National institutes of health stroke scale an alternative primary outcome measure for trials of acute treatment for ischemic stroke

Chalos, V.; Ende, N.A.M. van der; Lingsma, H.F.; Mulder, M.J.H.L.; Venema, E.; Dijkland, S.A.; ... ; MR CLEAN Investigators

National Institutes of Health Stroke Scale
An Alternative Primary Outcome Measure for Trials of Acute Treatment for Ischemic Stroke

Vicky Chalos, MD; Nadinda A.M. van der Ende, MD; Hester F. Lingsma, PhD;
Maxim J.H.L. Mulder, MD, PhD; Esmee Venema, MD; Simone A. Dijkland, MD;
Olvert A. Berkhemer, MD, PhD; Albert J. Yoo, MD, PhD; Joseph P. Broderick, MD, PhD;
Yuko Y. Palesch, PhD; Sharon D. Yeatts, PhD; Yvo B.W.E.M. Roos, MD, PhD;
Robert J. van Oostenbrugge, MD, PhD; Wim H. van Zwam, MD, PhD;
Charles B.L.M. Majoie, MD, PhD; Aad van der Lugt, MD, PhD; Bob Roozenbeek, MD, PhD;
Diederik W.J. Dippel, MD, PhD; on behalf of the MR CLEAN Investigators*

Background and Purpose—The modified Rankin Scale (mRS) at 3 months is the most commonly used primary outcome measure in stroke treatment trials, but it lacks specificity and requires long-term follow-up interviews, which consume time and resources. An alternative may be the National Institutes of Health Stroke Scale (NIHSS), early after stroke. Our aim was to evaluate whether the NIHSS assessed within 1 week after treatment could serve as a primary outcome measure for trials of acute treatment for ischemic stroke.

Methods—We used data from 2 randomized controlled trials of endovascular treatment for ischemic stroke: the positive MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands; N=500) and the neutral IMS (Interventional Management of Stroke) III trial (N=656). We used a causal mediation model, with linear and ordinal logistic regression adjusted for confounders, to evaluate the NIHSS 24 hours and 5 to 7 days after endovascular treatment as primary outcome measures (instead of the mRS at 3 months) in both trials. Patients who had died before the NIHSS was assessed received the maximum score of 42. NIHSS+1 was then log10-transformed.

Results—In both trials, there was a significant correlation between the NIHSS at 24 hours and 5 to 7 days and the mRS. In MR CLEAN, we found a significant effect of endovascular treatment on the mRS and on the NIHSS at 24 hours and 5 to 7 days. After adjustment for NIHSS at 24 hours and 5 to 7 days, the effect of endovascular treatment on the mRS decreased from common odds ratio 1.68 (95% CI, 1.22–2.32) to respectively 1.36 (95% CI, 0.97–1.91) and 1.24 (95% CI, 0.87–1.79), indicating that treatment effect on the mRS is in large part mediated by the NIHSS. In the IMS III trial there was no treatment effect on the NIHSS at 24 hours and 5 to 7 days, corresponding with the absence of a treatment effect on the mRS.

Conclusions—The NIHSS within 1 week satisfies the requirements for a surrogate end point and may be used as a primary outcome measure in trials of acute treatment for ischemic stroke, particularly in phase II(b) trials. This could reduce stroke-outcome assessment to its essentials (ie, neurological deficit), and reduce trial duration and costs. Whether and under which conditions it could be used in phase III trials requires a debate in the field with all parties.

Clinical Trial Registration—URL: http://www.isrctn.com. Unique identifier: ISRCTN10888758; https://www.clinicaltrials.gov. Unique identifier: NCT00359424. (***Stroke***. 2020;51:282-290. DOI: 10.1161/STROKEAHA.119.026791.)

Key Words: endovascular treatment ◼ informed consent ◼ NIHSS ◼ outcome ◼ stroke ◼ thrombectomy

Acute treatment for ischemic stroke has been rapidly evolving over the past 5 years, resulting in a drastic improvement of functional outcome after ischemic stroke in selected patients. However, a considerable number of patients do not recover to functional independence after acute treatment or are still not eligible for acute treatment.[1–3] To improve outcome and expand patient selection, new treatments or modifications to existing treatment modalities are continuously being tested in novel (randomized) clinical trials. One of the most important considerations in the design of a valid and useful clinical trial is the selection of an appropriate primary outcome measure.[4]

The most commonly used primary outcome measure in (ischemic) stroke treatment trials is the modified Rankin Scale (mRS). This 7-point ordinal scale describes the degree of global disability or dependence in daily life after stroke, that is, functional outcome.[5] It is known for its simplicity and its ease of interpretation.[6,7] However, the mRS has important practical limitations. Because the mRS measures functional outcome and has a floor effect in the acute setting (ie, patients will receive mRS scores of 4 or 5 because they are often bed-bound during hospital admission), it should ideally be assessed after patients have had the chance to resume their daily activities; typically after 3 months.[7,8] This long time span between treatment and outcome assessment may require intensive efforts to track down patients leading to increased trial duration and costs. Another undesirable result of this long time span is the risk of loss to follow-up. When investigators are reluctant to enroll patients who are at high risk for loss to follow-up, for example, because of socioeconomic factors or visitors from abroad, this could also lead to slower patient enrollment and selection bias.

Because of the increasing interest in—and the need for—rapid improvements in the acute treatment for ischemic stroke, efficient and cost-effective testing of new treatments is essential, especially for phase II(b) clinical trials, which are trials that are conducted to assess the efficacy of new treatments. Thus, early (surrogate) outcome measures are preferable for this purpose.

An alternative that may obviate the practical limitations of the mRS—the National Institutes of Health Stroke Scale (NIHSS)—is frequently used as an early secondary outcome measure in stroke trials. It is usually assessed 24 hours or 5 to 7 days after the treatment. It measures neurological deficit rather than functional outcome. NIHSS scores range from 0 to 42, with higher scores indicating more severe neurological deficit.[9] The NIHSS has a high intraobserver and interobserver reliability after only a few hours of training, is easy and quick to assess, and is a valid measure of stroke severity.[6,7,9] It reflects cerebral dysfunction by assessing several clinical items and is responsive to meaningful clinical change.[6,9] Importantly, early NIHSS scores have a strong prognostic value for long-term functional outcome after stroke.[10–12] However, the strong correlation between NIHSS and mRS scores does not ensure that the NIHSS is a valid surrogate end point (ie, able to replace the mRS as a measure of treatment effect). A surrogate end point should lie in the causal pathway between the intervention and the true end point.[13]

We used data from a positive and a neutral randomized controlled trial (RCT) of endovascular treatment (EVT) for ischemic stroke to evaluate whether the NIHSS within the first week after treatment could serve as a primary outcome measure for trials of acute treatment for ischemic stroke.

Methods

Data: MR CLEAN and IMS III Trial

Data were obtained from MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands)[14] and the IMS (Interventional Management of Stroke) III trial.[15] Anonymized trial data and methods that support our study findings are available for MR CLEAN upon reasonable request to mrclean@erasmusmc.nl and via the public dataset through National Institutes of Health for the IMS III trial (https://www.ninds.nih.gov/Current-Research/Research-Funded-NINDS/Clinical-Research/Archived-Clinical-Research-Datasets).

MR CLEAN was a phase III, multicenter, open-label RCT that evaluated the efficacy and safety of EVT plus usual care (intervention) compared with usual care alone (control) in patients with acute ischemic stroke. MR CLEAN enrolled 500 patients from 16 intervention centers in the Netherlands between December 2010 and March 2014. Enrolled patients were aged ≥18 years, had an ischemic stroke due to an intracranial large vessel occlusion in the anterior circulation with an NIHSS score of ≥2, and were able to undergo EVT within 6 hours after symptom onset. The central medical ethics committee and research board of each participating center approved this study. All patients or their legal representatives provided written informed consent before randomization.

The IMS III trial was a phase III, multicenter, open-label RCT, evaluating whether EVT combined with intravenous thrombolysis (IVT) with recombinant tissue-type plasminogen activator in a dose of 0.6 mg/kg (intervention) within 3 hours of symptom onset was superior to IVT alone (control). The IMS III trial enrolled 656 patients from 58 international centers between August 2006 and April 2012, aged 18 to 80 years with a moderate-to-severe ischemic stroke (NIHSS ≥10) before initiation of IVT. The study was approved by the ethics committee and research board of each participating center. Written informed consent was obtained from patients or their legal representative before enrollment in the study.

Causal Mediation Model

The causal pathway between the intervention (ie, treatment) and the true end point can be assessed with the Prentice criteria for surrogate end points,[16] similar to the mediation model described by Baron and Kenny.[17] Statistical validation of surrogate end points requires at least 4 conditions to be satisfied (Figure):

1. There is a significant treatment effect on the true end point (pathway c);
2. There is a significant treatment effect on the surrogate end point (pathway a);
3. There is a significant correlation between the surrogate end point and the true end point while controlling for treatment (pathway b);
4. The surrogate end point mediates the effect of treatment on the true end point, that is, the significant treatment effect on the true end point becomes not statistically significant or should be reduced (ie, partial mediation) after adjusting for the surrogate end point (pathway c′).[16,18]

In the current analysis, the treatment was EVT. The ordinal mRS at 3 months was considered the true end point. In both trials, the mRS was assessed by independent assessors blinded to treatment allocation. The NIHSS at 24 hours and at 5 to 7 days (or at discharge if earlier) after EVT were considered the potential surrogate end points, also called the mediating variables. In both trials, NIHSS scores were assessed by the treating physician.

This traditional approach of mediation ignores the issue of confounding, which may also occur in RCTs. As a result of randomization, pathway a and pathway c could be assumed to be free of confounding. However, pathway b may contain (known and unknown)
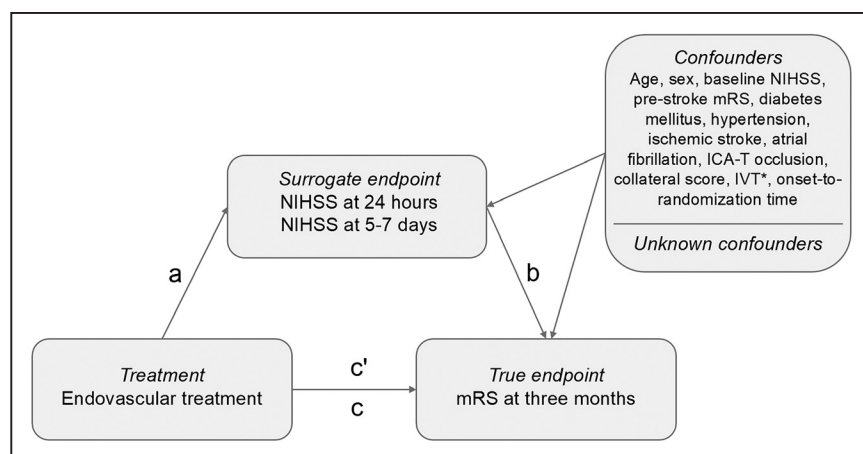
Figure. Applied causal mediation model. ICA-T indicates internal carotid artery terminus; IVT, intravenous thrombolysis; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale. *In MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands) only.

confounders because both the surrogate end points and true end point are outcomes of randomization.[18] Therefore, we adjusted for known confounders in pathway b and c′ (Figure).

The causal mediation model was applied to a trial with a positive treatment effect (MR CLEAN) and to a neutral trial (IMS III) to evaluate whether consistent relationships between pathway a and c were observed across the 2 trials.

Statistical Analyses

We compared baseline characteristics of patients in the intervention group versus the control group for both trials using descriptive statistics. Pathway a and c were tested with univariable linear and ordinal logistic regression, respectively. All pathways were tested with multivariable linear (pathway a) and ordinal logistic regression (pathway b, c, c′). Patients who had died before the time point of NIHSS assessment was reached, received the maximum NIHSS score of 42. NIHSS scores at 24 hours and 5 to 7 days were then log10-transformed to better meet the assumption of normally distributed residuals in linear regression, after adding 1 point to all NIHSS scores, so that the log10-transformed NIHSS score of 0 would remain 0. The mRS did not require transformation, as it was analyzed with ordinal logistic regression.

Pathway b was also tested with univariable logistic regression with functional independence (ie, mRS, 0–2) as the true end point, to calculate the sensitivity and specificity of the NIHSS at 24 hours and 5 to 7 days for mRS 0 to 2 by performing Receiver Operating Characteristic-curve analyses. Pathway c and c′ are presented as (adjusted) common odds ratios ([a]cOR) with 95% CI, which are the pooled estimates of the effect on each cutoff point on the mRS. Pathway a was estimated as an (adjusted regression coefficient beta; aβ) and is expressed as percentage increase or decrease of the NIHSS score in the intervention compared with the control group, with 95% CI. Pathway b is presented as the acOR for every 10% increase in the NIHSS score, with 95% CI.

Regression analyses of pathway a and c were adjusted for age, baseline NIHSS, collateral score on computed tomography angiography, and onset-to-randomization time. Regression analysis of pathway b and c′ were adjusted for age, sex, baseline NIHSS, prestroke mRS, diabetes mellitus, hypertension, ischemic stroke, atrial fibrillation, internal carotid artery terminus occlusion, collateral score on computed tomography angiography, IVT (in MR CLEAN only), and onset-to-randomization time.

Missing data of the confounders, the true end point, and the surrogate end points were replaced per trial by multiple imputation with regression based on relevant covariates and outcomes. We performed a sensitivity analysis of the causal mediation model, in which NIHSS scores of patients who had died before the time point of NIHSS assessment was reached were also imputed with multiple imputation with regression. Statistical analyses were performed with Stata/SE software version 15.1 (StataCorp, College Station, TX).

Results

All 500 patients in MR CLEAN and all 656 patients in the IMS III trial were included in this study. Baseline characteristics were similar in the intervention and control group for both trials (Table 1). Median age was 66 in MR CLEAN and 69 in the IMS III trial, and respectively, 58% and 52% of the patients were men.

In MR CLEAN, 12 patients who had died within 24 hours and 57 patients who had died within 5 to 7 days were assigned the maximum NIHSS score of 42. After this, missing NIHSS scores (8 at 24 hours and 18 at 5 to 7 days) were replaced by multiple imputation with regression. In the IMS III trial, the maximum NIHSS score of 42 was assigned to the 15 patients who died within 24 hours and 68 patients who died within 5 to 7 days. Missing NIHSS scores at 24 hours (n=31) and 5 to 7 days (n=29) were replaced by multiple imputation with regression.

Mediation Analysis

In present analysis of MR CLEAN, EVT was associated with a significant improvement of functional outcome, with an acOR of 1.68 (95% CI, 1.22–2.32; pathway c; Table 2). Patients treated with EVT had lower NIHSS scores than patients in the control group at 24 hours (aβ, −17% [95% CI, −26 to −6.4]) and at 5 to 7 days (aβ, −24% [95% CI, −36 to −11]; pathway a). This means that—for example—a patient in the control group with an NIHSS score of 15 at 24 hours or 5 to 7 days would have had an NIHSS score of 12 (15−15×0.17) at 24 hours or 11 (15−15×0.25) at 5 to 7 days after EVT. The NIHSS at 24 hours was correlated with the mRS (acOR, 0.79 [95% CI, 0.77–0.82]), as was the NIHSS at 5 to 7 days (acOR, 0.77 [95% CI, 0.75–0.80]; pathway b). The sensitivity and specificity of the NIHSS at 24 hours for mRS 0 to 2 was 85% at the optimal cutoff point (area under the Receiver Operating Characteristic-curve=0.91). For the NIHSS at 5 to 7 days, this was 88% (area under the Receiver Operating Characteristic curve=0.94; Figure I in the online-only Data Supplement). The effect of EVT on the mRS was not statistically significant after adjustment for NIHSS at 24 hours (acOR, 1.36 [95% CI, 0.97–1.91]) nor after adjustment for NIHSS at 5 to 7 days (acOR, 1.24 [95% CI, 0.87–1.79]; pathway c′).

In the IMS III trial, we did not find a significant treatment effect of EVT on the ordinal mRS (acOR, 1.30 [95% CI, 0.97–1.75]; pathway c), nor on the NIHSS at 24 hours (aβ

Table 1.   Baseline Characteristics of Patients in MR CLEAN and the IMS III Trial According to Treatment Allocation

	MR CLEAN		IMS III Trial	
	Intervention (N=233)	Control (N=267)	Intervention (N=434)	Control (N=222)
Age in years, median (IQR)	66 (55–76)	66 (56–76)	69 (58–76)	68 (57–77)
Men, n/N (%)	135/233 (58%)	157/267 (59%)	218/434 (50%)	122/222 (55%)
NIHSS, median (IQR)*	17 (14–21)	18 (14–22)	17 (13–20)	16 (13–21)
Systolic blood pressure in mm Hg, mean (SD)†	146 (26)	145 (24)	150 (28)	152 (29)
Intravenous thrombolysis, n/N (%)	203/233 (87%)	242/267 (91%)	434/434 (100%)	222/222 (100%)
Time from symptom onset to randomization in minutes, median (IQR)‡	204 (152–251)	196 (149–266)	144 (120–170)	140 (115–165)
Medical history				
Atrial fibrillation, n/N (%)	66/233 (28%)	69/267 (26%)	135/424 (32%)	62/219 (28%)
Diabetes mellitus, n/N (%)	34/233 (15%)	34/267 (13%)	94/432 (22%)	54/220 (25%)
Hypertension, n/N (%)	98/233 (42%)	129/267 (48%)	319/432 (74%)	171/220 (78%)
Previous ischemic stroke, n/N (%)	29/233 (12%)	25/267 (9.4%)	58/431 (13%)	32/222 (14%)
Prestroke modified Rankin Scale, n/N (%)				
0	190/233 (82%)	214/267 (80%)	379/434 (87%)	197/222 (89%)
1	21/233 (9.0%)	29/267 (11%)	35/434 (8.1%)	21/222 (9.5%)
2	12/233 (5.2%)	13/267 (4.9%)	19/434 (4.4%)	4/222 (1.8%)
≥3	10/233 (4.3%)	11/267 (4.1%)	1/434 (0.2%)	0/222 (0%)
Imaging				
ICA-T occlusion location, n/N (%)	59/233 (25%)	75/267 (28%)	39/410 (9.5%)	19/206 (9.2%)
ASPECTS on NCCT, median (IQR)§	9 (7–10)	9 (8–10)	8 (6–10)	8 (6–10)
Collateral score on CT angiography, n/N (%)‖				
0	9/231 (3.9%)	17/262 (6.5%)	12/134 (9.0%)	2/65 (3.1%)
1	72/231 (31%)	64/262 (24%)	26/134 (19%)	15/65 (23%)
2	88/231 (38%)	110/262 (42%)	33/134 (25%)	21/65 (32%)
3	62/231 (27%)	71/262 (27%)	63/134 (47%)	27/65 (42%)

ASPECTS indicates Alberta Stroke Program Early CT Score; CT, computed tomography; ICA-T, internal carotid artery terminus; IMS III, Interventional Management of Stroke III; IQR, interquartile range presented as the 25th and 75th percentile; MR CLEAN, Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands; NCCT, noncontrast computed tomography; and NIHSS, National Institutes of Health Stroke Scale.

*Missing data in the IMS III trial: 2.

†Missing data in the IMS III trial: 8.

‡Missing data in MR CLEAN: 2.

§The ASPECTS ranges from 0 to 10, with higher scores indicating fewer early ischemic changes; missing data in MR CLEAN: 4; missing data in the IMS III trial: 7.

‖The collateral score is a scale from 0 to 3, with 0 representing absent collateral flow and 3 good collateral flow.

−4.0% [95% CI, −17 to 11]) or 5 to 7 days (aβ −10% [95% CI, −25 to 7.4]; pathway a). The NIHSS at 24 hours and at 5 to 7 days were correlated with the mRS (both acOR, 0.79 [95% CI, 0.77–0.81]; pathway b). The sensitivities and specificities of the NIHSS at 24 hours and 5 to 7 days for mRS 0 to 2 were similar to those of the MR CLEAN (Figure I in the online-only Data Supplement). Because no significant treatment effect was found on pathway c, pathway c′ was not tested in these data.

Results of the sensitivity analysis, in which we imputed NIHSS scores for patients who had died instead of assigning them 42 points, were comparable to those of the main analysis (Table I in the online-only Data Supplement). Distributions of the NIHSS and log10-transformed NIHSS+1 in the main analysis and sensitivity analysis are given in Figures II through V in the online-only Data Supplement.

For pathways a and c of both trials, unadjusted results are provided in Table II in the online-only Data Supplement.

Discussion

We used a causal mediation model to assess the early NIHSS as a surrogate end point for the mRS at 3 months in a positive and a neutral RCT of EVT for ischemic stroke. We found the NIHSS to be a valid outcome measure for treatment effect that mediates the effect of EVT on the mRS.

Although this is the first study to formally evaluate the early NIHSS as a surrogate end point with a causal mediation

Table 2. Application of the Causal Mediation Model in MR CLEAN and the IMS III Trial*

	Pathway a Effect of EVT on NIHSS (aβ)		Pathway b Correlation Between NIHSS and mRS Adjusted for EVT (acOR†)		Pathway c Effect of EVT on mRS (acOR)	Pathway c′ Effect of EVT on mRS Adjusted for NIHSS (acOR)	
	NIHSS 24 h	NIHSS 5–7 days	NIHSS 24 h	NIHSS 5–7 days		NIHSS 24 h	NIHSS 5–7 days
MR CLEAN	−17% (−26 to −6.4)‡	−24% (−36 to −11)	0.79 (0.77 to 0.82)	0.77 (0.75 to 0.80)	1.68 (1.22 to 2.32)	1.36 (0.97 to 1.91)	1.24 (0.87 to 1.79)
IMS III trial	−4.0% (−17 to 11)	−10% (−25 to 7.4)	0.79 (0.77 to 0.81)	0.79 (0.77 to 0.81)	1.30 (0.97 to 1.75)	N/A	N/A

aβ indicates adjusted regression coefficient beta; acOR, adjusted common odds ratio; EVT, endovascular treatment; IMS III, Interventional Management of Stroke III; MR CLEAN, Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands; mRS, modified Rankin Scale; N/A, not applicable; and NIHSS, National Institutes of Health Stroke Scale.

*All acORs and aβs are reported with their 95% CIs. All pathways are adjusted for age, baseline NIHSS, collateral score, time from symptom onset to randomization. Pathway b and c′ are additionally adjusted for sex, prestroke mRS, diabetes mellitus, hypertension, ischemic stroke, atrial fibrillation, ICA-T occlusion, IVT (in MR CLEAN only). Pathway c′ was not tested in the IMS III trial since no significant treatment effect was found on pathway c.

†Expressed as an acOR for every 10% increase in the NIHSS.

‡Interpretation of aβ: a patient with an NIHSS score of 15 at 24 h after treatment with usual care alone, would have an NIHSS score of 12 at 24 h when treated with EVT + usual care (15−15×0.17).

model, our results are supported by previous findings. In a meta-analysis of 5 RCTs, the strong beneficial effect of EVT was shown on both the NIHSS at 24 hours (pathway a) and the mRS at 3 months (pathway c).[1] In 3 other RCTs of EVT that assessed the mRS at 3 months and the NIHSS at 24 hours or 7 days, treatment effect of EVT was similar on both outcome measures (ie, either both positive[19] or both neutral[20,21]). Moreover, the predictive value of the NIHSS within 1 week after ischemic stroke for the mRS at 3 months (pathway b) has been demonstrated before.[10–12] These previous findings provide reliable evidence of the validity of the NIHSS as a surrogate end point for functional outcome.[22] The high sensitivities, specificities, and corresponding areas under the Receiver Operating Characteristic-curves of the NIHSS predicting functional independence at 3 months in our study substantiate this as well.

Regarding the optimal timing of the NIHSS assessment, it has previously been pointed out that the 7-day relative neurological improvement on NIHSS can predict 90-day functional outcome after EVT better than the 24-hour relative neurological improvement.[23] The NIHSS at 24 hours might miss important evolution of ischemic damage or early complications. Although the NIHSS at 5 to 7 days mediated the effect of EVT on the mRS most, the importance of the NIHSS at 24 hours should not be underestimated as it is less inflicted by loss of patients because of early death, which was also visible by the distributions of the NIHSS scores. NIHSS at 24 hours may also be more useful in practice, simply because it is assessed early after trial inclusion.

The selection of an appropriate primary outcome measure is a critical and challenging step in the design of a clinical trial that concerns investigators, regulatory authorities, professional organizations writing guidelines, and trial sponsors. It depends on the disease and the expected mechanism and effect of the treatment under study. We studied the NIHSS as an outcome measure in 2 RCTs of EVT for ischemic stroke. Validation of surrogate end points is treatment specific. However, we think that our findings can be generalized to ischemic stroke trials investigating an acute treatments with an early expected effect, that is, to treatments in which the NIHSS lies on the causal pathway between the treatment and the mRS.[13,16] This is supported by the fact that in all previous RCTs of IVT or intraarterial (recombinant) tissue-type

plasminogen activator that assessed both the mRS at 3 months and NIHSS at 24 hours, treatment effect was similar on both outcome measures (ie, either both positive or both neutral).[24–28]

A primary outcome measure should ideally be simple; easy and quick to assess; reliable; valid; responsive to meaningful change; evaluating impairments, disabilities, handicaps, or quality of life; and be free of bias.[29] Although the mRS is widely considered to be the standard primary outcome measure in trials of acute treatment for ischemic stroke, it does not meet all these criteria. Apart from the practical limitations that are caused by the long time span between treatment and outcome assessment, the mRS also has other limitations. First, the mRS lacks specificity because it includes all kinds of disability— including disability not related to the stroke or treatment. This might be the case for adverse events that occur between day 7 and day 90 after stroke onset or for disability that existed before the stroke, which is influenced by patient comorbidity, socioeconomic factors, and cognitive abilities.[3] Second, the use of the mRS may influence trial results due to the moderate interobserver reliability.[30] Third, although a single-point change on the mRS can often be deemed clinically relevant, compared with other stroke-outcome measures with more items, such as the NIHSS, the mRS may be less responsive to change because of its limited number of categories.[6,7] Over time, several alternative early primary or surrogate outcome measures have been proposed, including the NIHSS at 24 hours or 5 to 7 days, mRS at 1 week, follow-up infarct volume at 1 week on computed tomography or magnetic resonance imaging, and reperfusion directly after treatment on digital subtraction angiography.[3,31–34] Follow-up infarct volume has been formally evaluated as a mediator of the mRS but did not meet the expectations of an early surrogate end point.[31,33]

Next to the advantages of the early NIHSS (ie, assessed during hospital stay, reliable, easy and quick to assess, valid measure of stroke severity, responsive to meaningful change),[6,7,9] using the NIHSS as a primary outcome measure in (randomized) clinical trials has some practical disadvantages as well. First, the NIHSS does not include death. Excluding deceased patients would bias treatment effect estimates substantially when the mortality is not equally distributed between treatment arms, which is likely when a treatment is effective. Therefore, we assigned deceased patients the maximum score of 42 and performed a sensitivity analyses in which we imputed NIHSS

scores of deceased patients. Although imputing deceased patients might statistically lead to better distributions of the (log10-transformed) NIHSS at 5 to 7 days, results of the mediation model were comparable, and intuitively stroke physicians and investigators may not consider it appropriate to impute outcomes of deceased patients. Including deceased patients in the NIHSS score, regardless of the approach that was used, leads to a non-normal distribution of the NIHSS, requiring a transformation. Whether our approach of the log10(NIHSS+1)-transformation is best for analyzing the NIHSS as a measure of treatment effect, while also taking into consideration how to easily interpret the outcome, needs further evaluation. The interpretation of the log10-transformed NIHSS+1 may sound challenging but comes down to expressing treatment effect as percentage increase or decrease of NIHSS scores in the intervention versus control group. A specific limitation of our study was that NIHSS scores were assessed in a nonblinded manner, which may have led to overestimation of the treatment effect on the NIHSS in MR CLEAN. Nevertheless, consistency in our results—specifically that no treatment effect was observed on the NIHSS in the IMS III trial—suggests that nonblinded assessment of the NIHSS did not substantially overestimate the observed treatment effect on the NIHSS in MR CLEAN. If the NIHSS is used as a surrogate end point in trials, blinded NIHSS scores could, for example, be obtained by video assessment.

For selection of the most appropriate primary outcome measure, it is also important to take the phase of the trial into consideration. The early NIHSS could be a very useful primary outcome measure in phase II(b) trials testing the effect of new therapeutic agents or interventions for ischemic stroke, as was also proposed in two previous simulation studies using RCT data of IVT.[32,34] Because of its assessment during hospital stay, using the NIHSS as a primary outcome measure may not only lead to reduced trial duration and costs, but as the NIHSS is a more direct measure of the effect of EVT (ie, restoring blood flow to ischemic brain tissue), it could be valuable in phase II(b) trials for a first assessment of the effect of new therapeutic agents or interventions, and for guiding decisions about whether this new treatment should be evaluated in a (larger) phase III trial.

Whether the NIHSS may also be useful in (confirmatory) phase III trials is a challenging question. In our study, we have proven the NIHSS to be a valid surrogate end point for the mRS. One might argue that based on our findings and on previous research, there is plenty of evidence that it could be used as a surrogate end point in phase III trials. This merits consideration, especially for confirmatory phase III trials, which test modifications of treatments that have already been proven to be effective and safe, such as the comparison of various (new) types of mechanical thrombectomy devices or fibrinolytic agents. In general, relying too quickly on surrogate end points as the primary source of efficacy information of new treatments may result in limited insights regarding efficacy, as well as less reliable estimates of safety and side effects.[22]

A more fundamental question in this debate is whether the early NIHSS is able to measure what we want to achieve with a specific treatment. Fleming et al[22] stated that when selecting the primary end point in phase III trials, the effects on such an end point should provide reliable evidence about whether a new treatment provides clinically meaningful benefit (ie, the primary outcome measure should be [1] "a clinical event relevant to the patient," or [2] an "end point that measures directly how a patient feels, functions or survives"). There is increasing interest in more patient-oriented outcomes such as Patient-Reported Outcome Measures, as it is believed that these matter most to patients. In stroke, the utility-weighted mRS was advocated as a more patient-oriented outcome.[35] In contrast, the NIHSS provides limited information on how a stroke affects patients in their daily lives (eg, on a functional and/or emotional level). An NIHSS score of 1 is generally considered to represent an excellent outcome, but even a partial hemianopia or moderate aphasia could have severe consequences for the patient's quality of life.[6] However, although the mRS better represents the impact of a treatment on a patient's daily life, the NIHSS provides a more direct measure of treatment effect and is more responsive to change than the mRS with its limited categories. Therefore, clinically relevant effects of new treatments may be captured more easily with the NIHSS. Improvement on one of the neurological functions measured with NIHSS will be meaningful to most, if not all, patients. This could particularly be considered relevant in phase II(b) and confirmatory phase III trials. Taken together, the NIHSS and mRS do not measure the exact same thing, but based on the statement of Fleming et al,[22] we can conclude that both the effects on the NIHSS and mRS could provide evidence whether a new treatment provides clinically meaningful benefit.

Even aside from considerations such as whether we owe it to our patients to fully establish improvement of functional outcome in the long run, we have to consider that effects of new treatments may be influenced by delayed effects of complications, such as pneumonia and deep venous thrombosis. Early outcome measures might miss that. Especially when effects in the weeks after start of the treatment are expected, the mRS may be the better choice.

All in all, selecting the most appropriate outcome measure for trials of acute treatment for ischemic stroke poses a major challenge for investigators, regulatory authorities, professional organizations writing guidelines, and trial sponsors, which also has important clinical implications for patients. Thus, before being able to recommend the use of the NIHSS in phase III trials, a debate, or maybe even a consensus agreement, in the field with all parties is required.

Conclusions

The NIHSS within 1 week after EVT fulfills the requirements for a surrogate end point. It may be used as a primary outcome measure in phase II(b) trials of acute treatment for ischemic stroke. This could reduce stroke-outcome assessment to its essentials and also reduce trial duration and costs. A debate in the field is required to determine whether and under which conditions the NIHSS could be used as a primary outcome measure in (confirmatory) phase III trials.

Appendix

MR CLEAN Investigators: Olvert A. Berkhemer, MD, Department of Radiology, Amsterdam UMC, Location AMC, University of Amsterdam, the Netherlands and Department of Neurology, Erasmus MC University Medical Center Rotterdam, the Netherlands;

Puck S.S. Fransen, MD, Department of Neurology, Erasmus MC University Medical Center Rotterdam, the Netherlands and Department of Radiology, Erasmus MC University Medical Center Rotterdam, the Netherlands; Debbie Beumer, MD, Department of Neurology, Erasmus MC University Medical Center Rotterdam, the Netherlands and Department of Neurology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM), the Netherlands; Lucie A. van den Berg, MD, Department of Neurology, Amsterdam UMC, Location AMC, University of Amsterdam, the Netherlands; Hester F. Lingsma, PhD, Department of Public Health, Erasmus MC University Medical Center Rotterdam, the Netherlands; Albert J. Yoo, MD, Department of Radiology, Massachusetts General Hospital, Boston, United States of America; Wouter J. Schonewille, MD, Department of Neurology, Sint Antonius Hospital, Nieuwegein, the Netherlands; Jan Albert Vos, MD, PhD, Department of Radiology, Sint Antonius Hospital, Nieuwegein, the Netherlands; Paul J. Nederkoorn, MD, PhD, Department of Neurology, Amsterdam UMC, Location AMC, University of Amsterdam, the Netherlands; Marieke J.H. Wermer, MD, PhD, Department of Neurology, Leiden University Medical Center, the Netherlands; Marianne A.A. van Walderveen, MD, PhD, Department of Radiology, Leiden University Medical Center, the Netherlands; Julie Staals, MD, PhD, Department of Neurology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM), the Netherlands; Jeannette Hofmeijer, MD, PhD, Department of Neurology, Rijnstate Hospital, Arnhem, the Netherlands; Jacques A. van Oostayen, MD, PhD, Department of Radiology, Rijnstate Hospital, Arnhem, the Netherlands; Geert J. Lycklama a Nijeholt, MD, PhD, Department of Radiology, MC Haaglanden, the Hague, the Netherlands; Jelis Boiten, MD, PhD, Department of Neurology, MC Haaglanden, the Hague, the Netherlands; Patrick A. Brouwer, MD, Department of Radiology, Erasmus MC University Medical Center Rotterdam, the Netherlands; Bart J. Emmer, MD, PhD, Department of Radiology, Erasmus MC University Medical Center Rotterdam, the Netherlands; Sebastiaan F. de Bruijn, MD, PhD, Department of Neurology, HAGA Hospital, the Hague, the Netherlands; Lukas C. van Dijk, MD, Department of Radiology, HAGA Hospital, the Hague, the Netherlands; L. Jaap Kappelle, MD, PhD, Department of Neurology, University Medical Center Utrecht, the Netherlands; Rob H. Lo, MD, Department of Radiology, University Medical Center Utrecht, the Netherlands; Ewoud J. van Dijk, MD, PhD, Department of Neurology, Radboud University Medical Center, Nijmegen, the Netherlands; Joost de Vries, MD, PhD, Department of Neurosurgery, Radboud University Medical Center, Nijmegen, the Netherlands; Paul L.M. de Kort, MD, PhD, Department of Neurology, Sint Elisabeth Hospital, Tilburg, the Netherlands; Willem Jan J. van Rooij, MD, PhD, Department of Radiology, Sint Elisabeth Hospital, Tilburg, the Netherlands; Jan S.P. van den Berg, MD, PhD, Department of Neurology, Isala Klinieken, Zwolle, the Netherlands; Boudewijn A.A.M. van Hasselt, MD, Department of Radiology, Isala Klinieken, Zwolle, the Netherlands; Leo A.M. Aerden, MD, PhD, Department of Neurology, Reinier de Graaf Gasthuis, Delft, the Netherlands; Rene J. Dallinga, MD, Department of Radiology, Reinier de Graaf Gasthuis, Delft, the Netherlands; Marieke C. Visser, MD, PhD, Department of Neurology, Amsterdam UMC, Location VUmc, University of Amsterdam, the Netherlands; Joseph C.J. Bot, MD, PhD, Department of Radiology, Amsterdam UMC, Location VUmc, University of Amsterdam, the Netherlands; Patrick C. Vroomen, MD, PhD, Department of Neurology, University Medical Center Groningen, the Netherlands; Omid Eshghi, MD, Department of Radiology, University Medical Center Groningen, the Netherlands; Tobien H.C.M.L. Schreuder, MD, Department of Neurology, Atrium Medical Center, Heerlen, the Netherlands; Roel J.J. Heijboer, MD, Department of Radiology, Atrium Medical Center, Heerlen, the Netherlands; Koos Keizer, MD, PhD, Department of Neurology, Catharina Hospital, Eindhoven, the Netherlands; Alexander V. Tielbeek, MD, PhD, Department of Radiology, Catharina Hospital, Eindhoven, the Netherlands; Heleen M. den Hertog, MD, PhD, Department of Neurology, Medical Spectrum Twente, Enschede, the Netherlands; Dick G. Gerrits, MD, Department of Neurology, Medical Spectrum Twente, Enschede, the Netherlands; Renske M. van den Berg-Vos, MD, PhD, Department of Neurology, Sint Lucas Andreas Hospital, Amsterdam, the Netherlands; Giorgos B. Karas, MD, Department of Radiology, Sint Lucas Andreas Hospital, Amsterdam, the Netherlands; Ewout W. Steyerberg, PhD, Department of Public Health, Erasmus MC University Medical Center Rotterdam, the Netherlands; H. Zwenneke Flach, MD, Department of Radiology, Isala Klinieken, Zwolle, the Netherlands; Henk A. Marquering PhD, Department of Radiology, Amsterdam UMC, Location AMC, University of Amsterdam, the Netherlands and Department of Biomedical Engineering and Physics, Academic Medical Center Amsterdam, the Netherlands; Marieke E.S. Sprengers, MD, PhD, Department of Radiology, Amsterdam UMC, Location AMC, University of Amsterdam, the Netherlands; Sjoerd F.M. Jenniskens, MD, PhD, Department of Radiology, Radboud University Medical Center, Nijmegen, the Netherlands; Ludo F.M. Beenen, MD, Department of Radiology, Amsterdam UMC, Location AMC, University of Amsterdam, the Netherlands; Rene van den Berg, MD, PhD, Department of Radiology, Amsterdam UMC, Location AMC, University of Amsterdam, the Netherlands; Peter J. Koudstaal, MD, PhD, Department of Neurology, Erasmus MC University Medical Center Rotterdam, the Netherlands; Wim H. van Zwam, MD, PhD, Department of Radiology, Maastricht University Medical Center, the Netherlands; Yvo B.W.E.M. Roos, MD, PhD, Department of Neurology, Amsterdam UMC, Location AMC, University of Amsterdam, the Netherlands Aad van der Lugt, MD, PhD, Department of Radiology, Erasmus MC University Medical Center Rotterdam, the Netherlands; Robert J. van Oostenbrugge, MD, PhD, Department of Neurology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM), the Netherlands; Charles B.L.M. Majoie, MD, PhD, Department of Radiology, Amsterdam UMC, Location AMC, University of Amsterdam, the Netherlands; and Diederik W.J. Dippel, MD, PhD, Department of Neurology, Erasmus MC University Medical Center Rotterdam, the Netherlands.

Disclosures

Dr Yoo reports grants from Cerenovus, Medtronic, Penumbra, Stryker, Genentech, personal fees from Penumbra personal fees from Cerenovus, and Genentech for core imaging laboratory activities and consultancy and has equity ownership in Insera Therapeutics outside the submitted work. Dr Broderick reports research monies

References

1. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al; HERMES collaborators. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet.* 2016;387:1723–1731. doi: 10.1016/S0140-6736(16)00163-X

2. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al; American Heart Association Stroke Council. 2018 guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the american Heart Association/American Stroke Association. *Stroke.* 2018;49:e46–e110. doi: 10.1161/STR.0000000000000158

3. Jovin TG, Albers GW, Liebeskind DS; STAIR IX Consortium. Stroke treatment academic industry roundtable: the next generation of endovascular trials. *Stroke.* 2016;47:2656–2665. doi: 10.1161/STROKEAHA.116.013578

4. Coster WJ. Making the best match: selecting outcome measures for clinical trials and outcome studies. *Am J Occup Ther.* 2013;67:162–170. doi: 10.5014/ajot.2013.006015

5. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke.* 1988;19:604–607. doi: 10.1161/01.str.19.5.604

6. Harrison JK, McArthur KS, Quinn TJ. Assessment scales in stroke: clinimetric and clinical considerations. *Clin Interv Aging.* 2013;8:201–211. doi: 10.2147/CIA.S32405

7. Kasner SE. Clinical interpretation and use of stroke scales. *Lancet Neurol.* 2006;5:603–612. doi: 10.1016/S1474-4422(06)70495-1

8. Dromerick AW, Edwards DF, Diringer MN. Sensitivity to changes in disability after stroke: a comparison of four scales useful in clinical trials. *J Rehabil Res Dev.* 2003;40:1–8. doi: 10.1682/jrrd.2003.01.0001

9. Brott T, Adams HP Jr, Olinger CP, Marler JR, Barsan WG, Biller J, et al. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke.* 1989;20:864–870. doi: 10.1161/01.str.20.7.864

10. Rangaraju S, Frankel M, Jovin TG. Prognostic value of the 24-hour neurological examination in anterior circulation ischemic stroke: a post hoc analysis of two randomized controlled stroke trials. *Interv Neurol.* 2016;4:120–129. doi: 10.1159/000443801

11. Sajobi TT, Menon BK, Wang M, Lawal O, Shuaib A, Williams D, et al; ESCAPE Trial Investigators. Early trajectory of stroke severity predicts long-term functional outcomes in ischemic stroke subjects: results from the ESCAPE Trial (Endovascular Treatment for Small Core and Anterior Circulation Proximal Occlusion With Emphasis on Minimizing CT to Recanalization Times). *Stroke.* 2017;48:105–110. doi: 10.1161/STROKEAHA.116.014456

12. Saver JL, Altman H. Relationship between neurologic deficit severity and final functional outcome shifts and strengthens during first hours after onset. *Stroke.* 2012;43:1537–1541. doi: 10.1161/STROKEAHA.111.636928

13. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996;125:605–613. doi: 10.7326/0003-4819-125-7-199610010-00011

14. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al; MR CLEAN Investigators. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* 2015;372:11–20. doi: 10.1056/NEJMoa1411587

15. Broderick JP, Palesch YY, Demchuk AM, Yeatts SD, Khatri P, Hill MD, et al; Interventional Management of Stroke (IMS) III Investigators. Endovascular therapy after intravenous t-PA versus t-PA alone for stroke. *N Engl J Med.* 2013;368:893–903. doi: 10.1056/NEJMoa1214300

16. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med.* 1989;8:431–440. doi: 10.1002/sim.4780080407

17. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986;51:1173–1182. doi: 10.1037//0022-3514.51.6.1173

18. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods.* 2013;18:137–150. doi: 10.1037/a0031034

19. Bracard S, Ducrocq X, Mas JL, Soudant M, Oppenheim C, Moulin T, et al; THRACE investigators. Mechanical thrombectomy after intravenous alteplase versus alteplase alone after stroke (THRACE): a randomised controlled trial. *Lancet Neurol.* 2016;15:1138–1147. doi: 10.1016/S1474-4422(16)30177-6

20. Ciccone A, Valvassori L; SYNTHESIS Expansion Investigators. Endovascular treatment for acute ischemic stroke. *N Engl J Med.* 2013;368:2433–2434. doi: 10.1056/NEJMc1304759

21. Muir KW, Ford GA, Messow CM, Ford I, Murray A, Clifton A, et al; PISTE Investigators. Endovascular therapy for acute ischaemic stroke: the Pragmatic Ischaemic Stroke Thrombectomy Evaluation (PISTE) randomised, controlled trial. *J Neurol Neurosurg Psychiatry.* 2017;88:38–44. doi: 10.1136/jnnp-2016-314117

22. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med.* 2012;31:2973–2984. doi: 10.1002/sim.5403

23. Pu J, Wang H, Tu M, Zi W, Hao Y, Yang D, et al. Combination of 24-hour and 7-day relative neurological improvement strongly predicts 90-day functional outcome of endovascular stroke therapy. *J Stroke Cerebrovasc Dis.* 2018;27:1217–1225. doi: 10.1016/j.jstrokecerebrovasdis.2017.11.042

24. IMS Study Investigators. Combined intravenous and intra-arterial recanalization for acute ischemic stroke: The interventional management of stroke study. *Stroke.* 2004;35:904–911

25. IMS II Trial Investigators. The Interventional Management of Stroke (IMS) II study. *Stroke.* 2007;38:2127–2135

26. Ogawa A, Mori E, Minematsu K, Taki W, Takahashi A, Nemoto S, et al; MELT Japan Study Group. Randomized trial of intraarterial infusion of urokinase within 6 hours of middle cerebral artery stroke: the middle cerebral artery embolism local fibrinolytic intervention trial (MELT) Japan. *Stroke.* 2007;38:2633–2639. doi: 10.1161/STROKEAHA.107.488551

27. Clark WM, Wissman S, Albers GW, Jhamandas JH, Madden KP, Hamilton S. Recombinant tissue-type plasminogen activator (Alteplase) for ischemic stroke 3 to 5 hours after symptom onset. The ATLANTIS Study: a randomized controlled trial. Alteplase Thrombolysis for Acute Noninterventional Therapy in Ischemic Stroke. *JAMA.* 1999;282:2019–2026. doi: 10.1001/jama.282.21.2019

28. National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med.* 1995;333:1581–1587

29. Lyden PD, Hantson L. Assessment scales for the evaluation of stroke patients. *J Stroke Cerebrovasc Dis.* 1998;7:113–127.

30. Quinn TJ, Dawson J, Walters MR, Lees KR. Reliability of the modified Rankin Scale: a systematic review. *Stroke.* 2009;40:3393–3395. doi: 10.1161/STROKEAHA.109.557256

31. Boers AMM, Jansen IGH, Brown S, Lingsma HF, Beenen LFM, Devlin TG, et al. Mediation of the relationship between endovascular therapy and functional outcome by follow-up infarct volume in patients with acute ischemic stroke. *JAMA Neurol.* 2019;76:194–202. doi: 10.1001/jamaneurol.2018.3661

32. Broderick JP, Lu M, Kothari R, Levine SR, Lyden PD, Haley EC, et al. Finding the most powerful measures of the effectiveness of tissue plasminogen activator in the NINDS tPA stroke trial. *Stroke.* 2000;31:2335–2341. doi: 10.1161/01.str.31.10.2335

33. Compagne KCJ, Boers AMM, Marquering HA, Berkhemer OA, Yoo AJ, Beenen LFM, et al; MR CLEAN Investigators. Follow-up infarct volume as a mediator of endovascular treatment effect on functional outcome in ischaemic stroke. *Eur Radiol.* 2019;29:736–744. doi: 10.1007/s00330-018-5578-9

34. Kerr DM, Fulton RL, Lees KR; VISTA Collaborators. Seven-day NIHSS is a sensitive outcome measure for exploratory clinical trials in acute stroke: evidence from the Virtual International Stroke Trials Archive. *Stroke.* 2012;43:1401–1403. doi: 10.1161/STROKEAHA.111.644484

35. Chaisinanunkul N, Adeoye O, Lewis RJ, Grotta JC, Broderick J, Jovin TG, et al; DAWN Trial and MOST Trial Steering Committees; Additional contributors from DAWN Trial Steering Committee. Adopting a patient-centered approach to primary outcome analysis of acute stroke trials using a utility-weighted modified rankin scale. *Stroke.* 2015;46:2238–2243. doi: 10.1161/STROKEAHA.114.008547