

UCLA

UCLA Previously Published Works

Title

Domain-Specific Outcomes for Stroke Clinical Trials: What the Modified Rankin Isn't Ranking.

Permalink

<https://escholarship.org/uc/item/0km716x5>

Journal

Neurology, 97(8)

ISSN

0028-3878

Authors

Braun, Robynne G
Heitsch, Laura
Cole, John W
[et al.](#)

Publication Date

2021-08-24

DOI

10.1212/wnl.00000000000012231

Peer reviewed

Domain-Specific Outcomes for Stroke Clinical Trials

What the Modified Rankin Isn't Ranking

Robynne G. Braun, MD, PhD, Laura Heitsch, MD, John W. Cole, MD, MS, Arne G. Lindgren, MD, PhD, Adam de Havenon, MD, Jason A. Dude, BS, Keith R. Lohse, PhD, Steven C. Cramer, MD, and Bradford B. Worrall, MD, MSc, on behalf of the GPAS Collaboration, Phenotyping Core

Correspondence

Dr. Braun
robynne.braun@ummm.edu

Neurology® 2021;97:367-377. doi:10.1212/WNL.0000000000012231

Abstract

Global outcome measures that are widely used in stroke clinical trials, such as the modified Rankin Scale (mRS), lack sufficient detail to detect changes within specific domains (e.g., sensory, motor, visual, linguistic, or cognitive function). Yet such data are vital for understanding stroke recovery and its mechanisms. Poststroke deficits in specific domains differ in their rate and degree of recovery and in their effects on overall independence and quality of life. For example, even in a patient with complete recovery of strength, persistent deficits in the nonmotor domains such as language and cognition may make a return to independent living impossible. In such cases, global measures based solely on the patient's degree of independence would overlook a complete recovery in the motor domain. Capturing these important aspects of recovery demands a domain-specific approach. If stroke outcomes trials are to incorporate finer-grained recovery metrics—which can require substantial time, effort, and expertise to implement—efficiency must be a priority. In this article, we discuss how commonly collected clinical data from the NIH Stroke Scale can guide the judicious selection of relevant recovery domains for more detailed testing. Our overarching goal is to make the implementation of domain-specific testing more feasible for large-scale clinical trials on stroke recovery.

From the Department of Neurology (R.G.B., J.W.C.), University of Maryland, Baltimore; Department of Emergency Medicine (L.H.), Washington University, St. Louis, MO; Department of Clinical Sciences, Lund (A.L.), Lund University; Department of Neurology (A.G.L.), Skåne University Hospital, Lund, Sweden; Departments of Neurology (A.d.H.) and Health and Kinesiology (J.A.D., K.R.L.), University of Utah, Salt Lake City; Department of Neurology (S.C.C.), University of California, Los Angeles; California Rehabilitation Institute (S.C.C.), Los Angeles; and Departments of Neurology and Public Health Sciences (B.B.W.), University of Virginia, Charlottesville.

Go to [Neurology.org/N](https://www.neurology.org/N) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Glossary

FM-UE = Fugl-Meyer upper extremity; **GPAS** = Genetics Platform for Acute Stroke; **ICC** = intraclass correlation coefficient; **ICF** = International Classification of Functioning; **ISGC** = International Stroke Genetics Consortium; **MoCA** = Montreal Cognitive Assessment Scale; **mRS** = modified Rankin Scale; **NIHSS** = NIH Stroke Scale; **NINDS** = National Institute of Neurological Disorders and Stroke; **SAFE** = Shoulder Abduction Finger Extension; **SRRR** = Stroke Recovery and Rehabilitation Roundtable; **tPA** = tissue plasminogen activator.

Clinical trials of acute stroke therapies have historically used global rating scales. The most commonly used end point measure in randomized control trials of acute stroke treatments is the modified Rankin Scale (mRS; table 1), a global measure of functional independence with scores ranging from 0 for no symptoms to 6 for death. Favored by clinical trialists for its ease of use, the mRS is endorsed as a primary end point by multiple regulatory agencies, including the US Food and Drug Administration¹ and National Institute of Neurological Disorders and Stroke (NINDS).²⁻⁴ Nearly every major acute stroke treatment trial conducted since 1995, including the NINDS tissue plasminogen activator (tPA) trial,⁵ IMSIII (Interventional Management of Stroke III),⁶ ALIAS (Albumin in Acute Ischemic Stroke),⁷ and MR CLEAN (Multi-center Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands),⁸ has included the 90-day mRS score. In practice, mRS scores are commonly reported in a binarized format as being “good” or “poor” (typically with an mRS of 0–1 or 0–2 defined as good outcome), further reducing the granularity of this scale from ordinal to dichotomous.

Another commonly used global measure is the NIH Stroke Scale (NIHSS), a nonlinear ordinal scale originally intended for measurement of stroke severity, but now also used to measure stroke outcome. Scores range from 0 to 42, where 0 indicates normal function, and higher scores indicate more severe deficits (table 2).^{9,10} The total NIHSS score is the sum of component scores spanning the sensory, motor, visual, linguistic, and cognitive domains. However, the individual NIHSS component scores generally are not reported because many clinical trials and administrative databases, including Get With the Guidelines Stroke,¹¹ only require reporting of the total NIHSS scores. Use of the NIHSS total score as a single value limits understanding of recovery by failing to delineate the relative contribution of sensory, motor, visual, linguistic, and cognitive domains individually, even though this is the level at which neurologic systems—and the neurorestorative/neurorehabilitation approaches targeting them—are organized.

Domain-specific outcome measures are important primary endpoints for clinical trials of stroke treatment and stroke recovery both in the acute and postacute phases of care. In 2007, Cramer et al.¹² illustrated this point in a deft summary of prior studies showing that poststroke deficits within particular domains recover at different rates and to differing

degrees, and can independently affect functional status and quality of life. Elsewhere, Felberg et al.¹³ reported on the recovery of poststroke deficits in the first 24 hours following tPA infusion and found that while all NIHSS component scores contributed to total NIHSS score decrease, some components such as aphasia, right-sided motor symptoms, and neglect responded less well than others. More recently, Mikulik et al.¹⁴ showed that not only the degree of recovery but also the rate of recovery was variable for different neurologic deficits. This fact underscores the importance of obtaining longitudinal measures that can capture recovery dynamics over time.

Domain-Specific Recovery Dynamics in the NINDS tPA Trial Data

To further illustrate these points regarding the domain-specific and time-dependent aspects of recovery, it is useful to plot NIHSS component scores in a format that shows the changes in relative proportions of patients with particular deficits over time. To this end, several alluvial plots of the NINDS tPA trial⁵ data are provided in Figure 1, depicting NIHSS component scores in the first 90 days poststroke for 487 patients with ischemic stroke. Patients deceased at 90 days or with missing data at any time point have been excluded. An interactive version of Figure 1 is available online with a video tutorial orienting readers to the interpretation of alluvial plots and to the figure's interactive features (<https://genestroke.wixsite.com/alliesinstroke/domainspecificmeasures>). When inspecting these plots, changes within the first 24 hours poststroke may be of particular interest for acute stroke clinicians since they reflect a primary effect of acute stroke interventions (e.g., tPA in this dataset). Changes between discharge and 90 days may particularly interest neurorehabilitation clinicians since they reflect continued recovery (or decline) within specific domains over time. From a patient perspective, having information about the relative persistence or resolution of different deficits may be helpful for understanding recovery prognosis. From the perspective of clinical trial design, evaluating these patterns could help to determine whether patients being enrolled into a trial are in fact likely to retain some treatable degree of impairment in the domain that the study treatment targets.

In figure 1, representative cases are rendered in black to show how recovery in specific domains may differentially contribute to the overall level of disability on the mRS. Take for example

Table 1 Modified Rankin Scale (mRS)

Score	Description
0	No symptoms.
1	No significant disability. Able to carry out all usual activities, despite some symptoms.
2	Slight disability. Able to look after own affairs without assistance, but unable to carry out all previous activities.
3	Moderate disability. Requires some help, but able to walk unassisted.
4	Moderately severe disability. Unable to attend to own bodily needs without assistance, and unable to walk unassisted.
5	Severe disability. Requires constant nursing care and attention, bedridden, incontinent.
6	Dead.

The original 1957 Rankin scale, used to assess patient outcomes in one of the first hospital stroke units, was scored from 1 to 5 (for no disability, slight, moderate, moderately severe, and severe disability), then later revised as the mRS to include a score of 0 for no symptoms and 6 for death.

2 patients with 90 days mRS ≥ 3 and 90 day NIHSS motor scores = 0 (a pattern that occurs in 45/487 cases). Both case 75 and case 211 show a similar pattern with recovery to mRS = 3 at 90 days (unable to live alone without help from another person). Both made a complete motor recovery, but neither patient fully recovered sensory or language function, and ultimately neither was able to return to independent living. These 2 cases both illustrate a classic neurorehabilitation dilemma, since patients with this deficit profile (i.e., complete motor recovery with persistent deficits in nonmotor domains) may not qualify for acute neurorehabilitation on the basis of their low physical therapy needs. It is possible that more intensive speech/language therapy might have allowed these patients to progress to an mRS = 2 (able to live alone with some outside assistance). This point is not likely to be recognized if only the total NIHSS is considered. To facilitate further data exploration of recovery dynamics within domains and variations in recovery patterns between domains, the reader is referred to the interactive version of figure 1 referenced above.

Domain-specific recovery dynamics cannot be discerned when using coarser global measures like the mRS or total NIHSS. This is an important consideration for studies designed to assess treatment efficacy, since the observed differences in the rate and degree of recovery across domains may reflect patient-specific differences in treatment responsiveness, lesion characteristics, neural repair processes, or the health of the premorbid neural substrate. Capturing these important aspects of recovery demands a domain-specific approach. The relative lack of domain-specific testing in stroke outcomes studies to date may be partly due to the perception that testing of multiple neurologic domains is impractical for large-scale, multicenter studies because it requires substantial time, effort, expertise, and expense.

However, we maintain that domain-specific measurements could be made more efficient and economical if that were a priority. To balance the need for finer-grained recovery metrics with the need for economical and efficient study operations, we have developed a protocolized approach that focuses on relevant domains, using commonly collected clinical data from the NIHSS. This approach was initially conceived as a means to harmonize data collection in multicenter studies examining the biology of stroke recovery, most notably for genetic studies of ischemic stroke. However, this framework could be readily applied to other types of stroke as well, including hemorrhagic stroke. The intent is to facilitate high-quality phenotypic data collection while minimizing burden on participants and research teams.

Screening for Domain-Specific Deficits

A schematic depiction of our overall approach is given in figure 2. Initial NIHSS scores are obtained at the time of acute hospital admission. These scores are then used to guide the selection of further domain-specific tests. Note that initial assessment scores (before any acute intervention) are used, since even in patients with post-treatment NIHSS = 0, subtle deficits may still be detected when more granular measures are applied. Depending on research priorities, additional sensory, motor, cognitive, linguistic, or visual testing can then be selected based on the impairments present at the time of hospital admission.

Domain-Specific Testing Options

A roster of domain-specific testing options is given in table 2. The testing options offered here have been selected with consideration given to feasibility, reproducibility, validity, inter- and intrarater reliability, and time/effort required to administer. They incorporate prior recommendations from the Stroke Recovery and Rehabilitation Roundtable (SRRR) consensus statement¹⁵ and the Research Outcome Measurement in Aphasia (ROMA) consensus statement.¹⁶ Regarding timing, we have also adhered to the SRRR recommendation for starting assessments within 7 days of stroke onset and continuing at set time points up to at least 3 months post-stroke. The list of testing options provided in table 2 is neither exhaustive nor proscriptive, but provides a solid starting point for researchers to select tests that are appropriate for assessing longer-term recovery. The rationale for our choice of testing instruments is fully detailed elsewhere (Standardized data collection in prospective genetic studies of ischemic stroke evolution and recovery; Global Alliance consensus statement as endorsed by the International Stroke Genetics Consortium [ISGC; <https://doi.org/10.1177/17474930211007288>]). Our rationale for test selection favors measures that are brief, simple, internationally familiar, and sensitive to change. Measures that have existing training materials to optimize inter- and intrarater reliability were selected wherever possible. Our pairing of specific

Table 2 NIH Stroke Scale (NIHSS) Items, Scoring, Attributes, and Domain-Specific Testing Options

NIHSS item	Domain/deficit tested	NIHSS κ^a	Scoring	Domain-specific testing options and durations
1a	LOC	0.50	0 = Alert, 2 = not alert, obtunded, 3 = unresponsive	MoCA ^c or Telephone MoCA (20 min)
1b	LOC questions	0.64	0 = Answers both correctly, 1 = answers one correctly, 2 = answers neither correctly	MoCA ^c or Telephone MoCA (20 min)
1c	LOC commands	0.41 ^b	0 = Performs both tasks correctly, 1 = performs one task correctly, 2 = performs neither task	NA
2	Gaze (horizontal extraocular movements)	0.33 ^b	0 = Normal, 1 = partial gaze palsy, 2 = total gaze palsy	NA
3	Visual fields	0.57	0 = No visual loss, 1 = partial hemianopsia, 2 = complete hemianopsia, 3 = bilateral hemianopsia	Portable perimetry app ^d
4	Facial palsy	0.22 ^b	0 = Normal, 1 = minor paralysis, 2 = partial paralysis, 3 = complete paralysis	NA
5a and 5b	Motor arm, left and right	0.77	0 = No drift, 1 = drift before 10 s, 2 = falls before 10 s, 3 = no effort against gravity, 4 = no movement	Fugl-Meyer ^c UE (20 min) or SAFE score (2 min)
6a and 6b	Motor leg, left and right	0.78	0 = No drift, 1 = drift before 5 s, 2 = falls before 5 s, 3 = no effort against gravity, 4 = no movement	10-meter walk test ^c (5 min) or Fugl-Meyer LE ^c
7	Limb ataxia	-0.16 ^b	0 = Absent, 1 = one limb, 2 = two limbs	NA
8	Sensation	0.50	0 = Normal, 1 = mild loss, 2 = severe loss	Fugl-Meyer Sensory ^c (10 min) or Nottingham Sensory Assessment (60 min)
9	Language/aphasia	0.79	0 = Normal, 1 = mild aphasia, 2 = severe aphasia, 3 = mute or global aphasia	WAB-R bedside form (5 min) or Cookie Thief language analysis (2 min)
10	Dysarthria	0.32 ^b	0 = Normal, 1 = mild, 2 = severe	NA
11	Extinction/inattention	0.61	0 = Normal, 1 = mild, 2 = severe	Star Cancellation Test (3 min) or Cookie Thief neglect analysis (2 min)

Abbreviations: LE = lower extremity; LOC = level of consciousness; MoCA = Montreal Cognitive Assessment Scale; SAFE = Shoulder Abduction Finger Extension; UE = upper extremity; WAB-R = Western Aphasia Battery–Revised.

The degree of interrater agreement expressed as κ may be interpreted as follows: <0, poor agreement; 0 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.8, substantial agreement; 0.81 to 1.0, near perfect agreement.

^a Interrater agreement (κ values) as reported by Goldstein et al. (1989).⁹

^b Items were considered unreliable (NA) for use as screening criteria based on κ values < 0.5.

^c Aligned with Stroke Recovery and Rehabilitation Roundtable and International Stroke Genetics Consortium Global Alliance recommendations for core measures.

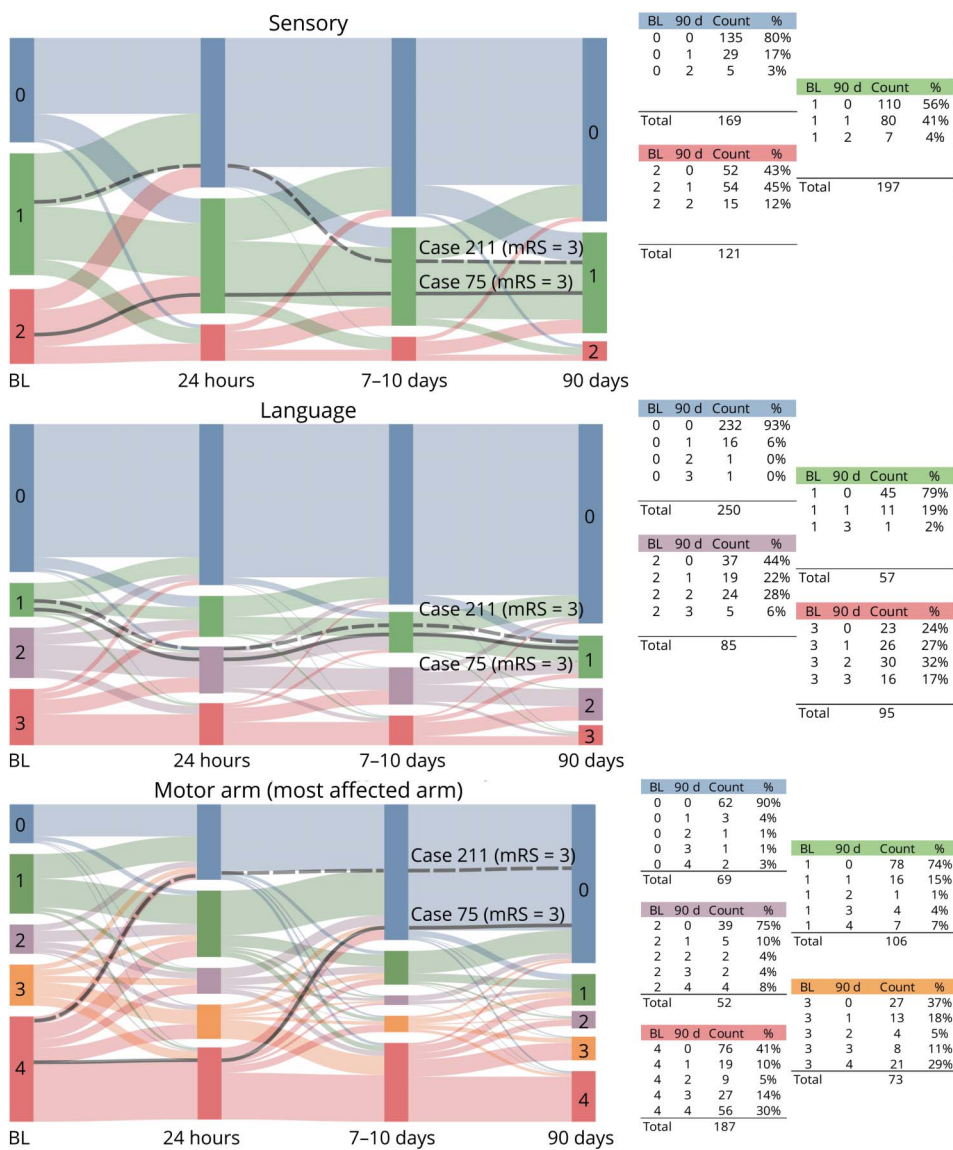
^d Exploratory.

NIHSS items to finer-grained follow-up measures is primarily on the basis of face validity and content validity without formal assessments for concurrent validity. Continuous variables were also favored to optimize statistical power and to avoid the pitfall of calculating percentages on ordinal or ratio data. The rationale for pairing items on the NIHSS to more detailed follow-up measures can be thought of as how they may load onto the NIHSS. A coarser measure like the NIHSS score can explain a portion of the variance across multiple finer-grained measures. The benefit of the coarser tool is that it can be given “at scale” even when subjects number in the thousands, while more time-intensive assessments (e.g., the Fugl-Meyer) provide unique information related to impairment in a specific context.

Domain-specific testing options include the Fugl-Meyer upper extremity (FM-UE)¹⁷ and Shoulder Abduction Finger Extension (SAFE) score¹⁸ for upper extremity motor deficits; Fugl-Meyer Sensory examination¹⁷ or Nottingham Sensory Assessment¹⁹ for somatosensory deficits; Montreal Cognitive

Assessment Scale (MoCA)²⁰ (or Telephone MoCA) for cognitive deficits; the Western Aphasia Battery–Revised bedside form or language analysis of the Cookie Thief picture for aphasia; and the Star cancellation test or analysis of the Cookie Thief picture using an alternative approach to assess neglect.^{21–23} For lower extremity motor deficits, at minimum a binary response (yes/no) should be recorded for whether the patient can walk 10 meters. If ambulatory, gait velocity can also be assessed using the 10-meter walk test.^{24,25} For patients who are unable to walk, the Fugl-Meyer lower extremity may be useful.¹⁷ Visual fields and hemispatial attention may warrant testing in tandem when abnormalities in either one are detected, since neglect and visual field cuts can co-occur, are often difficult to disambiguate, and require differing rehabilitation treatments.^{26,27} Several options for assessment of inattention/neglect and visual fields are available, including novel analyses of the NIHSS Cookie Thief picture,²¹ the Kessler Foundation Neglect Assessment Process,²⁸ and portable perimetry for assessing visual fields (available as an iPad

Figure 1 Stroke Outcomes Differ by Domain



Recovery dynamics in the first 90 days poststroke for 487 patients with ischemic stroke are displayed as alluvial plots of the NIH Stroke Scale (NIHSS) component scores. Two representative cases (in black) show how recovery patterns in specific domains can have different effects on global disability scores. Both cases showed the same degree of disability at 90 days with modified Rankin Scale (mRS) = 3 (unable to live alone without help from another person), and despite both making a complete motor recovery, neither was able to return to independent living due to persistent deficits in nonmotor domains. At right, the plotted NIHSS component scores in each domain are stratified by their baseline scores. For example, 75% of patients with baseline motor = 2 recovered to motor = 0 at 90 days. Patients with this recovery profile could potentially contribute to a ceiling effect if recruited in trials to treat motor deficits. An interactive version of figure 1 is available online to facilitate further data exploration of recovery dynamics within domains and variations in recovery patterns between domains (<https://genestroke.wixsite.com/allie-stroke/domainspecificmeasures>). A video tutorial of the interactive version provides an orientation on the interpretation of alluvial plots and the figure's interactive features. BL = baseline (at admission).

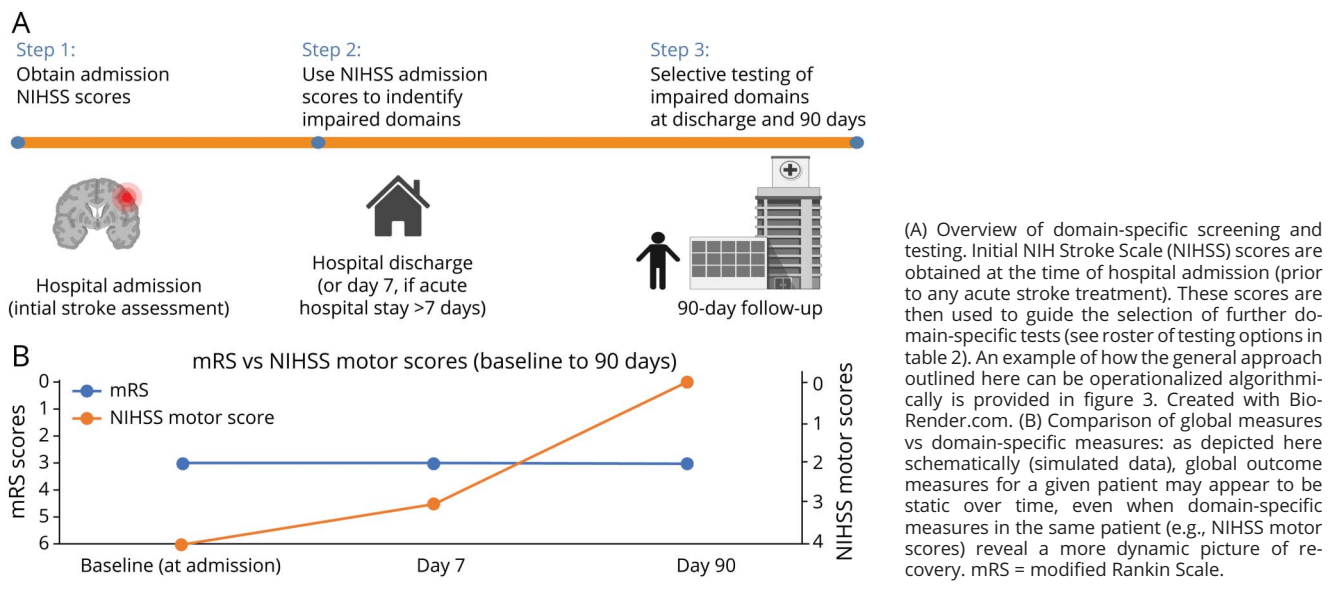
app).²⁹ The general approach to domain-specific test selection detailed here applies equally well to ischemic and hemorrhagic stroke, and can be adapted depending on a given study's research priorities to include outcomes in the sensory, motor, visual, cognitive, or linguistic domains. It may be less well-suited to studies of posterior circulation strokes given that the NIHSS itself is known to be weighted toward detection of anterior circulation symptoms, especially those reflecting left hemisphere damage.³⁰

The measures described in table 2 are geared toward assessing impairment (loss of body structure–function), rather than activity or participation, as defined by the WHO International Classification of Functioning (ICF). Our focus is primarily upon the design of studies to ascertain the biology of stroke recovery, including genetic influences, and we argue that

impairment is the level at which such influences are most likely to be detected. Recovery trajectories for strokes that primarily affect deep white matter (e.g., in hypertensive hemorrhage) are likely to entail biologically distinct repair processes compared to strokes that primarily involve cortex. While our main focus here is on large vessel ischemic stroke due to its prevalence and amenability to available interventions, an important follow-up would be to distinguish differences in biological repair processes and symptom trajectories between ischemic and hemorrhagic stroke. This would be an excellent opportunity to deploy genetic or genomic methods to study such biological differences in recovery mechanisms.

Environmental, social, and behavioral factors are also key determinants of recovery, but are less likely to have

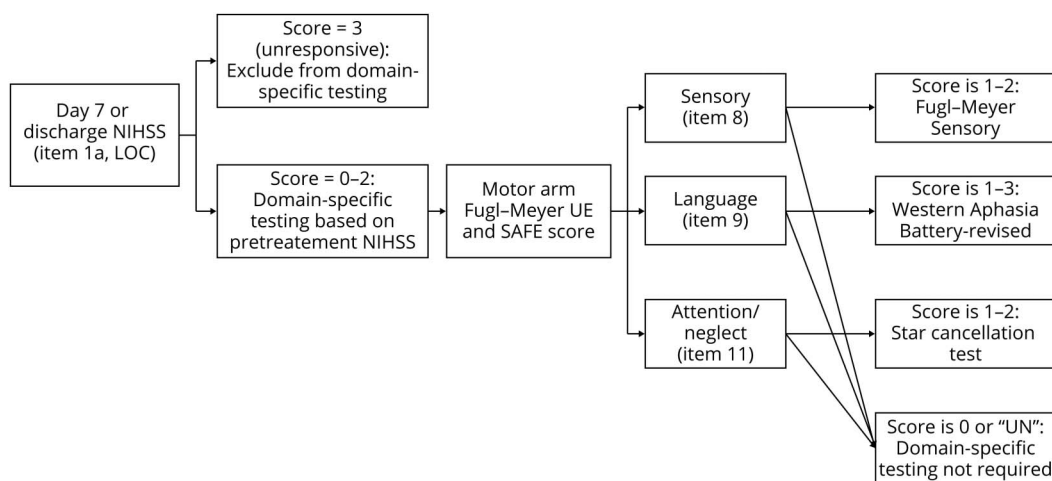
Figure 2 Overview of Domain-Specific Screening and Testing and Comparison of Global Measures vs Domain-Specific Measures



direct biological correlates. It is worth noting that the recovery evaluations of most interest to clinicians and researchers do not always align with those considered most important by patients. Several well-designed scales that assess these elements exist, including The Patient-Reported Outcome Measurement Information System (PROMIS), 36-Item Short Form Survey, EQ5D, and the Stroke Impact Scale,³¹ but are outside the immediate scope of this article.

We recognize that the field of recovery measures continues to evolve, thus there will be a continuing need to adopt new measures that may prove more efficient, reliable, or sensitive than those available at the time of this writing. For example, rapid advances in mobile/wearable technologies and telehealth are poised to offer new means for quantifying recovery and dose–response curves for activity-based therapies. For this reason, a regularly updated version and archival versions of recommended measures will be maintained on the Global

Figure 3 Example of an Algorithm for Domain-Specific Test Selection



This algorithm was developed by our research team for use in a large multisite genetics study of patients with ischemic stroke (Genetics Platform for Acute Stroke Drug Discovery), which uses arm motor recovery as the primary outcome. Arm motor testing is therefore conducted in all participants for this particular study. Patients who are unresponsive at day 7/discharge (i.e., NIH Stroke Scale [NIHSS] 1a = 3) do not undergo further domain-specific testing. LOC = level of consciousness; SAFE = Shoulder Abduction Finger Extension; UE = upper extremity; UN = untestable.

Example of an Algorithm for Domain-Specific Test Selection

In figure 3 we provide an example of how the general approach outlined in figure 2 can be operationalized algorithmically. The algorithm given here was developed by our research group for use in a large multisite genetics study of patients with ischemic stroke with anterior circulation large-vessel occlusion (the Genetics Platform for Acute Stroke [GPAS] Drug Discovery), which uses arm motor recovery as the primary outcome. In the domain-specific approach that we developed for GPAS, patients who are unresponsive at the day 7/discharge time point (i.e., have NIHSS 1a = 3), many of whom will transition to comfort care or hospice, do not undergo further domain-specific testing. For GPAS, arm motor testing is included on all participants since it is the primary outcome measure for this study and a high prevalence of arm motor deficits is expected in our population (likely $\geq 77\%$ based on prior studies).²¹ Adequate time for arm motor testing must therefore be budgeted. At minimum, we obtain the SAFE score, which is quick and simple to administer, and the finger extension component also serves as a proxy for corticospinal tract integrity. More intensive testing with the FM-UE takes approximately 15–20 minutes to administer (see times required for test administration in table 2). Another factor to be considered in study planning is that the FM-UE requires a small amount of testing equipment (reflex hammer and tennis ball). We require online training on the FM-UE for all examiners, which significantly improves inter- and intrarater reliability.³² A structured interview format that improves the reliability of the mRS is also recommended.⁴

The algorithmic approach detailed above could be readily integrated into clinical decision support tools where baseline (pretreatment) NIHSS scores would be input, and a list of domain-specific tests tailored for each patient could be generated. At present our algorithm is implemented manually and domain-specific tests are selected on a case-by-case basis, but ultimately for broader use this could be automated as a proprietary add-in to electronic medical records, or as a stand-alone app at the point of care. An additional feature of such an automated algorithm could be to ensure that the nonmotor domains are captured with adequate sample sizes taking into account the minimum clinically important difference for each test, as well as anticipated effect sizes.

Calculating the Degree of Recovery

To assess the degree of recovery, there are several potential approaches to be considered. The first and simplest is to calculate the degree of recovery as a percentage of the maximum possible score (S in the equation below) on a given test,

using the difference in day 7 vs day 90 values to calculate a percentage of available recovery achieved. This calculation essentially yields a scaled value, where an individual's remaining possible change is divided by the numerically achievable change available to that individual.

Formula 1 Percentage of available recovery achieved

$$1 - (S_{max} - S_{90d}) / (S_{max} - S_{7d}) * 100$$

This approach to calculating recovery finds precedent in other recent large-scale stroke recovery trials, such as those from the ENIGMA stroke recovery working group. Converting the scores to percentages as is done here offers one means to achieve a common scale for combining measures. Alternatively, Z scores could be used, but present an additional challenge when deciding the reference population to which scores should be standardized.

The approach given by formula 1 has several benefits, in that it allows domain-specific measures to be combined, similar to the Multiple Sclerosis Functional Composite score,³³ which combines measures of ambulation, hand/arm function, and cognition. It also considers the fact that impairment changes of a given magnitude may not carry the same practical significance at all points on a particular scale, a matter that is increasingly recognized in clinical trial design. For example, in the RATULS trial (Robot Assisted Training for the Upper Limb After Stroke), definitions of treatment “success” were scaled according to treatment effects that would be clinically meaningful relative to the patient's impairment severity.³⁴

Drawbacks to this approach are that it is not well-suited for use with ordinal scales that have unequal intervals between values, and results can be misinterpreted, especially in visual plots, unless the relative nature of the scale is taken into account. The impression may be that severely impaired patients who make large improvements have quantitatively less recovery than mildly impaired patients who make small improvements. If such scaled scores are used, we recommend that they be reported in tandem with their constituent nonscaled scores to limit the influence of this bias on interpretation.

An improvement on the above strategy for calculating the degree of recovery might be to use an approach based on regression rather than difference scores, such that 90-day and 7-day scores are normalized to the total points available on that scale, and the 7-day normalized scores are then controlled for as a baseline:

Formula 2 Percentage impairment at day 90, controlling for baseline impairment

$$([S_{90d}/S_{max}] * 100) \sim \beta_0 + \beta_1 ([S_{7d}/S_{max}] * 100)$$

This approach ameliorates some of the problems with using difference scores as in formula 1, but comes at the cost of

some additional statistical complexity, faces the same issues with treating ordinal data as interval or ratio data, and focuses on end points rather than change over time. To capture change over time, formal longitudinal models of normalized scores could be used (rather than simple difference scores or end points), but again, this comes with a trade-off of increased analytical complexity.^{35,36}

Thus, neither formula 1 nor formula 2 presents a perfect solution. With the caveats above, they offer a reasonable means—though certainly not the only means—to calculate recovery when pooling data from different scales. Different methods have also been proposed that warrant further critical evaluation, for example latent variable approaches, which conceive of recovery as a manifold in the multidimensional space of distinct structural/functional measures.^{37,38} Alternatively, there could be advantages to using adaptive testing, which would enhance efficiency, but to our knowledge few or no adaptive evaluations equivalent to the tests outlined in table 2 exist. A computerized adaptive testing system of the Fugl-Meyer motor scale has been evaluated,³⁹ but has not yet been broadly adopted or validated in other studies. Clearly the question of how best to quantitatively represent recovery is an area of active exploration, and the long-term solution requires continued study.

Practicalities and Advantages of Implementing a Domain-Specific Approach

From an implementation science point of view, domain-specific measures need to be simple and accessible, but sufficiently detailed to capture dynamic changes in the individual neural systems that contribute to stroke recovery. It is essential that they are implemented in a standardized format with rigorous training to assure high-quality data acquisition and are scored at least twice during recovery. We have outlined a method to accomplish this by using commonly collected clinical data from the NIHSS to guide the selection of domain-specific outcome testing.

Our proposed approach complements more global measures of disability such as the mRS by providing additional insight on how impairments in distinct domains each contribute to a patient's overall disability. Global and domain-specific measures are collected in parallel. In addition to capturing the dynamics of recovery more fully, collecting domain-specific data about impairments is essential for tracking patients' response to therapies. Ultimately, such metrics can aid in precision matching of deficits to targeted rehabilitation therapies. This is especially important in stroke rehabilitation trials, where patient selection and outcomes often focus on a specific domain, and rehabilitation therapies are also targeted to treat specific deficits (e.g., ankle dorsiflexor stimulation to improve

distal leg weakness, or surface EMG biofeedback to retrain swallowing).

Because many analyses, genetic or otherwise, will focus on specific recovery domains (e.g., motor or language), patients who have available data on those more detailed recovery measures can be combined for valid statistical analysis. When evaluating determinants of more global recovery, patients without deficits in particular domains would have missing data for those domains, for example, a patient with motor deficits but no language deficits on the NIHSS. One reasonable assumption might be that detailed language testing in this patient would not be especially informative, and could therefore sensibly be imputed as 0; modest deficits on the more detailed measure would likely be lacking in practical interest for recovery research, and finding any improvement large enough to unveil an associated biological mechanism would be unlikely with such modest deficits. An alternate approach would be to perform all of the finer-grained measures in a subset of patients to obtain a basis for more accurate imputation of missing data.

Domain-Specific Testing Can Inform Clinical Trial Design

A number of restorative therapies are currently under development.⁴⁰ Many of these target patients during the first few days poststroke, a time when the brain may be particularly galvanized to boost clinically useful plasticity in response to therapy.⁴¹ Restorative trials most often rely on domain-specific tests because recovery is linked with experience-dependent plasticity, and the experiences that constitute restorative therapies are themselves often domain-specific (e.g., treadmill training to improve leg motor function, or melodic intonation to improve language).

Domain-specific tests can also play an important role in clinical trial design by ensuring that the patients enrolled are likely to have at least some minimal degree of persistent treatable deficit in the target domain. Enrolling a high fraction of patients with excellent recovery makes it difficult to detect a difference between active and placebo therapies. Therefore, predicting a domain's late value (e.g., at 90 days) based on an early value (e.g., 24 hours poststroke, when many restorative trials enroll patients) can be of critical importance to identify suitable patients for hypothesis testing and to limit the enrollment of those patients who would be likely to have excellent 90-day outcomes independent of any study treatment. Conversely, domain-specific testing could also help to limit the enrollment of patients whose 90-day behavioral outcomes are likely to be abjectly poor, among whom it can also be difficult to demonstrate a treatment effect. A necessary first step toward developing such tools for study enrollment will be large-scale, high-quality, consistently gathered data and robust validation. The ISGC is planning a large-scale trial (GPAS Drug Discovery) that presents a prime opportunity for initial predictive tool derivation and validation.

Other Potential Uses for NIHSS Data

In addition to our proposed use of the NIHSS to screen for impaired domains and guide selective follow-up testing, we also recognize the potential for the NIHSS itself, either as total scores or component scores, to be used as a primary outcome measure. This approach warrants careful consideration, as is further discussed below. Proponents assert that the NIHSS is a good measure for acute stroke trials because it reflects impairment per se, and impairment (loss of body structure–function) is the WHO ICF level at which acute stroke treatment effects are likely to be the most pronounced. To the extent that postacute rehabilitation also aims for reversal of impairment (versus behavioral compensation), the same argument applies for stroke recovery and rehabilitation trials. Recently, Chalos et al.⁴² used a causal mediation model to assess whether early total NIHSS scores can serve as a surrogate end point for 3-month mRS scores. They demonstrated that total NIHSS at 1 week satisfied the Prentice criteria for surrogate end points with respect to the mRS, and qualifies for use as a primary outcome measure in this regard. There is also some precedent for using component NIHSS scores, either as outcome measures in and of themselves, or as predictors of more global outcome metrics.⁴³

Use of the NIHSS component scores themselves as outcome metrics could arguably offer a way to shift the focus in stroke research from global outcome measures to a more domain-specific perspective by allowing extraction of finer-grained impairment metrics from existing large datasets without the need for extensive additional testing. For example, in our work with the ISGC, ISGC stroke databanks could be queried using NIHSS scores stratified by domain to examine genetic determinants of recovery. This would allow investigators to approach questions about how neurobiological factors may differentially influence the sensory, motor, cognitive, and linguistic domains during stroke recovery and rehabilitation. This kind of approach is also highly relevant to the work of the Genetics of Ischaemic Stroke Functional Outcomes network, founded in 2017 to conduct the first international genome-wide association study of poststroke functional outcomes.⁴⁴

Cautions and Considerations

Before adopting any approach based on NIHSS component scores, it is important to consider the clinometric properties of the NIHSS items. While the overall interrater reliability for the total NIHSS has been reported at intraclass correlation coefficient (ICC) 0.95, the κ value for component scores varies broadly, from -0.16 to 0.79 (table 2). Those items with poor agreement ($\kappa < 0$) or slight agreement ($\kappa = 0$ – 0.20) stand in marked contrast to other domain-specific measures such as the FM-UE where—especially with diligent attention to rater training—ICC values as high as 0.98 to 0.99 have been reported.³² It could be argued, given the poor reliability for some of the NIHSS items and their lack of

granularity in many domains (perhaps most notably cognition), that another instrument (or combination of instruments) to more fully capture all important domains of function should be used at the outset of future stroke trials. In addition to considerations of clinometric properties, it must be noted that the SRRR consensus recommendations caution against reliance on the NIHSS as an outcome measure, instead placing preference on more detailed metrics. However, the tracking of NIHSS component scores could provide a readily adopted, albeit imperfect, first approach for acute stroke researchers to begin addressing the lack of detailed patient-level data on recovery dynamics as long as interpretations are tempered by the considerations mentioned above.

The lack of domain-specific outcome data in most large-scale clinical stroke trials reflects a reliance on global outcome measures, like the mRS, which are unable to delineate how deficits in specific domains contribute to a patient's overall level of disability. We have developed a practical means to streamline the testing and tracking of patients' recovery trajectories in a domain-specific manner. Our overarching goal is to make the implementation of domain-specific testing more feasible for large-scale clinical trials on stroke recovery. This approach complements, rather than supplants, more global outcome measures, and has potential to yield new discoveries and actionable insights that are relevant not only to acute stroke care but also to stroke rehabilitation. It also presents an opportunity to create a more unified research culture that spans the continuum of care, and thus can more fully capture the dynamics of stroke recovery as they occur over time.

Study Funding

Dr. Braun: NIH K12HD093427. Dr. Cole: NIH R01NS114045, American Heart Association, and Bayer Pharmaceuticals 17IBDG33700328. Dr. Cramer: NIH U01 NS086872, R01 NR015591, R01 HD095457, and R01HD062744. Dr. de Havenon: NIH K23NS105924. Dr. Heitsch: NIH K23NS099487-02. Dr. Lindgren: Swedish Government (under the “Avtal om Läkarutbildning och Medicinsk Forskning”), Swedish Heart and Lung Foundation, Region Skåne, Lund University, Skåne University Hospital, Sparbanksstiftelsen Färs och Frosta, and Freemasons Lodge of Instruction Eos in Lund. Dr. Worrall: NIH KL2 TR003016, R21 NS106480, U24 NS107222, the Australian-American Fulbright Commission, and the University of Newcastle as the 2018 Fulbright Distinguished Chair in Health.

Disclosure

Dr. Cramer performs consulting with AbbVie, Constant Therapeutics, MicroTransponder, Neuroolutions, Regenera, SanBio, Stemedica, Fujifilm Toyama Chemical Co., Biogen, and TRCare. Dr. de Havenon receives funding from AMAG and Regeneron pharmaceuticals for investigator-initiated clinical stroke research. Dr. Lindgren receives personal fees from Bayer, Astra Zeneca, BMS Pfizer, and Portola.

Dr. Worrall serves as Deputy Editor for *Neurology*[®]. The other authors have no relevant disclosures. Go to Neurology.org/N for full disclosures.

Publication History

Received by *Neurology* May 6, 2020. Accepted in final form April 20, 2021.

Appendix Authors

Name	Location	Contribution
Robynne G. Braun, MD, PhD	Department of Neurology, University of Maryland, Baltimore	Took the lead in writing the manuscript; conceived of the interactive data visualization and created the static figures; conceived of the presented idea with S.C. Cramer and B.B. Worrall; performed the analyses with input from A. de Havenon and L. Heitsch; wrote the manuscript and developed the theoretical framework with support from L. Heitsch, J.W. Cole, A.G. Lindgren, A. de Havenon, S.C. Cramer, and B.B. Worrall; provided critical feedback and helped shape the research, analysis, and manuscript
Laura Heitsch, MD	Department of Emergency Medicine, Washington University, St. Louis, MO	Performed the preliminary analyses; contributed to the manuscript and developed the theoretical framework with R.G. Braun, A.G. Lindgren, A. de Havenon, S.C. Cramer, and B.B. Worrall; provided critical feedback and helped shape the research, analysis, and manuscript
John W. Cole, MD, MS	Department of Neurology, University of Maryland, Baltimore	Contributed to the manuscript and developed the theoretical framework with R.G. Braun, L. Heitsch, A.G. Lindgren, A. de Havenon, S.C. Cramer, and B.B. Worrall; provided critical feedback and helped shape the research, analysis, and manuscript
Arne G. Lindgren, MD, PhD	Department of Clinical Sciences Lund, Neurology, Lund University, Sweden	Contributed to the manuscript and developed the theoretical framework with R.G. Braun, L. Heitsch, J.W. Cole, A. de Havenon, S.C. Cramer, and B.B. Worrall; provided critical feedback and helped shape the research, analysis, and manuscript
Adam de Havenon, MD	Department of Neurology, University of Utah, Salt Lake City	Performed the preliminary analyses; contributed to the manuscript and developed the theoretical framework with R.G. Braun, L. Heitsch, J.W. Cole, A.G. Lindgren, S.C. Cramer, and B.B. Worrall; provided critical feedback and helped shape the research, analysis, and manuscript

Appendix (continued)

Name	Location	Contribution
Jason A. Dude, BS	Department of Health and Kinesiology, University of Utah, Salt Lake City	Generated the interactive data visualization with direction from R.G. Braun and K.R. Lohse; provided critical feedback and helped shape the research, analysis, and manuscript
Keith R. Lohse, PhD	Department of Health and Kinesiology, University of Utah, Salt Lake City	Contributed to the manuscript and supervised the development of the interactive data visualization with R.G. Braun and J. Dude; provided critical feedback and helped shape the research, analysis, and manuscript
Steven C. Cramer, MD	Department of Neurology, University of California, Los Angeles; California Rehabilitation Institute, Los Angeles	Conceived of the presented idea with R.G. Braun and B.B. Worrall; supervised the project; contributed to the manuscript and developed the theoretical framework with R.G. Braun, L. Heitsch, J.W. Cole, A.G. Lindgren, A. de Havenon, and B.B. Worrall; provided critical feedback and helped shape the research, analysis, and manuscript
Bradford B. Worrall, MD, MSc	Departments of Neurology and Public Health Sciences, University of Virginia, Charlottesville	Conceived of the presented idea with R.G. Braun and S.C. Cramer; supervised the project; contributed to the manuscript and developed the theoretical framework with R.G. Braun, L. Heitsch, J.W. Cole, A.G. Lindgren, A. de Havenon, and S.C. Cramer; provided critical feedback and helped shape the research, analysis, and manuscript

References

- Hicks KA, Mahaffey KW, Mehran R, et al. 2017 Cardiovascular and stroke endpoint definitions for clinical trials. *Circulation*. 2018;137(9):961-972.
- Grinnon ST, Miller K, Marler JR, et al. National Institute of Neurological Disorders and Stroke common data element project: approach and methods. *Clin Trials*. 2012; 9(3):322-329.
- Taylor-Rowan M, Wilson A, Dawson J, Quinn TJ. Functional assessment for acute stroke trials: properties, analysis, and application. *Front Neurol*. 2018;9:191.
- Wilson JTL, Hareendran A, Hendry A, Potter J, Bone I, Muir KW. Reliability of the modified Rankin Scale across multiple raters: benefits of a structured interview. *Stroke*. 2005;36(4):777-781.
- National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med*. 1995;333(24): 1581-1587.
- Broderick JP, Palesch YY, Demchuk AM, et al. Endovascular therapy after intravenous t-PA versus t-PA alone for stroke. *N Engl J Med*. 2013;368(10):893-903.
- Ginsberg MD, Palesch YY, Hill MD, et al. High-dose albumin treatment for acute ischaemic stroke (ALIAS): a phase 3, randomised, double-blind, placebo-controlled trial. *Lancet Neurol*. 2013;12(11):1049-1058.
- Berkhemer OA, Fransen PSS, Beumer D, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med*. 2015;372(1):11-20.
- Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH Stroke Scale. *Arch Neurol*. 1989;46(6):660-662.
- Spilker J, Kongable G, Barch C, et al. Using the NIH Stroke Scale to assess stroke patients: the NINDS rt-PA stroke study group. *J Neurosci Nurs*. 1997;29(6): 384-392.
- Hong Y, LaBresh KA. Overview of the American Heart Association "Get with the Guidelines" programs: coronary heart disease, stroke, and heart failure. *Crit Pathw Cardiol*. 2006;5(4):179-186.

12. Cramer SC, Koroshetz WJ, Finklestein SP. The case for modality-specific outcome measures in clinical trials of stroke recovery-promoting agents. *Stroke*. 2007;38(4):1393-1395.
13. Felberg RA, Okon NJ, El-Mitwalli A, Burgin WS, Grotta JC, Alexandrov AV. Early dramatic recovery during intravenous tissue plasminogen activator infusion: clinical pattern and outcome in acute middle cerebral artery stroke. *Stroke*. 2002;33(5):1301-1307.
14. Mikulik R, Dusek L, Hill MD, et al. Pattern of response of National Institutes of Health Stroke Scale components to early recanalization in the CLOTBUST trial. *Stroke*. 2010;41(3):466-470.
15. Kwakkel G, van Wegen EEH, Burridge JH, et al. Standardized measurement of quality of upper limb movement after stroke: consensus-based core recommendations from the second stroke recovery and rehabilitation roundtable. *Neurorehabil Neural Repair*. 2019;33(11):951-958.
16. Wallace SJ, Worrall L, Rose T, et al. A core outcome set for aphasia treatment research: the ROMA consensus statement. *Int J Stroke*. 2019;14(2):180-185.
17. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The poststroke hemiplegic patient: 1: a method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7(1):13-11.
18. Nijland Rinske HM, van Wegen Erwin EH, Harmeling-van der Wel Barbara C, Kwakkel G; EPOS Investigators. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery. *Stroke*. 2010;41(4):745-750.
19. Lincoln N, Crow J, Jackson J, Waters G, Adams S, Hodgson P. The unreliability of sensory assessments. *Clin Rehabil*. 1991;5(4):273-282.
20. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699.
21. Agis D, Goggins MB, Oishi K, et al. Picturing the size and site of stroke with an expanded National Institutes of Health Stroke Scale. *Stroke*. 2016;47(6):1459-1465.
22. Halligan PW, Marshall JC, Wade DT. Visuospatial neglect: underlying factors and test sensitivity. *Lancet*. 1989;2(8668):908-911.
23. Wilson B, Cockburn J, Halligan P. Development of a behavioral test of visuospatial neglect. *Arch Phys Med Rehabil*. 1987;68:98-101.
24. Watson M. Refining the ten-metre walking test for use with neurologically impaired people. *Physiotherapy*. 2002;88(7):386-397.
25. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. *J Am Geriatr Soc*. 2006;54(5):743-749.
26. Barrett AM, Houston KE. Update on the clinical approach to spatial neglect. *Curr Neurol Neurosci Rep*. 2019;19(5):25.
27. Parton A, Malhotra P, Husain M. Hemispatial neglect. *J Neurol Neurosurg Psychiatry*. 2004;75(1):13-21.
28. Chen P, Chen CC, Hreha K, Goedert KM, Barrett AM. Kessler Foundation Neglect Assessment Process uniquely measures spatial neglect during activities of daily living. *Arch Phys Med Rehabil*. 2015;96(5):869-876.
29. Vingrys AJ, Healey JK, Liew S, et al. Validation of a tablet as a tangent perimeter. *Transl Vis Sci Technol*. 2016;5(4):3.
30. Woo D, Broderick JP, Kothari RU, et al. Does the National Institutes of Health Stroke scale favor left hemisphere strokes? NINDS t-PA stroke study group. *Stroke*. 1999;30(11):2355-2359.
31. Salinas J, Sprinkhuizen SM, Ackerson T, et al. An international standard set of patient-centered outcome measures after stroke. *Stroke*. 2016;47(1):180-186.
32. See J, Dodakian L, Chou C, et al. A standardized approach to the Fugl-Meyer assessment and its implications for clinical trials. *Neurorehabil Neural Repair*. 2013;27(8):732-741.
33. Fischer JS, Rudick RA, Cutter GR, Reingold SC. The Multiple Sclerosis Functional Composite measure (MSFC): an integrated approach to MS clinical outcome assessment: National MS Society Clinical Outcomes Assessment Task Force. *Mult Scler*. 1999;5(4):244-250.
34. Rodgers H, Bosomworth H, Krebs HI, et al. Robot Assisted Training for the Upper Limb After Stroke (RATULS): a multicentre randomised controlled trial. *Lancet*. 2019;394(10192):51-62.
35. Garcia TP, Marder K. Statistical approaches to longitudinal data analysis in neurodegenerative diseases: Huntington's disease as a model. *Curr Neurol Neurosci Rep*. 2017;17(2):14.
36. Lohse K, Shen J, Kozlowski A. Modeling longitudinal outcomes: a contrast of two methods. *J Mot Learn Dev*. 2019;8:1-21.
37. Hommel M, Detante O, Favre I, Touzé E, Jaillard A. How to measure recovery? Revisiting concepts and methods for stroke studies. *Transl Stroke Res*. 2016;7(5):388-394.
38. Saver JL. Optimal endpoints for acute stroke therapy trials: best ways to measure treatment effects of drugs and devices. *Stroke J Cereb Circ*. 2011;42(8):2356-2362.
39. Hou WH, Shih CL, Chou YT, et al. Development of a computerized adaptive testing system of the Fugl-Meyer motor scale in stroke patients. *Arch Phys Med Rehabil*. 2012;93(6):1014-1020.
40. Lin DJ, Finklestein SP, Cramer SC. New directions in treatments targeting stroke recovery. *Stroke*. 2018;49(12):3107-3114.
41. Cramer SC, Sur M, Dobkin BH, et al. Harnessing neuroplasticity for clinical applications. *Brain J Neurol*. 2011;134(pt 6):1591-1609.
42. Chalos V, van der Ende NAM, Lingsma HF, et al. National Institutes of Health Stroke Scale: an alternative primary outcome measure for trials of acute treatment for ischemic stroke. *Stroke*. 2020;51:282-290.
43. Abdul-Rahim AH, Fulton RL, Sucharew H, et al. National Institutes of Health Stroke Scale item profiles as predictor of patient outcome: external validation on safe implementation of thrombolysis in stroke-monitoring study data. *Stroke*. 2015;46(10):2779-2785.
44. Söderholm M, Pedersen A, Lorentzen E, et al. Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology*. 2019;92(12):e1271-e1283.

Manage Your Career | Recruit Top Talent

The AAN's Neurology Career Center is the largest job site specifically for neurologists. Visit careers.aan.com to find your next hire or search from hundreds of open positions in neurology.

Access the 2021 Annual Meeting Anytime, Anywhere

Access 300+ hours of content, 140+ programs with syllabi, 25+ exclusive bonus presentations, and integrated CME testing—anytime and anywhere through March 31, 2022—with 2021 Annual Meeting On Demand! Visit AAN.com/AMOD to order yours today.
