Utility-Weighted Modified Rankin Scale as Primary Outcome in Stroke Trials
A Simulation Study

Simone A. Dijkland, MD; Daphne C. Voormolen, MSc; Esmee Venema, MD;
Bob Roozenbeek, MD, PhD; Suzanne Polinder, PhD; Juanita A. Haagsma, PhD;
Daan Nieboer, MSc; Vicky Chalos, MD; Albert J. Yoo, MD; Jennifer Schreuders, MD;
Aad van der Lugt, MD, PhD; Charles B.L.M. Majoie, MD, PhD; Yvo B.W.E.M. Roos, MD, PhD;
Wim H. van Zwam, MD, PhD; Robert J. van Oostenbrugge, MD, PhD;
Ewout W. Steyerberg, PhD; Diederik W.J. Dippel, MD, PhD; Hester F. Lingsma, PhD;
on behalf of the MR CLEAN Investigators*

Background and Purpose—The utility-weighted modified Rankin Scale (UW-mRS) has been proposed as a new patient-centered primary outcome in stroke trials. We aimed to describe utility weights for the mRS health states and to evaluate the statistical efficiency of the UW-mRS to detect treatment effects in stroke intervention trials.

Methods—We used data of the 500 patients enrolled in the MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands). Utility values were elicited from the EuroQol Group 5-Dimension Self-Report Questionnaire assessed at 90 days after inclusion, simultaneously with the mRS. Utility weights were determined by averaging the utilities of all patients within each mRS category. We performed simulations to evaluate statistical efficiency. The simulated treatment effect was an odds ratio of 1.65 in favor of the treatment arm, similar for all mRS cutoffs. This treatment effect was analyzed using 3 approaches: linear regression with the UW-mRS as outcome, binary logistic regression with a dichotomized mRS (0–1/2–6, 0–2/3–6, and 0–4/5–6), and proportional odds logistic regression with the ordinal mRS. The statistical power of the 3 approaches was expressed as the proportion of 10 000 simulations that resulted in a statistically significant treatment effect ($P \le 0.05$).

Results—The mean utility values (SD) for mRS categories 0 to 6 were: 0.95 (0.08), 0.93 (0.13), 0.83 (0.21), 0.62 (0.27), 0.42 (0.28), 0.11 (0.28), and 0 (0), respectively, but varied substantially between individual patients within each category. The UW-mRS approach was more efficient than the dichotomous approach (power 85% versus 71%) but less efficient than the ordinal approach (power 85% versus 87%).

Conclusions—The UW-mRS as primary outcome does not capture individual variation in utility values and may reduce the statistical power of a randomized trial.   (***Stroke***. 2018;49:965-971. DOI: 10.1161/STROKEAHA.117.020194.)

Key Words: brain ischemia ■ odds ratio ■ quality of life ■ self-report ■ stroke

The modified Rankin Scale (mRS) is the most widely used primary outcome measure in trials for acute stroke interventions.[1,2] The mRS is an ordinal scale ranging from 0 (no symptoms) to 6 (death) measuring the degree of disability or dependence in everyday life.[3] Previously, dichotomizing the mRS into dead or dependent (mRS, 3–6) versus independent (mRS, 0–2) was common, but this results in a reduction in statistical power to detect relevant treatment effects.[4] Therefore,

statistical approaches preserving the ordinal nature of outcome measures, such as proportional odds logistic regression, have been recommended for stroke and other neurological disorders.[1,5–8]

Currently, the importance of incorporating quality of life (QoL) in outcome analysis in stroke trials is increasingly recognized.[9–11] For the mRS to reflect both treatment effect and patient perception, the utility-weighted mRS (UW-mRS) has been proposed and used as primary end point.[2,12,13] In the UW-mRS, utilities based on the EuroQol Group 5-Dimension Self-Report Questionnaire (EQ-5D-3L) values are assigned to the mRS health states. Two prior studies reported utility weights for the mRS health states: 1 representing the values of patients and 1 representing the values of clinicians. The utility weights that were proposed for the UW-mRS are based on these 2 studies.[12] Compared with the ordinal mRS, the UW-mRS showed similar statistical power to detect treatment effects in empirical data in a wide range of stroke trials.[12] However, because in empirical data, the true treatment effect is unknown, the only valid method to assess statistical power is simulation.

We aimed to describe utility weights for the mRS health states and to evaluate the statistical efficiency of the UW-mRS to detect treatment effects in stroke trials.

Methods

Anonymized trial data and analytic methods that support our study findings are available from the principal investigator (e-mail: mrclean@erasmusmc.nl) on reasonable request.

Study Population

We used individual patient data of the 500 patients enrolled in the MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands). MR CLEAN was a phase III, multicenter randomized clinical trial, designed to evaluate whether intra-arterial treatment (within 6 hours of symptom onset) plus usual care would be more effective than usual care alone in patients with acute ischemic stroke and a proximal arterial occlusion in the anterior cerebral circulation. The primary outcome was the mRS at 90 days, and the secondary outcome was the EQ-5D-3L at 90 days. In MR CLEAN, ethics approval was obtained from the local institutional review boards of the participating centers, and written informed consent was obtained from patients or legal representatives before randomization.[14]

Modified Rankin Scale

The mRS is a measure of functional outcome after stroke, evaluating the degree of disability or dependence in daily life. The scale is derived from clinical assessment by a trained nurse or a physician and consists of 7 grades ranging from 0 (no symptoms) to 6, with 5 indicating severe disability and 6 indicating death. A score of ≤2 indicates functional independence.[3]

Utilities

Utilities represent preferences for mRS health states and range from 0 (death) to 1 (perfect health). Utility values of poor outcome categories might even be negative, indicating that they are valued worse than death.[15] In MR CLEAN, utility values were elicited using the EQ-5D-3L responses of patient, proxy, or healthcare provider assessed at 90 days after inclusion, simultaneously with the mRS. The EQ-5D-3L consists of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with 3 levels each (no problems, some problems, and extreme problems), thus defining

243 (3^5) distinct health states.[16] Converting the EQ-5D-3L responses into utility values was done according to the Dutch tariff—a country-specific value set established based on the time trade-off method.[17] Patients who died before the follow-up interviews at 90 days received a utility value of zero. The utility values ranged from −0.33 to 1.00. We determined utility weights for each mRS category by averaging the derived utilities (including the negative values) of all patients within each mRS health state (eg, the utility weight for mRS=1 is the average of the utilities of all patients with mRS=1). Additionally, we matched the utility values proposed by Chaisinanunkul et al,[12] who collapsed mRS 5 to 6 by assigning a utility weight of zero to both categories, to our mRS values.

Simulations for Statistical Efficiency

Statistical efficiency was evaluated based on simulations that utilized the MR CLEAN database. For a single simulation, 500 patients were sampled at random with replacement. For each patient, the predicted probability of each possible outcome on the 7-point ordinal mRS was modeled as a function of the baseline covariates. These covariates were identical to those in MR CLEAN and included age, stroke severity (National Institutes of Health Stroke Scale) at baseline, time from stroke onset to randomization, status with respect to previous stroke, atrial fibrillation, diabetes mellitus, and occlusion of the internal carotid artery terminus (yes/no).[14]

Using these estimated probabilities, an actual outcome in terms of an mRS or UW-mRS was simulated. Treatment (yes/no) was randomly assigned, and the simulated treatment effect was an odds ratio (OR) of 1.65 (β=0.5) in favor of the treatment arm, similar for all mRS cutoffs. We also evaluated a scenario with no treatment effect, by simulating a treatment effect of OR=1.0 (β=0). During this process, samples of 500 subjects were generated representing 250 patients from the control group and 250 from the intervention group, with a known treatment effect. This was then repeated 10 000×.

The data were analyzed by 3 different statistical approaches. First, we dichotomized the 90-day mRS in 3 different ways of favorable versus unfavorable outcome: 0 to 1 versus 2 to 6, 0 to 2 versus 3 to 6, and 0 to 4 versus 5 to 6. The treatment effect on the dichotomized mRS was determined using binary logistic regression. Second, we used proportional odds logistic regression for analysis of the treatment effect on the ordinal mRS. We fitted a proportional odds logistic regression model with the 7-point ordinal mRS scale as outcome. The proportional odds model estimates a common OR over all health state transitions within the mRS. According to the proportional odds assumption, the common OR is an accurate reflection of the overall treatment effect if the ORs are the same for each health state transition. If there is agreement regarding the ordinality of the mRS, the common OR can be interpreted as a summary measure of treatment effect even if the proportional odds assumption is violated.[18] Third, treatment effect on the UW-mRS was analyzed using linear regression, as proposed by Chaisinanunkul et al.[12]

Each of the 3 approaches yielded either a significant ($P \leq 0.05$) or a nonsignificant treatment effect ($P > 0.05$, 2 sided). The power (or type 1 error in case of no treatment effect) of each statistical approach was estimated as the proportion of the 10 000 analyses, which resulted in a statistically significant treatment effect.

Associations were expressed as ORs or β with 95% confidence intervals (CIs), averaged over all simulations. All analyses were performed unadjusted and adjusted for the prespecified covariates identical to those mentioned above. Statistical analyses were performed with R software, version 3.3.2 (R Foundation for Statistical Computing, Vienna, Austria). Missing data on time from stroke to randomization (0.4%) and level of vessel occlusion (0.2%) was statistically imputed using simple imputation (replacement by mean or mode, as applicable).

Results

Study Population

All 500 participants from the MR CLEAN trial were included in our analysis. The mRS at 90 days was available for all

patients. The EQ-5D-3L assessments, and consequently the utility values, were available in 457 patients (including 108 patients who died before follow-up). In 43 patients (8.6%), mRS assessment could not be followed by an EQ-5D-3L assessment. In 192 patients (38%), the EQ-5D-3L was completed by a proxy.

The total study population had a mean age of 65 years (SD, 14 years), and most patients (58%) were men (Table 1). The intervention and control groups were similar in terms of baseline and treatment characteristics. The number of patients with poor outcome (mRS, 3–6) at 90 days was lower in the intervention group than in the control group (Figure 1).

Utility Weights

The mean utility values (SD) for mRS categories 0 to 6 were: 0.95 (0.08), 0.93 (0.13), 0.83 (0.21), 0.62 (0.27), 0.42 (0.28), 0.11 (0.28), and 0 (0), respectively (Table 2). We observed substantial variation in utility values within each mRS category (Figure 2). Within MR CLEAN, the mean UW-mRS for the intervention group was significantly higher when compared with the control group (Table 2).

Table 1.   Baseline Characteristics of the 500 Patients in the MR CLEAN Trial

Baseline Variable	Intervention (n=233) Intra-Arterial Treatment Plus Usual Care	Control (n=267) Usual Care
Age, y; median (IQR)	65.8 (54.5–76.0)	65.7 (55.5–76.4)
Male sex	135 (58%)	157 (59%)
NIHSS score, median (IQR)	17 (14–21)	18 (14–22)
Previous ischemic stroke	29 (12%)	25 (9%)
Atrial fibrillation	66 (28%)	69 (26%)
Diabetes mellitus	34 (15%)	34 (13%)
Prestroke mRS		
0	190 (82%)	214 (80%)
1	21 (9%)	29 (11%)
2	12 (5%)	13 (5%)
>2	10 (4%)	11 (4%)
Treatment with IV alteplase	203 (87%)	242 (91%)
Time from stroke onset to start of IV alteplase, min; median (IQR)	85 (67–110)	87 (65–116)
Occlusion of the internal carotid artery terminus*	59 (25%)	75 (28%)
Time from stroke onset to randomization, min; median (IQR)†	204 (152–251)	196 (149–266)
Time from stroke onset to groin puncture, min; median (IQR)	260 (210–313)	NA

IQR indicates interquartile range; IV, intravenous; mRS, modified Rankin Scale; NA, not applicable; and NIHSS, National Institutes of Health Stroke Scale.

*No vessel imaging in 1 patient in the control group.

†Data were missing for 2 patients in the intervention group.

Outcome Analysis in MR CLEAN

Ordinal analysis of the mRS showed improved functional outcomes in favor of the intervention, consistent throughout all categories of the mRS except for death (adjusted common OR, 1.67; 95% CI, 1.21–2.30; Figure 1). The dichotomous approach led to slightly stronger treatment effects for cutoffs mRS 0 to 1 and 0 to 2 (adjusted OR, 2.07 [95% CI, 1.07–4.02] and 2.16 [95% CI, 1.39–3.38], respectively). The fact that the ORs were not equal for the different cutoffs might imply that the proportional odds assumption did not hold perfectly in the empirical data. Linear analysis of the UW-mRS resulted in an adjusted β of 0.086 (95% CI, 0.033–0.131).

Simulations

For all 3 prespecified mRS dichotomizations, intra-arterial treatment was positively associated with better outcomes (adjusted OR, 1.66–1.68; Table 3). The estimated treatment effects were similar to the simulated (true) treatment effect of 1.65. When comparing the 3 different mRS cutoffs, the statistical efficiency for the cutoff of mRS 0 to 2 versus 3 to 6 was highest (power 71% versus 62% for mRS 0–1 and 35% for mRS 0–4). This could be explained by an almost equal distribution of patients among both categories for this cutoff (Table 3).

Ordinal analysis of the mRS estimated an adjusted treatment effect of common OR=1.66 (95% CI, 1.41–1.95; Table 3), similar to the dichotomous approach. However, the ordinal approach was statistically more efficient (power 87% versus 71%).

Linear regression analysis of the UW-mRS estimated an adjusted beneficial treatment effect of β=0.075 (95% CI, 0.027–0.125; Table 3). The UW-mRS approach was statistically less efficient in detecting treatment effects compared with the ordinal approach (power 85% versus 87%). Matching the utilities of Chaisinanunkul et al to the mRS values in MR CLEAN led to similar results (Tables 2 and 3). However, the assumptions of the linear model were not met because there was non-normality of the residuals (Figure I in the online-only Data Supplement).

In the simulations without a treatment effect, a proportion of false-positives (type 1 error) of around 5% was estimated for all 3 statistical approaches (data not shown).

Discussion

We evaluated the UW-mRS—a recently proposed patient-centered outcome measure in stroke. Our study, based on a Dutch stroke intervention trial, showed that the UW-mRS does not capture the individual variation in utility values within each mRS category. Moreover, our simulations revealed that the UW-mRS approach was more efficient in detecting treatment effects than dichotomous analysis of the mRS but less efficient than the ordinal approach.

Widely used functional outcome measures in stroke intervention trials, such as the mRS, have been extensively studied concerning their feasibility in measuring disability after stroke.[19,20] Nevertheless, more attention has recently been aimed at incorporating patient-reported QoL in stroke outcome measures.[10,11]
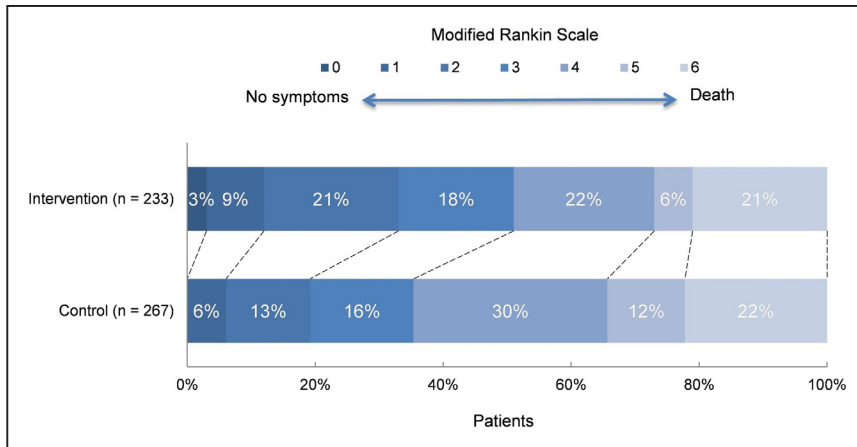
Figure 1. Distribution of the modified Rankin Scale at 90 days among intervention and control groups.

As part of this trend, the UW-mRS has been proposed as a new primary patient-centered outcome measure in acute stroke intervention trials. In empirical data, the UW-mRS was equally statistically efficient in detecting treatment effects compared with ordinal analysis of the mRS.[12] Based on that study, the UW-mRS was recently used as the primary end point in the DAWN trial (Diffusion-Weighted Imaging or Computerized Tomography Perfusion Assessment With Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention With Trevo),[13] and it is expected that more trials will follow. However, the study by Chaisinanunkul et al was only based on analyses of empirical sets of data. Because the true treatment effect in empirical data is unknown and different treatment effects on different outcome measures could be caused by random variation, the only valid method to assess the power of a statistical approach is a simulation study, as we performed.

Table 2. Mean Utility Values per mRS Category and Mean UW-mRS in MR CLEAN and the Study by Chaisinanunkul et al

	No. of Patients MR CLEAN	Mean (SD)	Chaisinanunkul et al[12], Mean Utility Values
mRS			
0	7	0.95 (0.08)	1.00
1	36	0.93 (0.13)	0.91
2	84	0.83 (0.21)	0.76
3	87	0.62 (0.27)	0.65
4	133	0.42 (0.29)	0.33
5	45	0.11 (0.28)	0.00
6	108	0.00	0.00
UW-mRS			
Overall	500	0.45 (0.32)	0.40
Intervention group	233	0.50 (0.33)*	0.46
Control group	267	0.41 (0.31)	0.36

MR CLEAN indicates Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands; mRS, modified Rankin Scale; and UW, utility weighted.

*Mean utility for the intervention group vs control group within MR CLEAN: $P=0.002$ (Mann–Whitney U test).

Intuitively, patient-centered outcomes, such as the UW-mRS, are clinically useful because they concern patient-reported measures combined with the perception of the general public. These outcomes reflect patient perception and respect the nonequality of health state transitions on an ordinal scale. Nevertheless, averaging utility values for each mRS category does not reflect individual valuation of these health states: all patients within 1 mRS category receive the same utility weight, irrespective of their own valuation of this health state (Figure 2). So, the UW-mRS is in fact a revaluation of the mRS. Moreover, the utility distribution with mRS=5 being worse than death for some patients does not support collapsing mRS categories 5 to 6 as proposed by Chaisinanunkul et al. To reflect true individual valuation of health states, QoL instruments should rather be used as outcome. However, utility values derived from the EuroQol Group 5-Dimension Self-Report Questionnaire may not cover the full range of limitations relevant to patients with stroke[21] and may, therefore, overestimate QoL in this group. An alternative would be to use utility values derived from QoL instruments designed specifically for patients with neurological disorders, such as Neuro-QoL.[22] Nevertheless, because QoL depends on many external factors, it might introduce noise, making it less suitable as a primary outcome measure.[23,24]

Our simulations revealed that the UW-mRS is not as statistically efficient as ordinal analysis of the mRS and may, therefore, cause a reduction in statistical power when used in randomized trials. Chaisinanunkul et al[12] analyzed the UW-mRS with a t test, implying a continuous outcome variable. We used linear regression, which is a comparable approach but allows for multivariable analysis. In theory, linear analysis is expected to be more efficient than ordinal analysis when the assumptions of the linear model are met. A linear model assumes that the errors between observed and predicted values, that is, the residuals of the regression, are normally distributed. In our analyses, however, we found non-normality of the residuals of the linear model for the UW-mRS. Because the UW-mRS remains a scale with 7 outcome categories, the assumption of normally distributed residuals can never be met. Non-normality of the residuals might cause bias because of underestimation of the standard error. Therefore, the actual power of the UW-mRS approach will be even <85%. Ordinal analysis also makes an assumption (the
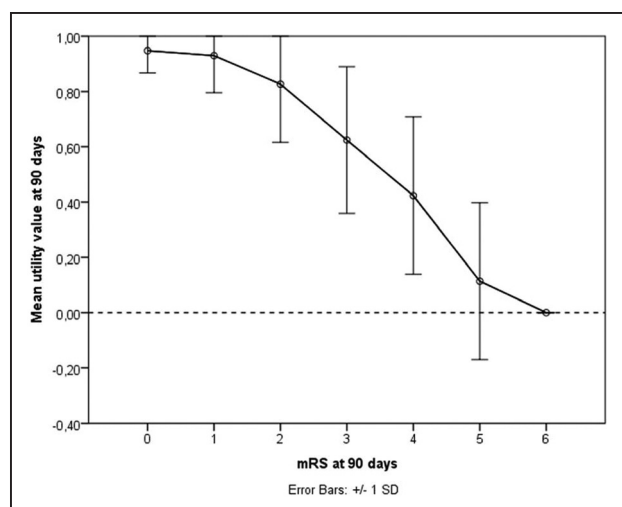
Figure 2. Mean EuroQol Group 5-Dimension Self-Report Questionnaire (EQ-5D-3L) utility values per modified Rankin Scale (mRS) category in MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands).

proportional odds assumption), but it should be noted that the assumption of a normal distribution of the residuals in a linear model is more difficult to fulfill than the assumption of ordinality in proportional odds analyses. In line with theoretical expectations, the UW-mRS showed to be exactly as efficient as the mRS when it was analyzed with a proportional odds model (data not shown).

Defining a beneficial treatment effect in terms of the UW-mRS, and, therefore, clinical interpretability, might be

difficult. Treatment effect on the UW-mRS scale is expressed as a difference in mean UW-mRS between treatment and control groups.[12] This difference can be converted into quality-adjusted life-years (QALYs) gained or lost by a certain treatment.[12,25] The QALY measure assumes that a year of life lived in perfect health is worth 1 QALY, and a year of life lived in a state less than perfect health is worth <1 QALY, proportional to its utility value (QALY=years of life×utility). QALYs can be used to calculate cost-effectiveness to select a certain intervention for funding.[26] Also, the QALY measure has been argued to be more intuitive to patients (healthy life-years gained) and, therefore, to improve communication of treatment effects.[12,25] However, when not converted into QALYs, treatment effects expressed as utility differences remain difficult to interpret. Moreover, clinicians and researchers are now used to working with the (common) OR.

Ordinal outcome scales are also used in other neurological disorders besides stroke. Examples are the Glasgow Outcome Scale in traumatic brain injury and the Guillain-Barre syndrome disability score in Guillain-Barre syndrome.[6,7,27] These ordinal outcomes could be transformed to patient-centered outcomes using utility values, similar to the UW-mRS. For randomized trials in patients with other neurological diseases, such as traumatic brain injury and Guillain-Barre syndrome, our study might, therefore, also implicate that ordinal analysis should remain the gold standard.

Our study has several strengths. The simulation study was based on data from the MR CLEAN trial, with relatively broad inclusion criteria.[14] As such, our findings should be generalizable to future stroke trials. Furthermore, simulation is the most adequate method to evaluate statistical power. Also, we used

Table 3.   Univariable and Multivariable Estimated Treatment Effects in Simulations (n=500)

		SE	Power		SE	Power
Binary logistic regression	Univariable OR (95% CI)*			Multivariable OR (95% CI)*†		
0–1 vs 2–6		0.205	56%		0.226	62%
0–1 (n=146)	1.54 (1.29–1.83)			1.67 (1.08–2.61)		
2–6 (n=354)	Reference			Reference		
0–2 vs 3–6		0.181	63%		0.203	71%
0–2 (n=272)	1.51 (1.30–1.74)			1.66 (1.12–2.48)		
3–6 (n=228)	Reference			Reference		
0–4 vs 5–6		0.303	32%		0.326	35%
0–4 (n=448)	1.58 (1.21–2.07)			1.68 (0.89–3.19)		
5–6 (n=52)	Reference			Reference		
Proportional odds logistic regression	Univariable OR (95% CI)*			Multivariable OR (95% CI)*†		
mRS at 90 d	1.53 (1.34–1.75)	0.159	76%	1.66 (1.41–1.95)	0.163	87%
Linear regression	Univariable β (95% CI)*			Multivariable β (95% CI)*†		
UW-mRS with MR CLEAN utilities	0.075 (0.020–0.131)	0.028	76%	0.075 (0.027–0.125)	0.025	85%
UW-mRS with utilities from Chaisinanunkul et al	0.076 (0.020–0.133)	0.029	75%	0.077 (0.026–0.128)	0.026	84%

CI indicates confidence interval; MR CLEAN, Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale; OR, odds ratio; and UW, utility weighted.

*Simulated treatment effect β=0.5 (OR, 1.65).

†Adjusted for age, NIHSS at baseline, time from stroke onset to randomization, status with respect to previous stroke, atrial fibrillation, diabetes mellitus, and occlusion of the internal carotid artery terminus (yes/no).

utility values derived using the recommended time trade-off method, which should be less prone to bias compared with other elicitation methods.[24]

Some limitations should also be acknowledged. As with all simulation studies, we do not know how far our findings may be extrapolated beyond the modeled situations. For instance, we only simulated a model with a uniform treatment effect across all mRS health state transitions, which, therefore, adheres perfectly to the proportional odds assumption. However, if the proportional odds assumption would be violated, and treatment effect would not be uniform across the different outcome categories, ordinal analysis would still be the most efficient.[6] Nevertheless, further validation of our results is required. Finally, we used the EuroQol Group 5-Dimension Self-Report Questionnaire assessed at 90 days after inclusion, which reflects neither short-term QoL nor the final health state. A better reflection of patient perception could be achieved by calculating QALYs based on multiple QoL measurements in 1 patient. Nevertheless, the aim of this study is not to describe QoL but to evaluate efficiency in detecting treatment effects.

In conclusion, the UW-mRS has been received as a promising new patient-centered outcome in stroke research. However, the UW-mRS does not capture individual variation in utilities within each mRS health state. Also, interpretation of treatment effect on the UW-mRS scale might be more challenging than was first suggested. Finally, clinicians and researchers should be aware of the reduction in power compared with ordinal analysis of the mRS when they use the UW-mRS as outcome measure in acute stroke intervention trials. More thorough evaluation of the UW-mRS in terms of its added value, analytic approach, and interpretation is required.

Sources of Funding

Disclosures

References

1. Lees KR, Bath PM, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al; European Stroke Organization Outcomes Working Group. Contemporary outcome measures in acute stroke research: choice of primary outcome measure. *Stroke*. 2012;43:1163–1170. doi: 10.1161/STROKEAHA.111.641423.

2. Lees KR, Khatri P; STAIR IX Collaborators. Stroke Treatment Academic Industry Roundtable recommendations for individual data pooling analyses in stroke. *Stroke*. 2016;47:2154–2159. doi: 10.1161/STROKEAHA.116.012966.

3. 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.

4. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332:1080. doi: 10.1136/bmj.332.7549.1080.

5. Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J; Optimising Analysis of Stroke Trials (OAST) Collaboration. Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. *Stroke*. 2007;38:1911–1915. doi: 10.1161/STROKEAHA.106.474080.

6. McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, et al. A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the IMPACT Project. *Clin Trials*. 2010;7:44–57. doi: 10.1177/1740774509356580.

7. Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, et al; IMPACT (International Mission on Prognosis and Clinical Trial Design in Traumatic Brain Injury) Study Group; CRASH (Corticosteroid Randomisation After Significant Head Injury) Trial Collaborators. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. *Crit Care*. 2011;15:R127. doi: 10.1186/cc10240.

8. Saver JL. Optimal end points for acute stroke therapy trials: best ways to measure treatment effects of drugs and devices. *Stroke*. 2011;42:2356–2362. doi: 10.1161/STROKEAHA.111.619122.

9. Ali M, Fulton R, Quinn T, Brady M; VISTA Collaboration. How well do standard stroke outcome measures reflect quality of life? A retrospective analysis of clinical trial data. *Stroke*. 2013;44:3161–3165. doi: 10.1161/STROKEAHA.113.001126.

10. Carod-Artal FJ, Egido JA. Quality of life after stroke: the importance of a good recovery. *Cerebrovasc Dis*. 2009;27(suppl 1):204–214. doi: 10.1159/000200461.

11. Kapoor A, Lanctôt KL, Bayley M, Kiss A, Herrmann N, Murray BJ, et al. "Good Outcome" isn't good enough: cognitive impairment, depressive symptoms, and social restrictions in physically recovered stroke patients. *Stroke*. 2017;48:1688–1690. doi: 10.1161/STROKEAHA.117.016728.

12. 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.

13. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al; DAWN Trial Investigators. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med*. 2018;378:11–21. doi: 10.1056/NEJMoa1706442.

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. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. *Med Decis Making*. 1994;14:9–18. doi: 10.1177/0272989X9401400102.

16. EuroQol Group. EuroQol–a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16:199–208.

17. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ*. 2006;15:1121–1132. doi: 10.1002/hec.1124.

18. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? *Stat Med*. 2009;28:3189–3209. doi: 10.1002/sim.3603.

19. 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.

20. 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.

21. Schreuders J, van den Berg LA, Fransen PS, Berkhemer OA, Beumer D, Lingsma HF, et al; MR CLEAN Investigators. Quality of life after intra-arterial treatment for acute ischemic stroke in the MR CLEAN trial-Update. *Int J Stroke*. 2017;12:708–712. doi: 10.1177/1747493017706244.

22. Gershon RC, Lai JS, Bode R, Choi S, Moy C, Bleck T, et al. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general

population testing. *Qual Life Res*. 2012;21:475–486. doi: 10.1007/s11136-011-9958-8.

23. Cnossen MC, Polinder S, Vos PE, Lingsma HF, Steyerberg EW, Sun Y, et al. Comparing health-related quality of life of Dutch and Chinese patients with traumatic brain injury: do cultural differences play a role? *Health Qual Life Outcomes*. 2017;15:72. doi: 10.1186/s12955-017-0641-9.

24. Post PN, Stiggelbout AM, Wakker PP. The utility of health states after stroke: a systematic review of the literature. *Stroke*. 2001;32:1425–1429.

25. Broderick JP, Adeoye O, Elm J. Evolution of the modified Rankin Scale and its use in future stroke trials. *Stroke*. 2017;48:2007–2012. doi: 10.1161/STROKEAHA.117.017866.

26. Prieto L, Sacristán JA. Problems and solutions in calculating quality-adjusted life years (QALYs). *Health Qual Life Outcomes*. 2003;1:80. doi: 10.1186/1477-7525-1-80.

27. Hughes RA, Newsom-Davis JM, Perkin GD, Pierce JM. Controlled trial prednisolone in acute polyneuropathy. *Lancet*. 1978;2:750–753.