JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

Journal of Movement Mechanics & Biomechanics Science (JMMBS)

Volume 3, Issue 2 (2026) | Article Type: **Review Article — Critical Narrative Review**
DOI: **10.66078/jmmbs.2026.v3i2.021** | ISSN 3070-3662 (Online) | JMMBS ID: JMMBS-2026-021-REV-v3-i2
Received: 20 October 2026 | Accepted: 10 November 2026 | Published: 11 November 2026

Wearable Assessment of Upper Limb Recovery After Stroke: From Movement Signals to Meaningful Endpoints

Sumit Chaudhary Khobey, BSc[1,*], Gergely Bíró, MEd[2], Sunita Malhotra, MSc[3], Pankaj Mehta, MSc[4], Josh Smith, PhD[5]

[1] Strength & Conditioning Department, BodyGNTX Rehab Institute, USA

[2] Sport Analytics, University of Szeged, Hungary; Teacher of Physical Education

[3] Clinical Research & Ethical Board Coordinator, MMSx Authority Institute, USA

[4] Department of Exercise Science, GFFI Fitness Academy, USA

[5] Independent Researcher, Clinical Rehabilitation & Sports Biomechanics, Canada

* Corresponding author

Corresponding Author
Sumit Chaudhary Khobey, BSc — correspondence via the JMMBS editorial office: info@jmmbs.org

ORCID iDs
Sumit Chaudhary Khobey — https://orcid.org/0009-0008-1523-6493
Gergely Bíró — https://orcid.org/0009-0002-3711-0628
Sunita Malhotra — https://orcid.org/0009-0007-2279-9764
Pankaj Mehta — https://orcid.org/0009-0009-5920-0158
Josh Smith — https://orcid.org/0009-0008-1412-7076

Source(s) of Support
No specific funding was received for this work.

Disclosure of Interest
The authors declare no competing interests.

Ethics and Consent
This review synthesises published literature. It reports no new participant recruitment, identifiable participant material or individual-level experimental data collection; ethical approval was therefore not required.

Data Availability
No new experimental dataset was generated. Sources are identified in the reference list. Source-verification and search records are held by the JMMBS editorial office and are available on reasonable request.

Use of AI-Assisted Tools
AI-assisted tools supported literature discovery, language editing, reference checking and document preparation. No participant data or experimental results were generated. The authors reviewed and verified the content and take full responsibility for the published article.

Number of Tables/Figures
4 Tables, 4 Figures · 36 References

Licence
© 2026 The Author(s). Published by JMMBS under the Creative Commons Attribution 4.0 International Licence (CC BY 4.0).

How to Cite

> Khobey, S. C., Bíró, G., Malhotra, S., Mehta, P., & Smith, J. (2026). Wearable assessment of upper limb recovery after stroke: From movement signals to meaningful endpoints. *Journal of Movement Mechanics & Biomechanics Science, 3*(2). https://doi.org/10.66078/jmmbs.2026.v3i2.021

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

ABSTRACT

Background: Wearable sensors characterise upper limb movement beyond the clinic, but movement quantity, functional activity, motor restitution, and participation are distinct outcomes.

Objective: To appraise the measurement chain connecting wearable signals to meaningful outcomes after stroke and propose a framework for endpoint specification and validation.

Methods: Targeted literature retrieval and citation verification were completed on 10 November 2026. Foundational measurement studies, primary accelerometry research, recent digital outcome studies, and methodological guidance were selected purposively. Evidence was organised by construct, reference, setting, responsiveness, and intended use. No pooled estimates or formal certainty grades were produced.

Synthesis: Bilateral accelerometry complements clinical capacity tests but cannot alone identify successful grasp, voluntary contribution, or restoration of joint coordination. Laterality ratios can improve when less affected arm activity declines. Classifiers, orientation metrics, finger sensors, and composite measures offer complementary approaches, with limits in activity labels, population coverage, and external validation. A recent digital arm performance scale supports further development; trial-efficiency estimates based on power analysis do not establish patient benefit or endpoint qualification. Measurement error, meaningful change, missingness, and usability require separate evaluation.

Conclusion: Credible endpoints require a defined patient-relevant construct, reproducible processing, validation in the intended setting, and a prespecified interpretation of change. Integrating wearable performance measures with impairment, capacity, and movement-strategy assessments supports more precise rehabilitation research and decision-making.

Keywords: stroke rehabilitation; upper limb; accelerometry; digital endpoints; motor recovery; measurement validity

HIGHLIGHTS

- Movement quantity, functional activity, motor restitution and participation are distinct constructs requiring distinct evidence.
- A laterality ratio can rise when less affected arm activity falls; absolute limb values must be retained.
- Classifier accuracy depends on validation design and activity prevalence; infrequent purposeful activity is easily overestimated.
- Credible digital endpoints need a patient-relevant construct, locked processing, an estimand and a prespecified interpretation of change.

1. INTRODUCTION

Upper limb rehabilitation after stroke seeks changes that matter in daily life: reaching for a cup, managing clothing, preparing food or returning to valued work. A person may regain the capacity to perform a task during examination without using that ability at home. Conversely, assistance, environmental adaptation or a compensatory strategy can improve independence without restoring the previous movement pattern. Assessment therefore requires clarity about what changed and why. The distinction between impairment, task performance under standardised conditions and everyday activity is foundational to clinical interpretation [1].

Wearable sensing extends the observation window from minutes to days and can reveal patterns unavailable during a clinic visit. This advantage is substantial, but it does not eliminate measurement problems. Continuous data can precisely quantify the wrong construct, conceal unrepresentative wear or reward an algorithm for detecting motion that has little relevance

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

to an intended outcome. The central question is therefore not whether a wrist sensor records movement accurately. It is whether the complete measurement process supports a specified inference about an individual or a treatment effect.

This review examines that process from physical signal to clinical endpoint. It distinguishes established empirical findings from methodological recommendations and gives particular attention to newer approaches that attempt to move beyond activity counts. Its practical contribution is a structured endpoint specification, together with an interpretation framework for discordant clinical and wearable changes. Neither framework is presented as an independently validated instrument or a replacement for clinical judgment.

2. EVIDENCE IDENTIFICATION AND CRITICAL APPRAISAL

Targeted web searches and citation checks were completed on 10 November 2026 using combinations of stroke, upper limb, accelerometry, functional arm use, wearable validation, digital endpoint, meaningful change and the titles of relevant foundational studies. Retrieval prioritised PubMed records, accessible full texts on publisher or research repository sites, and official methodological guidance. Reference lists of accessible papers informed additional targeted retrieval. No lower publication date limit was imposed; recent searches specifically included 2025 and 2026.

Selection was purposive. Studies were retained when they clarified a measurement construct, tested wearable outputs against a specified reference, examined daily performance longitudinally or addressed a practical requirement for endpoint use. Critical appraisal considered participant spectrum, setting, reference labels, independence of development and evaluation samples, aggregation, missing data and applicability to the proposed use. No formal risk of bias scale or evidence certainty system was applied across the heterogeneous study designs. The review does not claim exhaustive database coverage, duplicate screening, a prospectively registered protocol or a pooled accuracy estimate.

Bibliographic details were checked against primary publication records. Accessible full texts informed detailed appraisal of pivotal classification and orientation studies. For the 2026 digital arm performance scale and finger versus wrist study, the published indexed abstracts supported only the explicitly reported design and findings; their complete published methods and supplements were not available for this appraisal. No methodological details were transferred from an earlier preprint into the published study assessment. A source register records access level and claim support. These limits constrain the strength of the review's conclusions about the newest measures.

3. DEFINING THE OUTCOME BEFORE CHOOSING THE SENSOR

3.1 Impairment and standardised capacity

Motor impairment refers to limitations in movement generation and control. The Fugl-Meyer assessment provides a standardised examination of poststroke sensorimotor performance [2]. Capacity concerns what a person can accomplish under defined test conditions; the Action Research Arm Test assesses tasks involving grasp, grip, pinch and gross movement [1, 3]. These measures answer different questions from continuous monitoring. A short standardised examination can be appropriate when the aim is to control task demands, assistance and scoring conditions.

Neither a clinical total score nor a wearable output should be interpreted as a direct measure of neural repair. Recovery and compensation require definitions at the neural, movement and functional levels [4]. For example, increased trunk displacement can enable a successful reach despite persistent elbow control limitations. A person may also deliberately choose a compensatory strategy because it is useful and acceptable. The clinical value of that strategy and the biological mechanism underlying improvement are separate questions.

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

3.2 Daily performance and participation

Performance concerns activity in the usual environment. Participation concerns involvement in meaningful life situations. More movement does not necessarily imply greater independence, success, comfort or satisfaction. Equally, a reduction in redundant movements could accompany more efficient task completion. Sensor endpoints should therefore state whether they target movement duration, purposeful activity, task accomplishment, movement quality or participation. Table 1 separates these constructs and the evidence needed for each.

The Stroke Recovery and Rehabilitation Roundtable provides shared terminology and recommendations for standardised recovery measurement [5, 6]. A suitable assessment battery can pair an impairment measure and a capacity test with daily monitoring and selected movement analysis. Concordance supports a coherent interpretation; discordance can identify a clinically useful question. It should not automatically be resolved by declaring the continuously recorded measure more valid than the clinical examination.

Table 1. Constructs and the observations needed to support them

Construct	Suitable observation	Interpretation boundary
Motor impairment	Standardised sensorimotor examination, such as Fugl-Meyer assessment	Does not quantify everyday use or establish neural repair by itself
Task capacity	Standardised task performance, such as Action Research Arm Test	Does not establish transfer to ordinary daily activity
Daily movement	Wrist activity duration, magnitude and bilateral contribution	Does not identify purpose, successful grasp or independence
Functional activity	Validated classification against an explicit activity taxonomy	Meaning depends on included activities and reference labels
Movement strategy	Task-specific joint and trunk kinematics with appropriate controls	Does not by itself establish valued participation
Participation	Patient relevant activity, independence and assistance observations	Does not identify the neural or mechanical mechanism of improvement

Conceptual synthesis informed by clinical assessment and recovery frameworks [1, 2, 3, 4, 5, 6]. Measures should be selected for the question rather than combined into an undefined recovery label.

4. FROM PHYSICAL MEASUREMENT TO BILATERAL ACTIVITY SUMMARIES

4.1 What the sensor records

An accelerometer measures specific force in its local coordinate frame. The measured signal depends on translational acceleration, gravity, sensor orientation, bias and noise. In a simplified rigid attachment model,

$$\mathbf{f}_s(t) = \mathbf{R}_{ws}(t)^T\left[\mathbf{a}_w(t) - \mathbf{g}_w\right] + \mathbf{b}(t) + \boldsymbol{\varepsilon}(t).$$

Here $\mathbf{f}_s$ is the three axis sensor output in acceleration units; R_{ws} maps sensor coordinates to the world frame; $\mathbf{a}_w$ is sensor translational acceleration; $\mathbf{g}_w$ is gravitational acceleration; b is bias; and ε represents noise. The sign of the gravity term depends on the stated instrument convention. This expression does not include a guarantee of perfect attachment or calibration. Experimental evaluation of acceleration processing shows why gravity and movement cannot be perfectly separated by simple filtering during arbitrary rotation [7].

Consequently, a scalar acceleration magnitude is not a direct measure of muscle activation, joint moment, grip force, movement intent or successful object manipulation. Similar wrist motion can arise from different shoulder, elbow and trunk com-

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

binations. Passive assistance and arm swing can generate substantial signals, while holding a cup steadily can involve meaningful activity with little acceleration. These are consequences of the measurement geometry. A study concerned with restitution of joint coordination needs additional validated kinematic observations; a study of hand function may need information beyond the wrist [4, 6].

Different shoulder, elbow and trunk configurations can bring the wrist toward a similar target; interpreting wrist movement as a recovery strategy therefore requires additional kinematic and task observations (Figure 1).

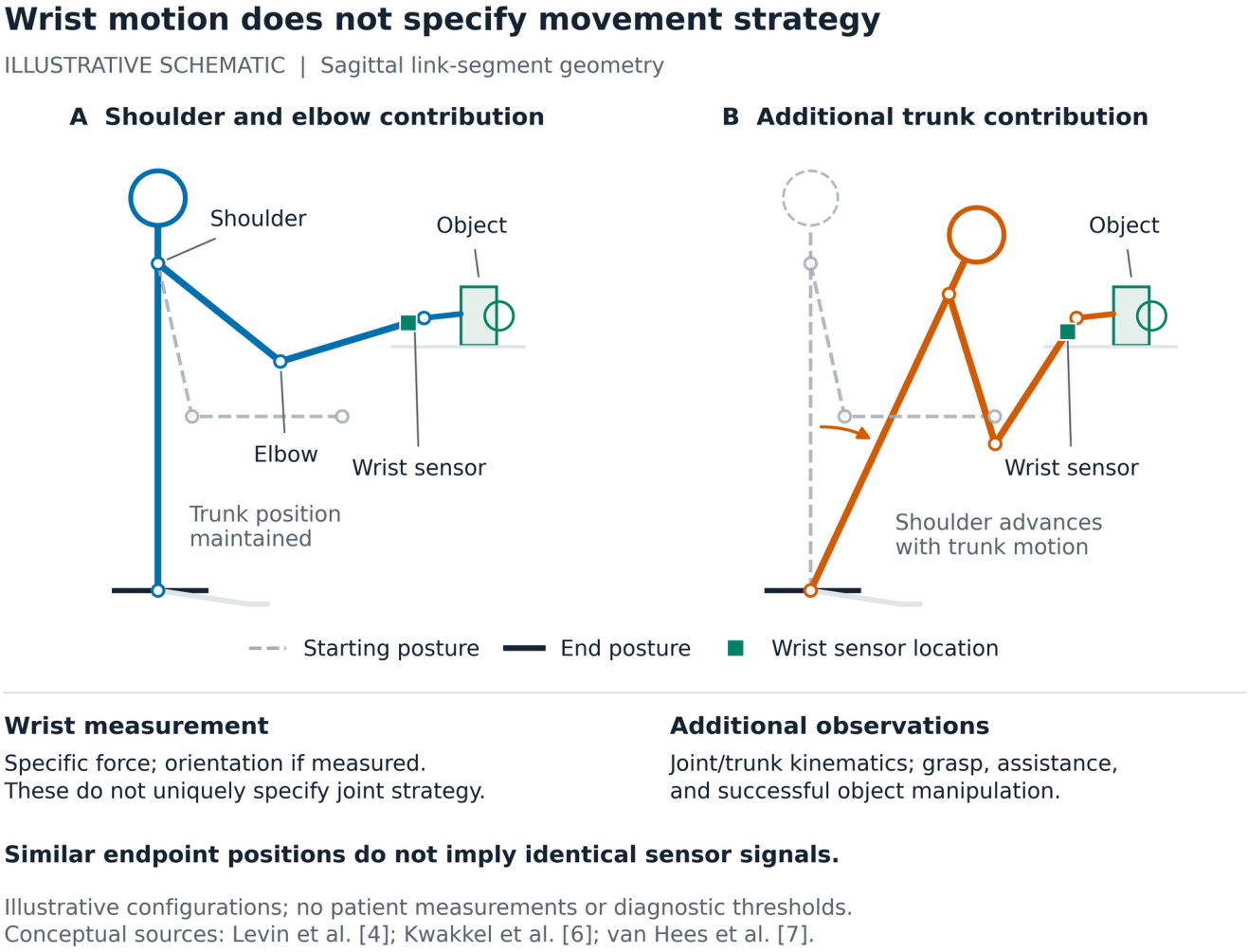


Figure 1. Wrist motion does not specify movement strategy. Illustrative sagittal link-segment configurations showing two ways of reaching toward an object. A, shoulder and elbow movement with the trunk position maintained; B, additional trunk motion advancing the shoulder. Dashed lines denote a shared starting configuration and solid lines the end configurations. Segment lengths are preserved within the schematic; the geometry is not measured participant data. Green squares locate a wrist sensor. Similar endpoint positions do not imply identical acceleration or orientation signals. Wrist motion alone does not uniquely determine joint coordination, assistance, grasp or successful object manipulation. Additional validated kinematics and task observations are needed for those inferences. Neither configuration is a diagnosis or a prescribed optimal strategy. Conceptual synthesis informed by recovery definitions, measurement recommendations and accelerometry principles [4, 6, 7].

4.2 What established accelerometry studies support

Early ambulatory studies demonstrated that wearable recordings could quantify arm activity outside the laboratory in chronic stroke [8]. The EXCITE validation study included 169 participants, 3–9 months after stroke, with mild to moderate arm impairment. Bilateral recordings over three days showed reliability and associations with observational and interview

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

measures of use [9]. These findings support activity monitoring in the studied context; they do not prove that every detected movement is purposeful or that the same processing performs equally in profound paresis.

Reference observations in adults without neurological impairment help describe ordinary bilateral patterns [10]. In a study of 74 nondisabled adults and 48 adults with chronic stroke, Bailey and colleagues used approximately one day of bilateral recordings to characterise intensity and relative limb contributions. Some participants with different clinical capacity scores showed similar daily activity patterns [11]. This supports complementary assessment rather than interchangeability of clinical scores and wearable outputs.

Studies early after stroke and in subacute populations also relate arm activity to impairment and self-care, while showing that these domains are not synonymous [12, 13]. Longitudinal evidence is particularly informative: a secondary analysis of 78 participants receiving task-specific training found no improvement in six accelerometry performance measures over the intervention, despite gains in capacity among some participants [14]. This finding concerns that intervention, population and measurement approach. It does not establish that daily performance cannot improve or that accelerometry detects every valuable change.

4.3 Ratios and aggregation require explicit definitions

Let A and L denote affected and less affected arm active time, measured in the same units over simultaneous valid wear. A duration use ratio is $U = A/L$ when $L > 0$. Let $m_A(t)$ and $m_L(t)$ denote comparable processed activity magnitudes. A bilateral magnitude is $B(t) = m_A(t) + m_L(t)$, while an instantaneous log magnitude ratio can be defined as $M(t) = \ln\left[m_A(t)/m_L(t)\right]$ when both magnitudes are positive. Published protocols describe related duration and magnitude summaries [15]. These quantities are not interchangeable: one emphasises duration, another overall intensity and another relative contribution.

The denominator creates a critical interpretive dependency. In an illustrative example, A = 2 hours and L = 5 hours gives U = 0.40. If A remains 2 hours while L falls to 3 hours, U rises to 0.67 without increased affected arm activity. This calculation is illustrative, not patient data. Symmetry can also coexist with bilateral inactivity. Reports should therefore retain both limb values, total activity and valid wear alongside a ratio. A zero denominator makes U undefined; it should not be silently replaced with a recovery score.

The less affected arm is a within-person comparator, not a guaranteed healthy reference. Its activity can change with dominance, bilateral task demands, assistance and the daily environment. Investigators should record pre-stroke handedness and affected side, describe how the comparator is used, and distinguish a longitudinal change in the denominator from a change in the target limb. Absolute values and bilateral distributions can preserve information lost by symmetry scores. These are analytical safeguards rather than a presumption that the less affected limb is clinically abnormal in every participant.

Study specifications must define filtering, threshold, epoch length, simultaneous wear rules and the treatment of zero or near-zero magnitudes. Adding a small constant to a log ratio changes its behaviour and requires justification. Aggregating a ratio of daily totals is not equivalent to averaging instantaneous ratios. Recording inactive periods, including sleep, or restricting analyses to active epochs also changes the target quantity. A reproducible name for an endpoint must be accompanied by its equation and implementation rather than inferred from a familiar label.

For the normalised difference R = (A - L)/(A + L), the same score can arise at different activity totals, and an increase can occur when only less affected activity declines; these analytical properties also require the absolute components to be retained (Figure 2).

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

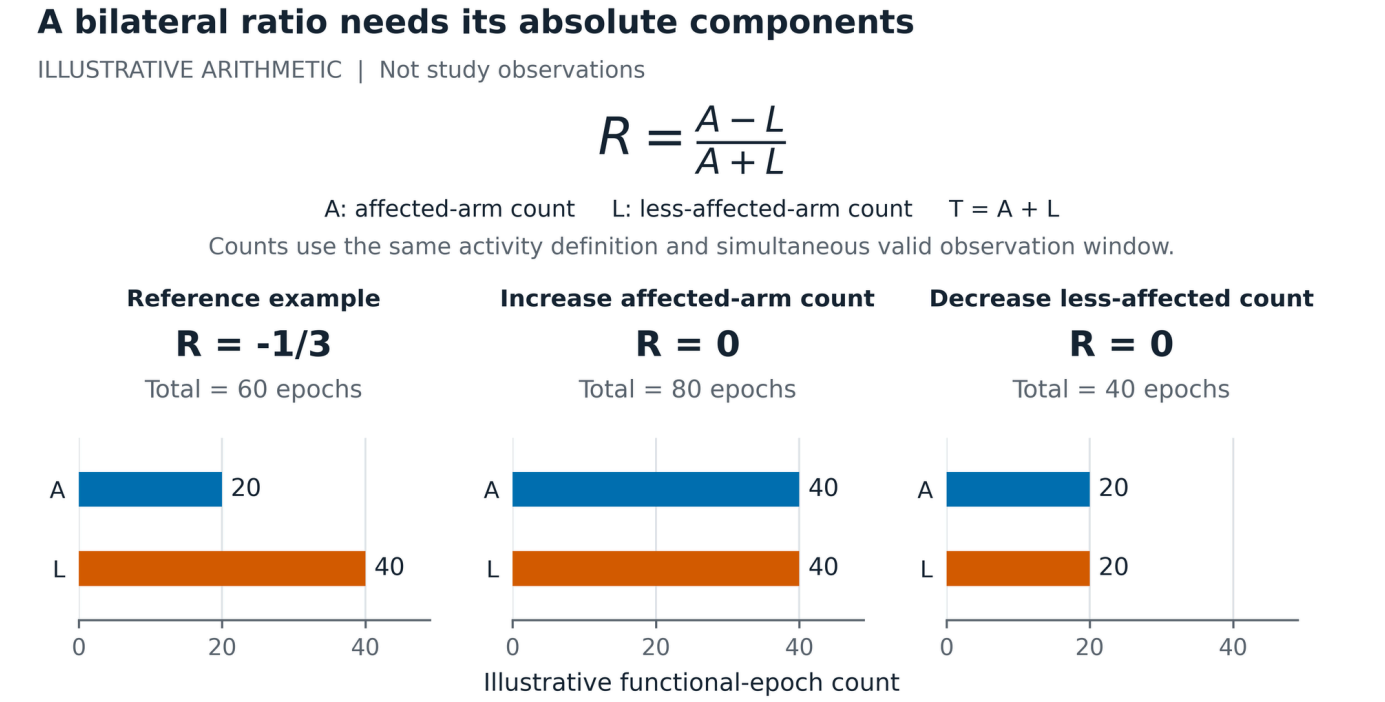


Figure 2. A bilateral ratio needs its absolute components. Illustrative arithmetic for the normalised difference R = (A - L)/(A + L), where A and L are affected- and less-affected-arm functional-epoch counts obtained with the same activity definition and simultaneous valid observation window. These values are not study observations. The reference example gives R = -1/3 from A = 20 and L = 40. Increasing A to 40 or decreasing L to 20 both yield R = 0, although the totals become 80 and 40 epochs, respectively. Multiplying both components by a positive factor preserves R. Thus a higher score or the same score can conceal different absolute activity patterns; neither establishes recovery. Nonnegative inputs require a positive total. This normalised difference [19] is distinct from the duration use ratio A/L [15].

5. BEYOND ACTIVITY COUNTS

5.1 Orientation and functional activity classification

Orientation based methods seek to reduce contamination from whole body motion. Leuenberger and colleagues evaluated a gross arm movement approach using wrist inertial measurements in 10 stroke survivors monitored for 48 hours [16]. Such methods identify movements meeting a specified geometric definition; they do not automatically include all useful upper limb behaviour. A forearm orientation criterion may miss low amplitude manipulation or sustained stabilisation. Its suitability depends on whether these exclusions match the intended construct.

Machine learning can classify patterns that simple thresholds cannot separate. Lum and colleagues compared classifiers against synchronized video labels in a simulated community setting. Their best reported paretic limb accuracy differed between within-person evaluation and evaluation across participants, approximately 93% versus 74% [17]. This difference il-

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

lustrates why personalisation, validation design and unfamiliar participants matter. It should not be read as a performance guarantee for unsupervised daily monitoring.

Published classifier performance differs between within-participant and across-participant evaluation, especially for the paretic arm (Figure 3).

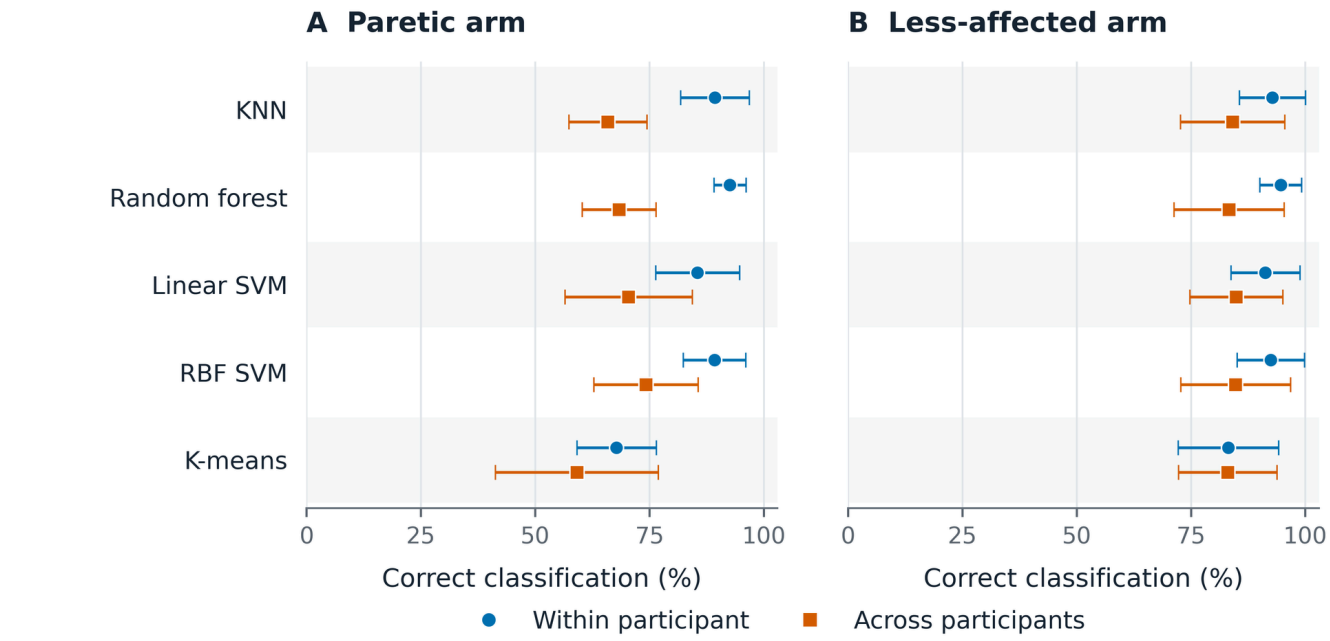


Figure 3. Classifier accuracy by validation design. Classification accuracy in ten participants after stroke. A, paretic arm; B, less-affected arm, measured in the same participants. Points and whiskers show the means and standard deviations (SDs) published in Lum et al. (2020), Table 2 [17]. Within-participant evaluation used stratified five-fold cross-validation; across-participant evaluation used leave-one-participant-out cross-validation. Functional versus nonfunctional labels came from video annotation during a simulated community activity session and included gesturing as functional activity. The source labels dispersion as SD without a more specific table-level variance definition; it has not been converted into a confidence interval. The common axis extends slightly above 100% to retain the reported mean-plus-SD endpoint of 100.02%; this is not an observed accuracy above 100%. KNN, k-nearest neighbors; SVM, support vector machine; RBF, radial basis function. These estimates do not represent independent unrestricted home validation.

The classification target itself can differ across studies. Lum and colleagues counted gesturing as functional activity, whereas the 2025 home comparison by Nieto and colleagues classified gestures outside their prehension focused functional category [17, 18]. Neither choice is automatically wrong; the difference means that superficially similar functional use outputs may estimate different behaviours. Annotation manuals should address stabilisation, gestures, assisted actions, unsuccessful attempts and ambiguous or occluded periods. Reference labels are operational judgments rather than error free observations of intent.

Nieto and colleagues studied four people completing activity scripts in a simulated apartment and in their homes. The work supports feasibility of testing transfer between settings, but within-person models and repeated scripts do not establish performance in new people during unrestricted daily life [18]. Their exclusion of blocks without sufficiently homogeneous labels

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

also matters when interpreting reported accuracy. Transitional and ambiguous periods remain part of ordinary behaviour and require explicit handling in eventual deployment.

5.2 Recent developments and their evidential boundaries

The 2026 FuncUseRatio study included 23 healthy participants and 25 people after stroke. Ratio agreement with video reached ICC 0.96 (95% confidence interval 0.90–0.98). Yet, using a 15° amplitude threshold and ±50° orientation range, absolute FuncUse underestimated video-labelled functional epochs by approximately 25% [19]. Strong ratio agreement therefore does not establish accuracy of the constituent limb values. Stroke assessments involved supervised occupational therapy; reference labels excluded stabilisation and inactivity, and parameters were explored within the study. Independent evaluation should test the selected rule in unrestricted settings and assess behaviours relevant to the intended endpoint. Its normalised difference ratio, $(F_A - F_L)/(F_A + F_L)$ for affected and less affected functional movement counts with a positive sum, also differs from A/L.

At matching algorithm parameters, agreement differs for the functional-use ratio and its absolute limb components (Figure 4).

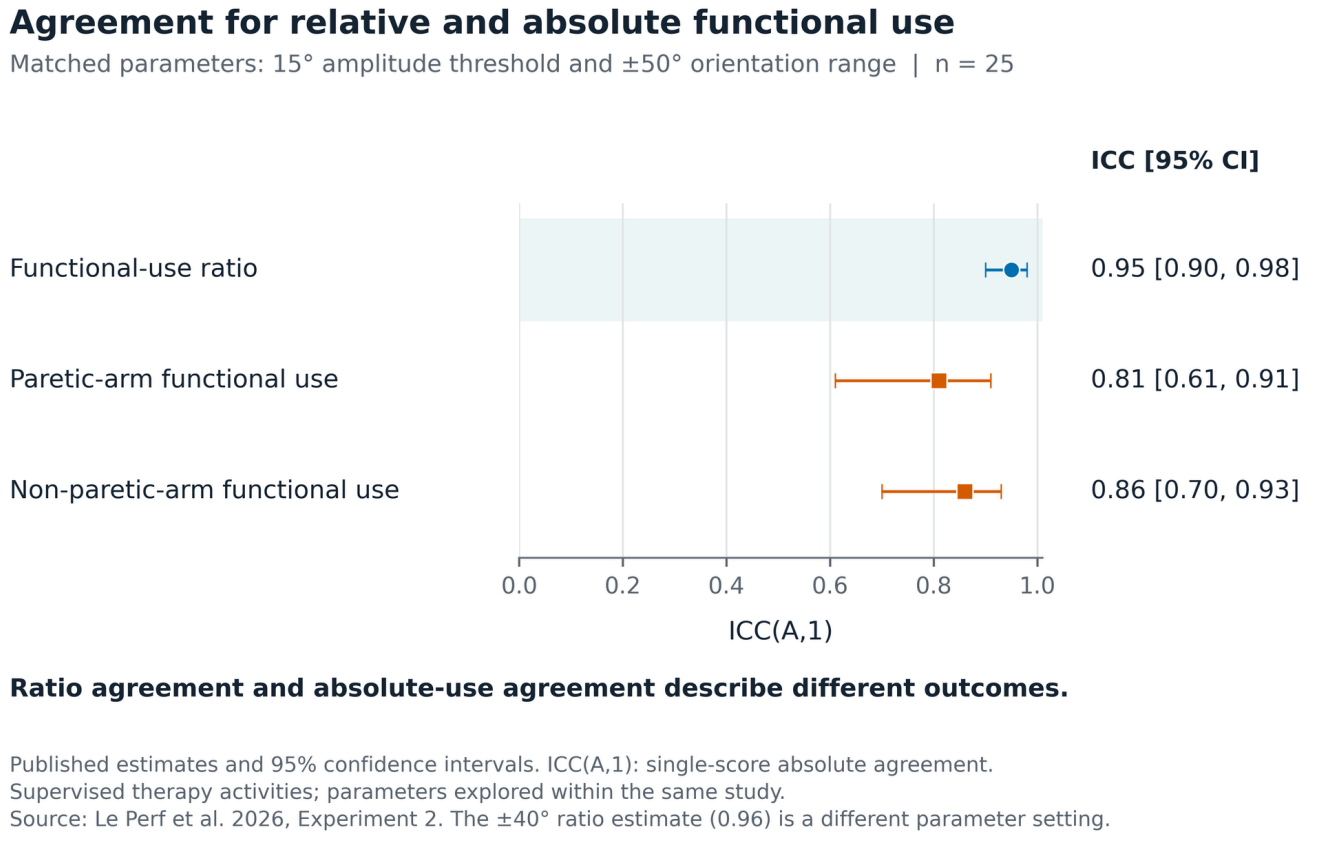


Figure 4. Agreement for relative and absolute functional use. Agreement with video-based assessment in 25 participants after stroke, using a 15-degree amplitude threshold and a plus-or-minus 50-degree orientation range. Points and whiskers show published single-score, two-way mixed-effects, absolute-agreement ICC(A,1) estimates and 95% confidence intervals. The ratio is a normalised difference between paretic and non-paretic functional-use counts; the absolute measures count functional two-second epochs. The matched ratio estimate is 0.95. The separately reported 0.96 ratio estimate uses a plus-or-minus 40-degree orientation range and is not substituted here. Higher ratio agreement coexisted with approximately 25% underestimation of absolute functional use, reported as an approximate bias in the source. Assessments involved supervised occupational-therapy activities; stabilisation was excluded from functional labels, and parameter selection used the same study data. Replotted from Le Perf et al. (2026), Experiment 2 [19]. The intervals do not test differences between these dependent estimates.

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

A 2026 finger versus wrist study analysed 20 participants with chronic stroke and mild to moderate impairment. Its published abstract reports stronger measurement properties for fine hand movement measures and improved properties with longer monitoring, identifying four days as sufficient within that study [20]. This supports investigating distal sensors when dexterity is the target. It does not establish a universal four day protocol across devices, recovery stages or severe impairment, nor does it remove the need to evaluate comfort and adherence.

Wang and colleagues developed the digital arm performance scale from approximately 23,000 hours of data in 215 participants. The published abstract describes movement segmentation, aggregated features and a linear mixed-effects model, with favourable reliability, validity and responsiveness. Its estimate of more than 60% fewer participants for trials came from power analysis [21]. That estimate is not an observed recruitment saving in independent prospective trials. It depends on the assumed treatment effect, variability and analysis. This abstract-level evidence does not establish the details of partitioning, calibration or uncertainty estimation, so the present appraisal makes no claims about those procedures.

Contextual sampling offers a complementary approach. Senadheera and colleagues combined bilateral sensing with repeated experience sampling over seven days in 30 stroke survivors and 30 comparison participants. Activity and environmental context helped interpret differences in arm use [22]. These observational associations do not establish that changing the context will improve recovery, but they show why a sensor score alone cannot explain behaviour. Table 2 summarises selected primary studies and their boundaries without ranking heterogeneous methods by a single accuracy statistic.

Table 2. Selected primary evidence and limits of interpretation

Study	Design and population	Contribution and principal boundary
Uswatte 2006 [9]	Validation in 169 participants, 3–9 months after stroke	Supports bilateral activity monitoring; selected mild to moderate impairment and indirect functional comparators
Bailey 2015 [11]	74 nondisabled adults and 48 with chronic stroke	Describes bilateral daily patterns; brief observation and cross-sectional comparisons
Waddell 2017 [14]	Longitudinal secondary analysis of 78 trial participants	Demonstrates capacity–performance discordance in the studied intervention; not proof that all transfer fails
Lum 2020 [17]	Classifiers tested against video in simulated community tasks	Functional labels improve specificity; personalised and across-person performance differ
Nieto 2025 [18]	Four participants completing laboratory and home scripts	Tests setting transfer; small case series with within-person models and scripted activities
Le Perf 2026 [19]	23 healthy participants and 25 participants after stroke	Orientation and ratio validation; supervised tasks, selected definitions and parameter exploration
Dhamrongsirivadh 2026 [20]	20 participants with chronic stroke in final analysis	Supports distal sensing and multiday evaluation; published abstract only in this review
Wang 2026 [21]	Composite development and evaluation using 215 participants	Promising digital arm performance scale; power-analysis efficiency, published abstract only here
Senadheera 2025 [22]	Seven-day monitoring in 30 stroke survivors and 30 comparators	Connects movement with sampled context; observational associations do not establish causation

This purposive evidence map is not an exhaustive review, ranking of devices or formal risk of bias assessment. Sample sizes refer to people, not sensor epochs.

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

6. VALIDATION MUST FOLLOW THE INTENDED INFERENCE

6.1 Analytical evaluation and model transport

For a classifier, frame or epoch accuracy is insufficient by itself. A model that correctly labels most inactive epochs may still estimate active time poorly. Evaluation should report class-specific sensitivity and precision, a confusion matrix, performance for ambiguous segments and error in the final person-level endpoint. Disagreement should be examined by activity and impairment, not only averaged across all observations. Where output probabilities are used, their calibration also requires assessment. These are proposed analytical requirements whose thresholds should be justified by the intended decision.

The effect of classification error depends on activity prevalence. For a binary activity label, expected classified active fraction is

$$\mathrm{E}\left(\widehat{p}\right) = \mathrm{Se}\, p + \left(1 - \mathrm{Sp}\right)\left(1 - p\right),$$

where p is the true active fraction, Se is sensitivity and Sp is specificity under the relevant operating conditions. In a hypothetical period with p = 0.05 and both sensitivity and specificity equal to 0.90, the expected classified active fraction is 0.14. This example is a mathematical illustration, not an estimate for any published device. It shows why apparently favourable classification percentages can substantially overestimate infrequent purposeful activity. Validation should therefore preserve realistic activity proportions and assess errors in the derived time or proportion endpoint. The equation assumes stable conditional error rates; changes in task mix or impairment can make those rates vary.

Participant and session independence must be established before feature selection, normalisation and model tuning. Splitting adjacent windows from one recording between training and evaluation risks exploiting person-specific or temporally dependent patterns. A separate test site is valuable when the intended claim concerns transport to another service; holding out windows is not an equivalent test. The same principle applies to composite scores trained against clinical scales. Correlation with the scale used during development may be partly expected and should be complemented by independent, patient relevant observations.

TRIPOD+AI provides reporting guidance for clinical prediction models, while PROBAST+AI addresses quality, bias and applicability of prediction models [23, 24]. Their directly applicable elements should be used for eligible prediction components; using machine learning alone does not make every activity summary a clinical prediction model. Neither checklist substitutes for validation of sensor acquisition, activity labels or the final outcome construct. Authors should state the component assessed and avoid presenting reporting completeness as evidence of clinical utility.

6.2 Verification and clinical validation

The V3 framework separates verification of the technology, analytical validation of the derived measure and clinical validation for the intended clinical concept [25]. V3+ adds usability validation [26]. This addition is especially relevant after stroke because successful monitoring can depend on one handed donning, comprehension, assistance and device maintenance. A technically accurate system that systematically fails in the intended users can produce biased clinical evidence.

Clinical validation should begin with patient relevant goals. The framework proposed by Manta and colleagues links meaningful aspects of health to a concept of interest, an outcome and a research endpoint [27]. For example, independent participation in meal preparation may motivate measuring affected arm contribution during ordinary food related tasks. Total wrist activity is a possible input, but it cannot inherit that meaning without evidence connecting the signal to the stated concept. Interviews and observations can identify activities missed by an apparently comprehensive composite.

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

A digital measure becomes an endpoint when its variable, timing, aggregation and role in a study are specified. Whether it is used as a biomarker or clinical outcome assessment depends on its context and intended interpretation; neither label establishes fitness for a particular trial [28]. FDA guidance on remote digital health technology use further emphasises suitability, validation and participant usability within the investigation [29]. Regulatory qualification, where relevant, is a separate evidence based process. Publication of an algorithm or use of a compliant device does not by itself qualify every derived endpoint.

7. SEPARATING MEASUREMENT ERROR FROM MEANINGFUL CHANGE

Reliability, agreement and responsiveness answer different questions. An intraclass correlation coefficient reflects the ability to distinguish people relative to observed variability; it can be high in a heterogeneous sample despite substantial absolute error. The ICC formulation, confidence interval, repeated measurement interval and clinical stability assumptions should be reported [30]. Device repeatability on a bench also differs from repeatability of a daily performance summary, which includes real variation in routines.

Method comparison should assess the size and distribution of disagreement, including systematic bias and whether error changes with the outcome level [31]. Repeated windows from a person must not be treated as independent pairs. A classifier may agree at the daily total level while mislabeling particular activities, or classify individual epochs well while introducing bias in a person's aggregate. The assessment should match the intended use: descriptive profiling, individual monitoring and group treatment comparisons can require different error tolerances.

Under a suitable stable-population model with approximately normal, independent errors and equal error variance at two assessments and no unaccounted systematic shift, an illustrative minimal detectable change is $\mathrm{MDC}_{95} = 1.96\sqrt{2}\,\mathrm{SEM}$. SEM is the standard error of measurement in the endpoint's original units [30, 32]. The formula requires revision when error depends strongly on baseline level, measurement schedules differ or errors are correlated. A clinically changing early recovery cohort is not automatically an appropriate sample for estimating stable test-retest error.

Detectable change does not define important change. Anchor based interpretation requires an external criterion that represents a relevant improvement and has adequate reliability and association with the target construct [32]. Early stroke research has estimated clinically important differences for several upper limb outcomes, including accelerometry, but these estimates remain tied to the instrument, anchor and recovery context [33]. They should not be transferred to a new composite score by numerical analogy.

For a wearable endpoint, the relevant anchor might concern improved task independence, valued activity participation or a patient rated change in a carefully specified domain. A global rating of overall recovery may be too broad if the sensor captures only arm movement. Multiple anchors and uncertainty intervals are preferable to presenting a single universal threshold without qualification. Group mean treatment effects and individual responder classifications also require separate interpretation: a statistically precise group difference need not exceed an individual MDC, and an individual's detectable change need not be beneficial.

8. OBSERVATION WINDOWS AND MISSING DATA

Monitoring duration should be estimated for the intended population and endpoint. Day-to-day variation may reflect therapy, weekday obligations, fatigue, social opportunity or travel. A longer interval may stabilise a mean while concealing a clinically relevant short event. Repeated baseline periods can quantify how many valid days support adequate precision, but the choice should preserve the outcome's clinical meaning. There is no single sufficient wear duration for all upper limb measures [20, 29].

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

Validity rules should distinguish actual inactivity from nonwear. Severe paresis, sleep, quiet stabilisation and device removal can all produce low acceleration. A nonwear rule developed in active adults may therefore misclassify the patients for whom monitoring is most needed. Wear diaries, contact sensors or complementary data can help, but each introduces burden and potential error. The rule and any manual adjudication should be finalised before treatment groups are compared.

Wear-adjusted metrics address one problem while potentially introducing another. Minutes of activity per valid hour are not equivalent to total daily activity. A participant who wears the device mainly during therapy may have a high rate and an unrepresentative daily sample. Simultaneous bilateral wear is essential for paired summaries. Reports should provide planned observation, captured time, analyzable time, valid days and exclusions by participant and group, preserving the reasons for loss rather than reporting a single average adherence percentage.

Missingness may be related to disability, treatment burden, perceived benefit or device usability. Complete-case analysis can select a population different from the one randomised. Imputation should respect the endpoint distribution, the timing of missing data and variables associated with both missingness and outcome; it cannot reconstruct unobserved functional behaviour without assumptions. Sensitivity analyses should examine plausible departures from the primary assumptions. Device failure, nonwear and an intercurrent clinical event require separate descriptions because they affect interpretation differently [29, 34].

9. DESIGNING A TRIAL ENDPOINT AND TESTING CLINICAL USEFULNESS

The estimand framework identifies the treatment conditions, population, variable, intercurrent event strategy and population-level summary that define the treatment effect of interest [34]. This structure is useful for rehabilitation endpoints even when a study is not intended for regulatory submission. Hospitalization, recurrent stroke, additional therapy and treatment discontinuation can change what a wearable outcome represents. A treatment policy question about everyday performance after allocation differs from a hypothetical question about performance without an intervening hospitalization. An event that makes the outcome undefined cannot be resolved merely by assigning zero activity.

A credible trial should prespecify its primary wearable endpoint, analysis window, aggregation, baseline adjustment and handling of these events. The algorithm should be locked before evaluation. Sample size depends on independent participants, effect size, outcome variance, attrition and the planned analysis; millions of epochs cannot compensate for a narrow participant sample. Multiple sensors and aggregation rules create opportunities for selective reporting, so exploratory comparisons should remain distinguishable from confirmatory tests.

Treatment plausibility must be separated from efficacy. A dose-response trial of task-specific training in chronic stroke did not establish a monotonic benefit of greater prescribed movement practice on capacity [35]. The ICARE trial did not show superiority of its structured task-oriented programme over the comparison care on the primary outcome [36]. These studies do not settle the value of all rehabilitation approaches. They show why improved measurement should support well-designed comparisons rather than convert a plausible mechanism or a rising sensor score into a treatment claim.

If the device provides feedback, monitoring becomes part of the intervention. A trial should distinguish passive observation from reminders, goals or therapist review, because these can alter behaviour. The target should reward useful activity rather than movements performed only to raise a score. This is a design concern for prospective testing, not an assertion that a particular system causes harmful behaviour. Persistence of benefit after feedback withdrawal may also be relevant if sustained daily participation is the goal.

The proposed endpoint specification in Table 3 links these considerations to explicit decisions. A feasible development sequence would begin with concept elicitation and usability, continue through representative reference observations and repeated monitoring, and then evaluate a locked endpoint at independent sites. Clinical usefulness requires a further compar-

ison: does access to the wearable result improve decisions, patient outcomes or resource use compared with the assessment otherwise available? Better agreement or responsiveness alone does not answer that question.

Table 3. Proposed minimum specification for a wearable trial endpoint

Domain	Required decision	Evidence or reporting needed
Concept	Which patient relevant aspect of functioning is targeted	Patient input, explicit construct and activities included or excluded
Population	Recovery stage, impairment range and setting	Representation, exclusions and subgroup uncertainty
Acquisition	Placement, sampling, calibration and simultaneous wear	Device and firmware version, verification and failure criteria
Processing	Filters, epochs, classifier and aggregation	Reproducible equations or code, locked parameters and denominator rules
Observation	Planned days, daily window and valid wear	Representativeness, captured time and reasons for missingness
Interpretation	Intended direction and meaning of change	Absolute error, relevant anchors and uncertainty
Treatment effect	Population, treatment contrast, variable, events and summary	Prespecified estimand, analysis and sensitivity analyses
Usability	Tasks required from patients and other users	Intended-user testing, assistance, burden and completion failures
Transport	Claim for unfamiliar people and settings	Independent evaluation and version-specific performance

Proposed application of digital measurement and trial principles [25, 26, 27, 28, 29, 34], not a qualification certificate or validated checklist.

10. INTEGRATING RESULTS INTO REHABILITATION REASONING

Wearable and clinical findings should be interpreted as a pattern. Increased capacity with unchanged daily movement can prompt examination of opportunity, habits, fatigue or measurement coverage. Increased daily activity with stable capacity may indicate better use of existing ability, changed assistance or a different routine. Improvement in both still does not establish restitution without movement-level evidence. Table 4 offers a proposed interpretation matrix with alternative explanations and the next observation needed; it is not a diagnostic algorithm.

Reports should state what was observed in concrete terms. For example, affected arm active time increased during valid daytime wear, whereas task capacity and reported assistance needs were unchanged. This statement preserves the distinction between measurement and interpretation. It also allows the patient to judge whether the observed behaviour served a valued activity. A simple rising recovery score would obscure those relationships.

Implementation should evaluate people with aphasia, cognitive or visual difficulties, severe paresis, sensory impairment and limited support, or clearly restrict the intended population when evidence is unavailable. Accommodation might include assisted setup, accessible instructions or offline collection. Investigators should report who could not participate and why. V3+ usability validation requires attention to the tasks performed by patients, caregivers and clinical staff, rather than assuming that successful wear in a supervised study will transfer to routine care [26].

Data governance is part of reproducibility. Studies should specify consent for contextual video, access to raw signals, secondary algorithm development, retention and version tracking. A firmware or processing change can alter an endpoint without biological change. Where raw video cannot be shared, annotation definitions, processing code, synthetic examples and justified access arrangements can still improve scrutiny. These are proposed reporting practices; privacy constraints should be described rather than used to imply that an opaque processing pipeline is reproducible.

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

Table 4. Proposed interpretation of discordant changes

Observed pattern	Plausible interpretations	Additional observation needed
Capacity improves, movement unchanged	Limited transfer, unchanged opportunity or insensitive movement summary	Activity context, patient goals and coverage of the target behaviour
Movement increases, capacity unchanged	Greater use of existing ability, altered routine, assistance or compensation	Functional labels, assistance and task success
Use ratio improves, affected activity unchanged	Less affected arm activity declined	Both absolute limb values, total activity and valid wear
Task success improves, joint control unchanged	Effective compensation or environmental adaptation	Kinematics and patient valued independence
Composite improves, valued participation unchanged	Detectable change outside the prioritised activity or anchor mismatch	Construct alignment, meaningful-change evidence and patient interview
Daily score falls during reduced wear	Measurement selection, real activity change or both	Wear timing, missingness reasons and sensitivity analysis

These are competing hypotheses to examine, not diagnoses or observed treatment effects. Clinical assessment, bilateral monitoring and contextual evidence should be considered together.

11. LIMITATIONS AND RESEARCH PRIORITIES

This review is selective and does not estimate the total literature or provide pooled comparative performance. Devices, labels, populations and observation schedules are heterogeneous. Several foundational studies have limited participant spectra, while some recent reports remain small or require fuller published-method appraisal. The newest DAPS and finger-worn findings were evaluated at abstract level, and the source register makes that limitation explicit. The proposed tables integrate measurement reasoning and existing frameworks but have not been validated as a consensus checklist.

Priority studies should test locked algorithms in independent participants and sites, include unstructured daily behaviour, quantify endpoint error across severity and link longitudinal changes to patient relevant anchors. Labels should distinguish purposeful manipulation, stabilisation and other valuable activity without assuming that a shared name means a shared construct. Research should also examine whether increased sensor detail adds actionable information beyond established assessment, and whether that information improves decisions when used in practice.

12. CONCLUSION

Upper limb wearables can strengthen stroke rehabilitation research by describing activity beyond the clinic. Their credibility depends on preserving the distinctions between acceleration, functional activity, capacity, participation and restitution. A useful digital endpoint specifies its construct, measurement pipeline, observation window, uncertainty and intended treatment effect before interpretation. Combining validated daily performance measures with clinical and movement-level assessments offers a defensible route to more informative trials and more precise rehabilitation reasoning.

REFERENCES

1. Lang CE, Bland MD, Bailey RR, Schaefer SY, Birkenmeier RL. Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. J Hand Ther. 2013;26(2):104–114; quiz 115. https://doi.org/10.1016/j.jht.2012.06.005
2. 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. https://doi.org/10.2340/1650197771331
3. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int J Rehabil Res. 1981;4(4):483–492. https://doi.org/10.1097/00004356-198112000-00001

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

4. Levin MF, Kleim JA, Wolf SL. What do motor “recovery” and “compensation” mean in patients following stroke? Neurorehabil Neural Repair. 2009;23(4):313–319. https://doi.org/10.1177/1545968308328727
5. Bernhardt J, Hayward KS, Kwakkel G, Ward NS, Wolf SL, Borschmann K, et al. Agreed definitions and a shared vision for new standards in stroke recovery research: the Stroke Recovery and Rehabilitation Roundtable taskforce. Neurorehabil Neural Repair. 2017;31(9):793–799. https://doi.org/10.1177/1545968317732668
6. Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. Neurorehabil Neural Repair. 2017;31(9):784–792. https://doi.org/10.1177/1545968317732662
7. van Hees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One. 2013;8(4):e61691. https://doi.org/10.1371/journal.pone.0061691
8. Uswatte G, Foo WL, Olmstead H, Lopez K, Holand A, Simms LB. Ambulatory monitoring of arm movement using accelerometry: an objective measure of upper-extremity rehabilitation in persons with chronic stroke. Arch Phys Med Rehabil. 2005;86(7):1498–1501. https://doi.org/10.1016/j.apmr.2005.01.010
9. Uswatte G, Giuliani C, Winstein C, Zeringue A, Hobbs L, Wolf SL. Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial. Arch Phys Med Rehabil. 2006;87(10):1340–1345. https://doi.org/10.1016/j.apmr.2006.06.006
10. Bailey RR, Lang CE. Upper-limb activity in adults: referent values using accelerometry. J Rehabil Res Dev. 2013;50(9):1213–1222. https://doi.org/10.1682/JRRD.2012.12.0222
11. Bailey RR, Klaesner JW, Lang CE. Quantifying real-world upper-limb activity in nondisabled adults and adults with chronic stroke. Neurorehabil Neural Repair. 2015;29(10):969–978. https://doi.org/10.1177/1545968315583720
12. Lang CE, Wagner JM, Edwards DF, Dromerick AW. Upper extremity use in people with hemiparesis in the first few weeks after stroke. J Neurol Phys Ther. 2007;31(2):56–63. https://doi.org/10.1097/NPT.0b013e31806748bd
13. Thrane G, Emaus N, Askim T, Anke A. Arm use in patients with subacute stroke monitored by accelerometry: association with motor impairment and influence on self-dependence. J Rehabil Med. 2011;43(4):299–304. https://doi.org/10.2340/16501977-0676
14. Waddell KJ, Strube MJ, Bailey RR, Klaesner JW, Birkenmeier RL, Dromerick AW, et al. Does task-specific training improve upper limb performance in daily life poststroke? Neurorehabil Neural Repair. 2017;31(3):290–300. https://doi.org/10.1177/1545968316680493
15. Lang CE, Waddell KJ, Klaesner JW, Bland MD. A method for quantifying upper limb performance in daily life using accelerometers. J Vis Exp. 2017;(122):55673. https://doi.org/10.3791/55673
16. Leuenberger K, Gonzenbach R, Wachter S, Luft A, Gassert R. A method to qualitatively assess arm use in stroke survivors in the home environment. Med Biol Eng Comput. 2017;55(1):141–150. https://doi.org/10.1007/s11517-016-1496-7
17. Lum PS, Shu L, Bochniewicz EM, Tran T, Chang LC, Barth J, et al. Improving accelerometry-based measurement of functional use of the upper extremity after stroke: machine learning versus counts threshold method. Neurorehabil Neural Repair. 2020;34(12):1078–1087. https://doi.org/10.1177/1545968320962483
18. Nieto EM, Lujan E, Mendoza CA, Arriaga Y, Fierro C, Tran T, et al. Accelerometry and the capacity–performance gap: case series report in upper-extremity motor impairment assessment post-stroke. Bioengineering (Basel). 2025;12(6):615. https://doi.org/10.3390/bioengineering12060615
19. Le Perf G, Pratola B, Mottet D, Laffont I, Bakhti K. Validity of FuncUseRatio for ecological monitoring of functional arm use after stroke: toward precision neurorehabilitation. J Neuroeng Rehabil. 2026;23:230. https://doi.org/10.1186/s12984-026-01978-9
20. Dhamrongsirivadh R, Pugliese BL, Civeriati V, Piela K, Fabara E, Vergara-Diaz G, et al. More than just arm movement: finger-worn accelerometers provide a valid and sensitive alternative to wrist-worn accelerometers for measuring real-world upper-limb performance in stroke survivors. Arch Phys Med Rehabil. 2026 Aug 31. Online ahead of print. https://doi.org/10.1016/j.apmr.2026.08.015
21. Wang R, Lang CE, Stoykov ME, Bonato P, Lee SI. Assessment of upper-limb motor recovery after stroke using a wrist-worn accelerometer digital biomarker. Sci Transl Med. 2026;18(869):eadw3644. https://doi.org/10.1126/scitranslmed.adw3644
22. Senadheera I, Hettiarachchi P, Haslam B, Nawaratne R, Pollack M, Hillier S, et al. Combined individual experience and accelerometry measurement of upper limb use in daily activities in real time after stroke. Sensors (Basel). 2025;25(23):7330. https://doi.org/10.3390/s25237330
23. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

24. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505. https://doi.org/10.1136/bmj-2024-082505

25. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). NPJ Digit Med. 2020;3:55. https://doi.org/10.1038/s41746-020-0260-4

26. Bakker JP, Barge R, Centra J, Cobb B, Cota C, Guo CC, et al. V3+ extends the V3 framework to ensure user-centricity and scalability of sensor-based digital health technologies. NPJ Digit Med. 2025;8(1):51. https://doi.org/10.1038/s41746-024-01322-2

27. Manta C, Patrick-Lake B, Goldsack JC. Digital measures that matter to patients: a framework to guide the selection and development of digital measures of health. Digit Biomark. 2020;4(3):69–77. https://doi.org/10.1159/000509725

28. Goldsack JC, Dowling AV, Samuelson D, Patrick-Lake B, Clay I. Evaluation, acceptance, and qualification of digital measures: from proof of concept to endpoint. Digit Biomark. 2021;5(1):53–64. https://doi.org/10.1159/000514730

29. US Food and Drug Administration. Digital health technologies for remote data acquisition in clinical investigations. Guidance for industry, investigators, and other stakeholders. December 2023. Accessed 10 November 2026. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations

30. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–240. https://doi.org/10.1519/15184.1

31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;327(8476):307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

32. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. 2006;4:54. https://doi.org/10.1186/1477-7525-4-54

33. Lang CE, Edwards DF, Birkenmeier RL, Dromerick AW. Estimating minimal clinically important differences of upper-extremity measures early after stroke. Arch Phys Med Rehabil. 2008;89(9):1693–1700. https://doi.org/10.1016/j.apmr.2008.02.022

34. International Council for Harmonisation. E9(R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Final version. 20 November 2019. Accessed 10 November 2026. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf

35. Lang CE, Strube MJ, Bland MD, Waddell KJ, Cherry-Allen KM, Nudo RJ, et al. Dose response of task-specific upper limb training in people at least 6 months poststroke: a phase II, single-blind, randomized, controlled trial. Ann Neurol. 2016;80(3):342–354. https://doi.org/10.1002/ana.24734

36. Winstein CJ, Wolf SL, Dromerick AW, Lane CJ, Nelsen MA, Lewthwaite R, et al. Effect of a task-oriented rehabilitation program on upper extremity recovery following motor stroke: the ICARE randomized clinical trial. JAMA. 2016;315(6):571–581. https://doi.org/10.1001/jama.2016.0276

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

MULTILINGUAL ABSTRACTS · RESUMEN · الملخص · सार

Español

Evaluación con dispositivos portátiles de la recuperación del miembro superior tras un ictus: de las señales de movimiento a criterios de valoración significativos

Antecedentes: Los sensores portátiles caracterizan el movimiento del miembro superior más allá de la consulta, pero la cantidad de movimiento, la actividad funcional, la restitución motora y la participación son resultados distintos.

Objetivo: Evaluar la cadena de medición que conecta las señales de los dispositivos portátiles con resultados significativos tras el ictus y proponer un marco para la especificación y la validación de criterios de valoración.

Métodos: La recuperación dirigida de la literatura y la verificación de citas se completaron el 10 de noviembre de 2026. Se seleccionaron de forma intencional estudios fundacionales de medición, investigación primaria con acelerometría, estudios recientes de resultados digitales y guías metodológicas. La evidencia se organizó por constructo, referencia, entorno, sensibilidad al cambio y uso previsto. No se produjeron estimaciones agrupadas ni graduaciones formales de la certeza.

Síntesis: La acelerometría bilateral complementa las pruebas clínicas de capacidad, pero por sí sola no puede identificar un agarre exitoso, la contribución voluntaria ni la restauración de la coordinación articular. Los cocientes de lateralidad pueden mejorar cuando disminuye la actividad del brazo menos afectado. Los clasificadores, las métricas de orientación, los sensores de dedo y las medidas compuestas ofrecen enfoques complementarios, con límites en las etiquetas de actividad, la cobertura poblacional y la validación externa. Una escala digital reciente de desempeño del brazo respalda su desarrollo posterior; las estimaciones de eficiencia de ensayos basadas en análisis de potencia no establecen un beneficio para el paciente ni la cualificación del criterio de valoración. El error de medición, el cambio significativo, los datos faltantes y la usabilidad requieren una evaluación independiente.

Conclusión: Los criterios de valoración creíbles requieren un constructo definido y relevante para el paciente, un procesamiento reproducible, la validación en el entorno previsto y una interpretación preespecificada del cambio. Integrar las medidas portátiles de desempeño con evaluaciones de la deficiencia, la capacidad y la estrategia de movimiento favorece una investigación y una toma de decisiones más precisas en rehabilitación.

Palabras clave: rehabilitación del ictus; miembro superior; acelerometría; criterios de valoración digitales; recuperación motora; validez de la medición

JMMBS
JOURNAL OF MOVEMENT MECHANICS & BIOMECHANICS SCIENCE
Research, reviews, and methodological work in biomechanics, movement mechanics, rehabilitation, and performance science

العربية

التقييم بالأجهزة القابلة للارتداء لتعافي الطرف العلوي بعد السكتة الدماغية: من إشارات الحركة إلى نقاط نهاية ذات معنى

الخلفية: تصف المستشعرات القابلة للارتداء حركة الطرف العلوي خارج العيادة، غير أن كمية الحركة والنشاط الوظيفي والاستعادة الحركية والمشاركة نواتجُ متمايزة.

الهدف: تقييم سلسلة القياس التي تربط إشارات الأجهزة القابلة للارتداء بالنواتج ذات المعنى بعد السكتة الدماغية، واقتراح إطار لتحديد نقاط النهاية والتحقق منها.

المنهجية: اكتمل الاسترجاع الموجّه للأدبيات والتحقق من الاستشهادات في 10 نوفمبر 2026. واختيرت بصورة هادفة دراساتُ قياس تأسيسية، وأبحاث أولية في قياس التسارع، ودراسات حديثة للنواتج الرقمية، وإرشادات منهجية. ونُظِّمت الأدلة حسب البنية والمرجع والبيئة والاستجابة للتغير والاستخدام المقصود. ولم تُحسب تقديرات مُجمَّعة ولا تصنيفات رسمية لدرجة اليقين.

التركيب: يُكمّل قياس التسارع الثنائي اختباراتِ القدرة السريرية، لكنه لا يستطيع وحده تحديد الإمساك الناجح أو المساهمة الإرادية أو استعادة التناسق المفصلي. وقد تتحسن نسب الجانبية عندما ينخفض نشاط الذراع الأقل تأثرًا. وتقدّم المصنِّفات ومقاييس الاتجاه ومستشعرات الأصابع والمقاييس المركّبة مقاربات متكاملة، مع حدود تتعلق بتسميات النشاط وتغطية الفئات السكانية والتحقق الخارجي. ويدعم مقياسٌ رقمي حديث لأداء الذراع مزيدًا من التطوير؛ أما تقديرات كفاءة التجارب المستندة إلى تحليل القوة الإحصائية فلا تُثبت فائدةً للمريض ولا تأهيلَ نقطة النهاية. ويتطلب كلٌّ من خطأ القياس والتغير ذي المعنى والبيانات المفقودة وقابلية الاستخدام تقييمًا مستقلًا.

الخلاصة: تتطلب نقاط النهاية الموثوقة بنيةً محدّدة ذات صلة بالمريض، ومعالجةً قابلة للاستنساخ، وتحققًا في البيئة المقصودة، وتفسيرًا مُسبق التحديد للتغير. ويدعم دمجُ مقاييس الأداء القابلة للارتداء مع تقييمات الاعتلال والقدرة واستراتيجية الحركة بحثًا وقراراتٍ تأهيلية أكثر دقة.

الكلمات المفتاحية: تأهيل السكتة الدماغية؛ الطرف العلوي؛ قياس التسارع؛ نقاط النهاية الرقمية؛ التعافي الحركي؛ صلاحية القياس

हिन्दी

स्ट्रोक के बाद ऊपरी अंग की रिकवरी का वियरेबल-आधारित मूल्यांकन: गति-संकेतों से सार्थक एंडपॉइंट तक

पृष्ठभूमि: वियरेबल सेंसर क्लिनिक के बाहर ऊपरी अंग की गति का वर्णन करते हैं, परंतु गति की मात्रा, कार्यात्मक गतिविधि, मोटर पुनर्स्थापन और सहभागिता अलग-अलग परिणाम हैं।

उद्देश्य: स्ट्रोक के बाद वियरेबल संकेतों को सार्थक परिणामों से जोड़ने वाली माप-शृंखला का मूल्यांकन करना और एंडपॉइंट के निर्धारण व सत्यापन के लिए एक ढाँचा प्रस्तावित करना।

विधियाँ: लक्षित साहित्य-प्राप्ति और उद्धरण-सत्यापन 10 नवंबर 2026 को पूर्ण किए गए। आधारभूत माप-अध्ययन, प्राथमिक एक्सेलेरोमेट्री शोध, हालिया डिजिटल परिणाम-अध्ययन और पद्धतिगत दिशानिर्देश उद्देश्यपूर्ण ढंग से चुने गए। साक्ष्य को संरचना, संदर्भ, परिवेश, परिवर्तन-संवेदनशीलता और अभिप्रेत उपयोग के अनुसार व्यवस्थित किया गया। कोई संयुक्त अनुमान या औपचारिक निश्चितता-ग्रेड तैयार नहीं किए गए।

संश्लेषण: द्विपक्षीय एक्सेलेरोमेट्री नैदानिक क्षमता-परीक्षणों की पूरक है, परंतु अकेले सफल पकड़, स्वैच्छिक योगदान या संधि-समन्वय की पुनर्स्थापना की पहचान नहीं कर सकती। कम प्रभावित भुजा की गतिविधि घटने पर पार्श्विकता-अनुपात बेहतर दिख सकते हैं। क्लासिफ़ायर, अभिविन्यास-मेट्रिक्स, उँगली-सेंसर और संयुक्त माप पूरक दृष्टिकोण प्रदान करते हैं, जिनकी सीमाएँ गतिविधि-लेबल, जनसंख्या-कवरेज और बाहरी सत्यापन से जुड़ी हैं। एक हालिया डिजिटल आर्म परफ़ॉर्मेंस स्केल आगे के विकास का समर्थन करता है; पावर-विश्लेषण पर आधारित परीक्षण-दक्षता अनुमान रोगी-लाभ या एंडपॉइंट-योग्यता स्थापित नहीं करते। माप-त्रुटि, सार्थक परिवर्तन, लुप्त डेटा और उपयोगिता — प्रत्येक का अलग मूल्यांकन आवश्यक है।

निष्कर्ष: विश्वसनीय एंडपॉइंट के लिए रोगी-प्रासंगिक परिभाषित संरचना, पुनरुत्पादनीय प्रसंस्करण, अभिप्रेत परिवेश में सत्यापन और परिवर्तन की पूर्व-निर्दिष्ट व्याख्या आवश्यक है। वियरेबल प्रदर्शन-मापों को क्षति, क्षमता और गति-रणनीति के आकलनों के साथ जोड़ना अधिक सटीक पुनर्वास-शोध और निर्णय-निर्माण का समर्थन करता है।

मुख्य शब्द: स्ट्रोक पुनर्वास; ऊपरी अंग; एक्सेलेरोमेट्री; डिजिटल एंडपॉइंट; मोटर रिकवरी; माप वैधता