phase, ceiling or floor effects on some measurement scales cannot be ruled out.^{63,74,75} Most scales were meta-analysed together by reviews authors using SMD (eg, FMA-UE, ARAT, BBT) adding further heterogeneity and less comparability of review findings. For this reason, we focused on primary outcome (upper limb function) as a proxy for effectiveness and applied an outcome-centered approach with FMA-UE for the clinically interpretation of results.

We did not attempt to quantitatively synthesize the results using indirect comparison techniques, such as network meta-analysis; therefore, we cannot assess the best modality of VR intervention (immersive, non-immersive or semi-immersive). Categorization of interventions (eg, immersive versus non-immersive) was not performed since more than one-third of the reviews did not describe the intervention in detail, which is a symptom of poor planning and suboptimal reporting of systematic reviews and RCTs in neurorehabilitation.⁷⁶ A template checklist for intervention description and replication (TIDieR) can help trialists and reviewers avoid this pitfall in the interpretation of findings in clinical practice.⁷⁷ Despite missing descriptions, and assuming a certain degree of heterogeneity across interventions, the reviews were consistent about keeping a broad definition of VR.

We did not examine comparisons of co-interventions (VR alone or VR combined with other interventions), duration or intensity, as our description was primarily derived from the details provided in the reviews. Nevertheless, it should be noted that undergoing VR and conventional therapy may extend overall therapy duration, so it is possible that the observed effects might be attributed to co-interventions that increase repetitive activities.⁷⁸ Conventional therapy is usually offered to stroke patients and can be considered as an add-on therapy.

We also found some meta-analyses that used the SMD measure though they explored a typical outcome scale (eg, FMA-UE). To allow for usability of the entire body of evidence, we had to back translate the SMDs to a typical scale (eg, 0–66 for FMA-UE). Systematic review authors should prioritize the MD instead of the SMD when performing a meta-analysis on the same scale when trials report final values and changes. Final value and change scores should not be conflated in standardized mean differences, since the difference in standard deviation does not reflect differences in measurement scale but rather the differences in the reliability of measurements.⁷⁹

Finally, our interpretation of clinical relevance was driven by the assumption that conventional therapy in trials reached the MID and because we determined the threshold of statistical relevance for VR and conventional therapy using a distribution-based approach. MID is known to vary depending on the anchor or the distribution approach.^{80,81}

Our data indicate that multiple meta-analyses agreed on the superiority of VR with or without conventional therapy over conventional therapy on FMA-UE for upper limb, with low to moderate certainty of evidence and probable to definite clinical relevance. For secondary outcomes, there is still uncertainty about the superiority or no difference of VR with or without conventional therapy due to the heterogeneity of measurement scales (eg, methodological choices). As a safe intervention, clinicians should consider learning to use and embed VR technologies into their practice and adapt VR interventions according to patients' needs and preferences. Caution in the interpretation of findings is warranted, however, given the poor methodological quality of the reviews.

Contributors

SG, GC, SB conceptualised this review. SB, SS was responsible for library resources and supervision of the resource retrieval from databases. SB and SS conducted the relevant searches. SB, SG, GC were responsible for data collection and analysis. SG, GC, SB were responsible for methodology, validation and supervision. SB, SS, DC, AT, SG, GC prepared the initial draft of the paper, and SB, SS, FA, GB, DC, MF, ES, AT, SG, GC contributed to the development and refinement of subsequent drafts. All authors read and approved the final manuscript. SB, GC and SG have accessed and verified the underlying data.

Data sharing statement

The full dataset is freely available online in OSF (<https://osf.io/wtymk/>), a secure online repository for research data.

Declaration of interests

FA received research grant for San Raffaele Hospital from Italian Minister of Health, AriSLA (Fondazione Italiana di Ricerca per la SLA), the European Research Council, Foundation Research on Alzheimer Disease; consulting fees, Payment for meeting and lectures from Biogen Idec, Roche, Zambon and Italfarmaco. DC received research grant for San Raffaele Hospital from Italian Ministry of Health Grant: Ricerca Finalizzata 2018 (GR-2018-12366005) for Virtual Reality in Parkinson disease. MF received research grant for San Raffaele Hospital from Biogen Idec, Merck Serono, Novartis, Roche, Italian Ministry of Health, Fondazione Italiana Sclerosi Multipla, and AriSLA; consulting fees, Payment for meeting and lectures from Biogen Idec, Merck Serono, Novartis, Roche, Italian Ministry of Health, Fondazione Italiana Sclerosi Multipla, and AriSLA, Bayer, Biogen, Celgene, Chiesi Italia SpA, Eli Lilly, Genzyme, Janssen, Merck-Serono, Neopharmed Gentili, Novartis, Novo Nordisk, Roche, Sanofi, Sanofi-Genzyme, Takeda, Bristol-Myers Squibb, and TEVA. In addition, Participation on a Data Safety Monitoring Board or Advisory Board from Alexion, Biogen, Bristol-Myers Squibb, Merck, Novartis, Roche, Sanofi, Sanofi-Aventis, Sanofi-Genzyme, Takeda. ES received research grant for San Raffaele Hospital from Italian Ministry of Health Grant: Ricerca Finalizzata 2021 (GR-2021-12374005) for Parkinson disease and Italian Ministry of Health Grant: PNRR 2022 (PNRR-MAD-2022-12376826) for functional motor disorders. DC, AT, GC, and SG report Italian Physiotherapy Association Board membership (unpaid). All other authors declare no competing interests.

Acknowledgements

The authors wish to thank Kenneth Adolf BRITSCH, Avicenna Snc, the external English service for language revision and assistance in manuscript preparation and Dr Francesco Bonetti for his help in data extraction. The study was supported by the Italian Ministry of Health (L2059-Revisioni sistematiche della letteratura scientifica in ortopedia, traumatologia e riabilitazione funzionale).

Appendix B. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclinm.2023.102220>.

References

- 1 Katan M, Luft A. Global burden of stroke. *Semin Neurol*. 2018;38(2):208–211.
- 2 Collaborators GBDS. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol*. 2021;20(10):795–820.
- 3 Gorelick PB. The global burden of stroke: persistent and disabling. *Lancet Neurol*. 2019;18(5):417–418.
- 4 Virtual reality, a neuroergonomic and neurorehabilitation tool for promoting neuroplasticity in stroke survivors: a systematic review with meta-analysis. In: *Proceedings of the 21st congress of the international ergonomics association*. IEA; 2021:495–508.
- 5 Bleyenheuft Y, Gordon AM. Precision grip in congenital and acquired hemiparesis: similarities in impairments and implications for neurorehabilitation. *Front Hum Neurosci*. 2014;8:459.
- 6 Nichols-Larsen DS, Clark PC, Zeringue A, Greenspan A, Blanton S. Factors influencing stroke survivors' quality of life during subacute recovery. *Stroke*. 2005;36(7):1480–1484.
- 7 Al-Qazzaz NK, Ali SH, Ahmad SA, Islam S, Mohamad K. Cognitive impairment and memory dysfunction after a stroke diagnosis: a post-stroke memory assessment. *Neuropsychiatr Dis Treat*. 2014;10:1677–1691.
- 8 Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. *Lancet*. 2011;377(9778):1693–1702.
- 9 Ursin MH, Bergland A, Fure B, et al. Gait and balance one year after stroke; relationships with lesion side, subtypes of cognitive impairment and neuroimaging findings—a longitudinal, cohort study. *Physiotherapy*. 2019;105(2):254–261.
- 10 Jin R, Pillozzi A, Huang X. Current cognition tests, potential virtual reality applications, and serious games in cognitive assessment and non-pharmacological therapy for neurocognitive disorders. *J Clin Med*. 2020;9(10):3287.
- 11 Imbimbo I, Coraci D, Santilli C, et al. Parkinson's disease and virtual reality rehabilitation: cognitive reserve influences the walking and balance outcome. *Neurol Sci*. 2021;42(11):4615–4621.
- 12 Shen J, Xiang H, Luna J, et al. Virtual reality-based executive function rehabilitation system for children with traumatic brain injury: design and usability study. *JMIR Serious Games*. 2020;8(3):e16947.
- 13 Mumford N, Wilson PH. Virtual reality in acquired brain injury upper limb rehabilitation: evidence-based evaluation of clinical research. *Brain Inj*. 2009;23(3):179–191.
- 14 Taylor MJ, McCormick D, Shawis T, Impson R, Griffin M. Activity-promoting gaming systems in exercise and rehabilitation. *J Rehabil Res Dev*. 2011;48(10):1171–1186.
- 15 Weiss P, Kizony R, Feintuch U, Katz N. Virtual reality in neuro-rehabilitation. In: Selzer M, Cohen L, Gage F, Clarke S, Duncan P, eds. *Textbook of neural repair and rehabilitation*. Cambridge University Press; 2006:182–197.
- 16 Riva G. Virtual reality in psychotherapy: review. *Cyberpsychol Behav*. 2005;8(3):220–230. discussion 31–40.
- 17 Merians AS, Jack D, Boian R, et al. Virtual reality-augmented rehabilitation for patients following stroke. *Phys Ther*. 2002;82(9):898–915.
- 18 Slater M. Immersion and the illusion of presence in virtual reality. *Br J Psychol*. 2018;109(3):431–433.
- 19 Riva G, Mantovani F, Gaggioli A. Presence and rehabilitation: toward second-generation virtual reality applications in neuropsychology. *J Neuroeng Rehabil*. 2004;1(1):9.
- 20 Dockx K, Bekkers EM, Van den Bergh V, et al. Virtual reality for rehabilitation in Parkinson's disease. *Cochrane Database Syst Rev*. 2016;12(12):Cd010760.

- 21 Verschure PF. Neuroscience, virtual reality and neuro-rehabilitation: brain repair as a validation of brain theory. *Annu Int Conf IEEE Eng Med Biol Soc.* 2011;2011:2254–2257.
- 22 Cameirão MS, Badia SB, Duarte E, Frisoli A, Verschure PF. The combined impact of virtual reality neurorehabilitation and its interfaces on upper extremity functional recovery in patients with chronic stroke. *Stroke.* 2012;43(10):2720–2728.
- 23 Jin M, Pei J, Bai Z, et al. Effects of virtual reality in improving upper extremity function after stroke: a systematic review and meta-analysis of randomized controlled trials. *Clin Rehabil.* 2022;36(5):573–596.
- 24 Zhang B, Li D, Liu Y, Wang J, Xiao Q. Virtual reality for limb motor function, balance, gait, cognition and daily function of stroke patients: a systematic review and meta-analysis. *J Adv Nurs.* 2021;77(8):3255–3273.
- 25 Aminov A, Rogers JM, Middleton S, Caeyenberghs K, Wilson PH. What do randomized controlled trials say about virtual rehabilitation in stroke? A systematic literature review and meta-analysis of upper-limb and cognitive outcomes. *J Neuroeng Rehabil.* 2018;15(1):29.
- 26 Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. *Cochrane Database Syst Rev.* 2017;11(11):Cd008349.
- 27 Chen J, Jin W, Zhang XX, Xu W, Liu XN, Ren CC. Tele-rehabilitation approaches for stroke patients: systematic review and meta-analysis of randomized controlled trials. *J Stroke Cerebrovasc Dis.* 2015;24(12):2660–2668.
- 28 Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: overviews of reviews. In: Higgins JPT, Thomas J, Chandler J, et al., eds. *Cochrane handbook for systematic reviews of interventions version 6.3 (updated February 2022)*; Cochrane; 2022. Available from: www.training.cochrane.org/handbook.
- 29 Pollock M, Fernandes RM, Pieper D, et al. Preferred Reporting Items for Overviews of Reviews (PRIOR): a protocol for development of a reporting guideline for overviews of reviews of healthcare interventions. *Syst Rev.* 2019;8(1):335.
- 30 Gates M, Gates A, Pieper D, et al. Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement. *BMJ.* 2022;378:e070849.
- 31 Reeves BC, Deeks JJ, Higgins J, Shea B, Tugwell P, Wells GA. Chapter 24: including non-randomized studies on intervention effects. Available from: <https://training.cochrane.org/handbook/current/chapter-24>.
- 32 Bernhardt J, Hayward KS, Kwakkel G, et al. Agreed definitions and a shared vision for new standards in stroke recovery research: the Stroke Recovery and Rehabilitation Roundtable taskforce. *Int J Stroke.* 2017;12(5):444–450.
- 33 Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev.* 2021;10(1):39.
- 34 National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE). Available from: https://ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/ctcae_v3.pdf. Accessed September 9, 2023.
- 35 Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358:j4008.
- 36 Lunny C, Brennan SE, McDonald S, McKenzie JE. Toward a comprehensive evidence map of overview of systematic review methods: paper 2-risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. *Syst Rev.* 2018;7(1):159.
- 37 Pollock A, Farmer SE, Brady MC, et al. An algorithm was developed to assign GRADE levels of evidence to comparisons within systematic reviews. *J Clin Epidemiol.* 2016;70:106–110.
- 38 Pérez-Bracchiglione J, Meza N, Bangdiwala SI, et al. Graphical representation of overlap for Overviews: GROOVE tool. *Res Synth Methods.* 2022;13(3):381–388.
- 39 Pieper D, Antoine SL, Mathes T, Neugebauer EA, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. *J Clin Epidemiol.* 2014;67(4):368–375.
- 40 Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ.* 1997;156(10):1411–1416.
- 41 Mauri M, Elli T, Caviglia G, Uboldi G, Azzi M. RAWGraphs: a visualisation platform to create open outputs. In: *Proceedings of the 12th biannual conference on Italian SIGCHI chapter.* New York, NY, USA: ACM; 2017:28:1–28:5. <https://doi.org/10.1145/3125571.3125585>.
- 42 Mlake-Lye IM, Mak S, Lee J, et al. Massage for pain: an evidence map. *J Altern Complement Med.* 2019;25(5):475–502.
- 43 Belur J, Tompson L, Thornton A, Simon M. Interrater reliability in systematic review methodology: exploring variation in coder decision-making. *Socio Methods Res.* 2018;50(2):837–865.
- 44 Moja L, Fernandez del Rio MP, Banzi R, et al. Multiple systematic reviews: methods for assessing discordances of results. *Intern Emerg Med.* 2012;7(6):563–568.
- 45 Lucenteforte E, Moja L, Pecoraro V, et al. Discordances originated by multiple meta-analyses on interventions for myocardial infarction: a systematic review. *J Clin Epidemiol.* 2015;68(3):246–256.
- 46 Mayo-Wilson E, Fusco N, Li T, et al. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. *J Clin Epidemiol.* 2017;86:39–50.
- 47 Organization WH. *Towards a common language for functioning, disability, and health: ICF. The international classification of functioning, disability and health.* 2002.
- 48 Fugl-Meyer AR, Jaasko 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.
- 49 Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' guides to the medical literature: XXV. Evidence-Based medicine: principles for applying the users' guides to patient care. Evidence-based medicine working group. *JAMA.* 2000;284(10):1290–1296.
- 50 Yoshii A, Plaut DA, McGraw KA, Anderson MJ, Wellik KE. Analysis of the reporting of search strategies in Cochrane systematic reviews. *J Med Libr Assoc.* 2009;97(1):21–29.
- 51 Hiragami S, Inoue Y, Harada K. Minimal clinically important difference for the Fugl-Meyer assessment of the upper extremity in convalescent stroke patients with moderate to severe hemiparesis. *J Phys Ther Sci.* 2019;31(11):917–921.
- 52 Arya KN, Verma R, Garg RK. Estimating the minimal clinically important difference of an upper extremity recovery measure in subacute stroke patients. *Top Stroke Rehabil.* 2011;18(Suppl 1):599–610.
- 53 Aromataris E, Fernandez R, Godfrey C, Holly C, Khalil H, Bhatrasakoon P. *Chapter 10: umbrella reviews.* 2020.
- 54 Cohen J. *Statistical power analysis for the behavioral sciences.* Hillsdale, New Jersey: Lawrence Erlbaum Associates; Routledge; 1988.
- 55 Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J.* 2007;7(5):541–546.
- 56 Man-Son-Hing M, Laupacis A, O'Rourke K, et al. Determination of the clinical importance of study results. *J Gen Intern Med.* 2002;17(6):469–476.
- 57 Berard CH, Cavalier M, Godart N. Use of virtual reality-based therapy in the field of chronic stroke rehabilitation to improve hand function: asystematic review. *Sport i Turystyka Środkowoeuropejskie Czasopismo Naukowe.* 2022;5(3):115–138.
- 58 Cabrales SX Jr. *Exploring the effectiveness of virtual reality therapy compared to mirror therapy in treating individuals with upper extremity dysfunction following stroke: a meta analysis.* Fresno: California State University; 2018.
- 59 Ekechukwu END, Nzeakuba IC, Dada OO, et al. *Virtual reality, a neuroergonomic and neurorehabilitation tool for promoting neuroplasticity in stroke survivors: a systematic review with meta-analysis.* Congress of the International Ergonomics Association; 2022. Springer; 2022:495–508.
- 60 Chen J, Or CK, Chen T. Effectiveness of using virtual reality-supported exercise therapy for upper extremity motor rehabilitation in patients with stroke: systematic review and meta-analysis of randomized controlled trials. *J Med Internet Res.* 2022;24(6):e24111.
- 61 Garay-Sánchez A, Suarez-Serrano C, Ferrando-Margeli M, Jimenez-Rejano JJ, Marcén-Román Y. Effects of immersive and non-immersive virtual reality on the static and dynamic balance of stroke patients: a systematic review and meta-analysis. *J Clin Med.* 2021;10(19):4473.
- 62 Chen P, Liu TW, Tse MMY, Lai CKY, Tsoh J, Ng SSM. The predictive role of hand section of fugl-meyer assessment and motor activity Log in action research arm test in people with stroke. *Front Neurol.* 2022;13:926130.
- 63 Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. *Phys Ther.* 2008;88(5):559–566.
- 64 Voinescu A, Sui J, Stanton Fraser D. Virtual reality in neuro-rehabilitation: an umbrella review of meta-analyses. *J Clin Med.* 2021;10(7):1478.

- 65 Pollock A, Farmer SE, Brady MC, et al. Interventions for improving upper limb function after stroke. *Cochrane Database Syst Rev*. 2014;2014(11):Cd010820.
- 66 Li Z, Han XG, Sheng J, Ma SJ. Virtual reality for improving balance in patients after stroke: a systematic review and meta-analysis. *Clin Rehabil*. 2016;30(5):432–440.
- 67 Australian and New Zealand Living Clinical Guidelines for Stroke Management - Chapter 5 of 8: Rehabilitation. v10.0 published on 09/12/2022. <https://app.magicapp.org/#/guideline/6659>
- 68 Kaur G, English C, Hillier S. Physiotherapists systematically overestimate the amount of time stroke survivors spend engaged in active therapy rehabilitation: an observational study. *J Physiother*. 2013;59(1):45–51.
- 69 Higgins JPT, Thomas J, Chandler J, et al., eds. *Cochrane handbook for systematic reviews of interventions version 6.3 (updated February 2022)*; Cochrane; 2022. Available from: www.training.cochrane.org/handbook.
- 70 Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg*. 2021;88:105906.
- 71 Heidari S, Babor TF, De Castro P, Tort S, Curno M. Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. *Res Integr Peer Rev*. 2016;1:2.
- 72 Accounting for sex and gender makes for better science. *Nature*. 2020;588(7837):196.
- 73 Grassini S, Laumann K. Are modern head-mounted displays sexist? A systematic review on gender differences in HMD-mediated virtual reality. *Front Psychol*. 2020;11:1604.
- 74 Lin JH, Hsueh IP, Sheu CF, Hsieh CL. Psychometric properties of the sensory scale of the Fugl-Meyer Assessment in stroke patients. *Clin Rehabil*. 2004;18(4):391–397.
- 75 Kristersson T, Persson HC, Alt Murphy M. Evaluation of a short assessment for upper extremity activity capacity early after stroke. *J Rehabil Med*. 2019;51(4):257–263.
- 76 Feller D, Pedri C, Gozzer P, Innocenti T, Trentin F. The reporting of somatic sensory training interventions in individuals following a stroke is suboptimal: a systematic review and meta - research study. *Am J Phys Med Rehabil*. 2023;102(8):701–706.
- 77 Vassar M, Page MJ, Glasbey J, et al. Evaluation of the completeness of intervention reporting in Cochrane surgical systematic reviews using the TIDieR-SR checklist: a cross-sectional study. *BMJ Evid Based Med*. 2021;26(2):51–52.
- 78 Silver B. Virtual reality versus reality in post-stroke rehabilitation. *Lancet Neurol*. 2016;15(10):996–997.
- 79 Deeks JHJ, Altman DG. Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, et al., eds. *Cochrane handbook for systematic reviews of interventions version 6.3 (updated February 2022)*; Cochrane; 2022. Available from: www.training.cochrane.org/handbook.
- 80 Mouelhi Y, Jouve E, Castelli C, Gentile S. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health Qual Life Outcomes*. 2020;18(1):136.
- 81 Wells G, Beaton D, Shea B, et al. Minimal clinically important differences: review of methods. *J Rheumatol*. 2001;28(2):406–412.