How Well Do Standard Stroke Outcome Measures Reflect Quality of Life?
A Retrospective Analysis of Clinical Trial Data
Myzoon Ali, PhD; Rachael Fulton, MSc; Terry Quinn, MD; Marian Brady, PhD; on behalf of the VISTA Collaboration*

Background and Purpose—Quality of life (QoL) is important to stroke survivors yet is often recorded as a secondary measure in acute stroke randomized controlled trials. We examined whether commonly used stroke outcome measures captured aspects of QoL.

Methods.—We examined primary outcomes by National Institutes of Health Stroke Scale (NIHSS), Barthel Index (BI) and modified Rankin Scale (mRS), and QoL by Stroke Impact Scale (SIS) and European Quality of Life Scale (EQ-5D) from the Virtual International Stroke Trials Archive (VISTA). Using Spearman correlations and logistic regression, we described the relationships between QoL mRS, NIHSS, and BI at 3 months, stratified by respondent (patient or proxy). Using χ² analyses, we examined the mismatch between good primary outcome (mRS ≤1, NIHSS ≤5, or BI ≥95) but poor QoL, and poor primary outcome (mRS ≥3, NIHSS ≥20, or BI ≤60) but good QoL.

Results.—Patient-assessed QoL had a stronger association with mRS (EQ-5D weighted score n=2987, P<0.0001, r=−0.7, r²=0.53; SIS recovery n=2970, P<0.0001, r=−0.71, r²=0.52). Proxy responses had a stronger association with BI (EQ-5D weighted score n=837, P<0.0001, r=0.78, r²=0.63; SIS recovery n=867, P<0.0001, r=0.68, r²=0.48). mRS explained more of the variation in QoL (EQ-5D weighted score=53%, recovery by SIS v3.0=52%) than NIHSS or BI and resulted in fewer mismatches between good primary outcome and poor QoL (P<0.0001, EQ-5D weighted score=8.5%; SIS recovery=10%; SIS-16=4.4%).

Conclusions.—The mRS seemed to align closely with stroke survivors’ interests, capturing more information on QoL than either NIHSS or BI. This further supports its recommendation as a primary outcome measure in acute stroke randomized controlled trials. (Stroke. 2013;44:3161-3165.)

Key Words: Barthel Index ■ modified Rankin Scale ■ outcome ■ quality of life ■ stroke ■ trial

Traditional acute stroke randomized controlled trial (RCT) outcome measures often focus on impairments and activity, with functional outcome at 3 months commonly used as an end point. However, for stroke survivors, perception of recovery will have a much broader focus, and patient experience of stroke recovery may be better captured by quality of life (QoL) measures. Ideally, the assessment tools selected for use in RCTs should reflect a range of health-related function after stroke and should capture meaningful changes both for those affected by stroke and for the purposes of the trial. In choosing an outcome, the trialists must balance comprehensive assessment against the burden and cost of excessive testing. This can present a challenge because no single stroke outcome measure describes all dimensions of impairment, activity limitation, or participation. Furthermore, meaningful recovery varies between patients: what is considered as a good outcome by one stroke survivor may not be similarly rated by another. Despite this complexity, there is a growing recognition of the importance of understanding the impact of acute stroke interventions on QoL, in addition to effects on bodily functions.

Previous research has highlighted the need to identify and use appropriate outcome measures in stroke RCTs. The National Institutes of Neurological Disorders and Stroke Common Data Elements (NINDS CDE) project and the European Stroke Organisation Outcomes Working Group have recommended the modified Rankin Scale (mRS) as a primary outcome measure in trials of acute stroke interventions; the Barthel Index (BI) and the National Institutes of Health Stroke Scale (NIHSS) are also commonly recorded.

As QoL measures are not routinely collected as primary outcomes in trials of acute stroke interventions, it is important to assess whether existing measures should be qualified...
by additional QoL metrics. We, therefore, described the relationship between 2 frequently used QoL measures (choosing exemplars of generic- and stroke-specific QoL scales) and common primary outcome measures (mRS, BI, and NIHSS) at 3 months after stroke. We hypothesized that existing primary outcome measures may also reflect patient QoL.

Methods

Data

We accessed data from the Virtual International Stroke Trials Archive (VISTA), a repository for anonymized, completed stroke trials.2 We extracted data on patient age, initial stroke severity measured using the baseline NIHSS score, outcomes by European Quality of Life Scale (EQ-5D), 2 forms of the Stroke Impact Scale (SIS version 3.0 and SIS-16), mRS, NIHSS, and BI at 3 months. The EQ-5D is a generic health-related QoL scale. It assesses QoL, measuring domains such as mobility, self-care, usual activities, anxiety/depression, and pain/discomfort; a visual analog scale enables the patients to assess their own health state. These domains can then be summarized into a single weighted index. The SIS is a stroke-specific QoL scale that measures physical problems, memory, emotions, communication, activities of daily living, mobility, participation, hand function and patients’ perceptions of recovery. The SIS-16 is based on the physical functioning domains of the SIS. Both the visual analog scale of the EQ-5D and the recovery item of the SIS version 3.0 use a scale of 0 to 100 to indicate QoL and recovery, respectively. A score of 0 indicates the worst possible outcome, whereas a score of 100 indicates the best possible health state/recovery.

Analysis

We analyzed the strength of association between the weighted score for EQ-5D, visual analog scale of the EQ-5D, SIS-16, recovery by SIS version 3.0, and the mRS, NIHSS, and BI at 3 months after stroke. These associations were assessed using partial correlations, adjusting for age and initial stroke severity (baseline NIHSS). Spearman rank correlation coefficients were used to take into account the ordinal and nonparametric nature of the scales. Using logistic regression and calculating Nagelkierke generalized coefficient of determination ($r^2$) values, we examined the proportion of variation in QoL explained by each primary outcome measure, adjusting for age and initial stroke severity (baseline NIHSS).

Using $\chi^2$ analyses, we examined the mismatch between primary outcomes and QoL by describing the proportion of patients who not only achieved a good primary outcome, but also scored poorly on QoL. To determine QoL as a score in the worst quartile ($\leq 25$%) and good QoL as a score in the highest quartile ($\geq 75$%) of the distribution for this population.

We analyzed data from surviving patients at 3 months after stroke, where assessments were conducted for each of the QoL and primary outcome measures of interest. The 3-month follow-up period was selected as this is a commonly used end point in acute stroke trials; by this time point, most patients will have reached a plateau in functional recovery. Recognizing that properties of scales may differ by additional QoL metrics, we performed subgroup analyses limited to assessments completed by the subject and those completed by an informant or proxy.

Results

Data Set

Our analysis data set contained 4946 patients who were enrolled into acute stroke trials within 6 hours of stroke onset. By 3 months, 817 (16.5%) had died and 128 (2.6%) were lost to follow-up. Complete data were available for each of the outcome measures of interest in at least 3787 patients. The median age of patients in our data set was 71 years (interquartile range, 60–78); 54.9% were male (Table 1). EQ-5D scores were available for 3932 patients at 3 months; 3005 (76.4%) patients completed this assessment for themselves, whereas 857 (21.8%) required help from a proxy. At 3 months, SIS scores were available for 3872 patients; 2990 (77.2%) had completed assessments by themselves, and 882 (22.8%) required assistance from a proxy.

Correlations and Proportion of Variation in QoL Explained by Primary Outcome Measure

Analysis of data from all patients revealed that at 3 months, QoL defined using the visual analog scale of EQ-5D ($n=3828$, $P<0.0001$, $r=-0.64$) and recovery by SIS version 3.0 ($n=3838$, $P<0.0001$, $r=-0.73$) had a stronger association with mRS, with 47% and 59% of the variation in these QoL measures, respectively, being explained by mRS at 3 months, age, and initial stroke severity. The SIS-16 ($n=3856$, $P<0.0001$, $r=0.85$; Table 2) had a stronger association with BI, with 81% of the variation being explained by the BI score at 3 months, age, and initial stroke severity.

Patient Versus Proxy Respondents

We observed that the majority of patients’ responses on QoL had a stronger association with outcomes by mRS (EQ-5D weighted score $P<0.0001$, $r=-0.7$; EQ-5D visual analog scale $P<0.0001$, $r=-0.59$; SIS version 3.0 recovery $P<0.0001$, $r=-0.71$). The mRS also explained a higher proportion of the variation in QoL than either the NIHSS or BI, accounting for 53% of the variation in EQ-5D weighted scores, 39% of the variation in EQ-5D visual analog scale scores, and 52% of the variation in recovery by SIS version 3.0 (Table 2). The SIS-16 had a marginally stronger association with the BI ($P<0.0001$, $r=0.8$; Table 2).

Proxy responses had a stronger association with BI at 3 months (EQ-5D weighted score $P<0.0001$, $r=0.78$; EQ-5D weighted score $P<0.0001$, $r=0.80$; Table 2).

Table 1. Patient Demographics

<table>
<thead>
<tr>
<th></th>
<th>All Patients (n=4946)</th>
<th>Those Requiring Proxy Respondents (n=839)</th>
<th>Those Able to Self-Report (n=2594)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (IQR)</td>
<td>71 (60–78)</td>
<td>75 (66–80)</td>
<td>68 (58–76)</td>
</tr>
<tr>
<td>Baseline NIHSS score,</td>
<td>12 (8–17)</td>
<td>15 (11–20)</td>
<td>10 (7–14)</td>
</tr>
<tr>
<td>median (IQR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex, male, frequency n%</td>
<td>2715 (54.9)</td>
<td>419 (49.9)</td>
<td>1685 (57)</td>
</tr>
<tr>
<td>Atrial fibrillation, present, frequency n%</td>
<td>1271 (25.7)</td>
<td>259 (30.9)</td>
<td>622 (21.1)</td>
</tr>
<tr>
<td>Myocardial infarction, present, frequency n%</td>
<td>641 (13.0)</td>
<td>84 (10)</td>
<td>374 (12.7)</td>
</tr>
<tr>
<td>Hypertension, present, frequency n%</td>
<td>3665 (74.1)</td>
<td>630 (75.1)</td>
<td>2172 (73.5)</td>
</tr>
<tr>
<td>Diabetes mellitus, present, frequency n%</td>
<td>1135 (23.0)</td>
<td>214 (25.5)</td>
<td>621 (21)</td>
</tr>
</tbody>
</table>

IQR indicates interquartile range; and NIHSS, National Institutes of Health Stroke Scale.
Table 2. Strength of Association and Proportion of Variation in Quality of Life Explained by Primary Outcome Measures at 3 Months

<table>
<thead>
<tr>
<th></th>
<th>EQ-5D Weighted Scores at 3 mo</th>
<th>EQ-5D Visual Analog Scale at 3 mo</th>
<th>Stroke Impact Scale Recovery at 3 mo</th>
<th>Stroke Impact Scale-16 at 3 mo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>r</td>
<td>$r^2$</td>
<td>n</td>
</tr>
<tr>
<td>All patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI</td>
<td>3823</td>
<td>0.76</td>
<td>0.61</td>
<td>3827</td>
</tr>
<tr>
<td>mRS</td>
<td>3824</td>
<td>−0.75</td>
<td>0.62</td>
<td>3828</td>
</tr>
<tr>
<td>NIHSS</td>
<td>3787</td>
<td>−0.65</td>
<td>0.49</td>
<td>3790</td>
</tr>
<tr>
<td>Able to self-report</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI</td>
<td>2986</td>
<td>0.69</td>
<td>0.49</td>
<td>2987</td>
</tr>
<tr>
<td>mRS</td>
<td>2987</td>
<td>−0.7</td>
<td>0.53</td>
<td>2988</td>
</tr>
<tr>
<td>NIHSS</td>
<td>2976</td>
<td>−0.59</td>
<td>0.39</td>
<td>2977</td>
</tr>
<tr>
<td>Require proxy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI</td>
<td>837</td>
<td>0.78</td>
<td>0.63</td>
<td>840</td>
</tr>
<tr>
<td>mRS</td>
<td>837</td>
<td>−0.74</td>
<td>0.59</td>
<td>840</td>
</tr>
<tr>
<td>NIHSS</td>
<td>811</td>
<td>−0.65</td>
<td>0.43</td>
<td>813</td>
</tr>
</tbody>
</table>

All correlations were statistically significant at P<0.0001. BI indicates Barthel Index; EQ-5D, European Quality of Life Scale; mRS, modified Rankin Scale; and NIHSS, National Institutes of Health Stroke Scale.

Mismatch Between Primary Outcomes and QoL

**Good Primary Outcome and Poor QoL**

Using $\chi^2$ analysis, we examined the proportion of patients who were classified as having a good primary outcome (mRS ≤1, NIHSS ≤5, or BI ≥20) but who were also rated in the lowest quartile for QoL (≤25%). Use of the mRS (EQ-5D weighted score: $P<0.0001$, proportion with poor QoL=8.5%; SIS-16: $P<0.0001$, proportion with poor QoL=4.4%; Table 3), seemed to be superior to the NIHSS (EQ-5D weighted score $P<0.0001$, proportion with poor QoL=11.3%; SIS-16, $P<0.0001$, proportion with poor QoL=5%), in producing fewer mismatches with poor QoL.

**Poor Primary Outcome and Good QoL**

We described the proportion of patients who were classified as having a poor primary outcome (mRS ≥3, NIHSS ≥20, or BI ≤60) but who also rated their QoL in the highest quartile. Our $\chi^2$ analysis revealed that when using NIHSS ≥20 to describe poor primary outcome, no patients reported good QoL ($P<0.0001$ for all observations), compared with mRS ≥3 (EQ-5D weighted score $P<0.0001$, proportion with good QoL=1.1%; SIS-16 $P<0.0001$, proportion with good QoL=8.4%) or BI ≤60 (EQ-5D weighted score $P<0.0001$, proportion with good QoL=24%; SIS-16 $P<0.0001$, proportion with good QoL=0.2%; Table 3). Overall, there were fewer mismatches for poor primary outcome and good QoL than for good primary outcome and poor QoL.

**Discussion**

We observed that the primary outcome measures routinely used in acute stroke trials, mRS explained a higher proportion of the variation in patient-assessed QoL. Use of mRS also corresponded with fewer descriptions of poor QoL than either the NIHSS or BI. Proxy assessments of QoL were more strongly associated with BI. However, the BI has floor and ceiling effects and may lack sensitivity to assess deficits through the range of expected outcomes. Those with lower BI scores, in turn, are more likely to require assistance from a proxy. As the BI is based on activities that can be observed and assessed by proxies, performance in basic activities of daily living is likely to inform a proxy’s views on QoL, contributing to our observations of a stronger association between proxy responses on BI and QoL.

Previous work undertaken by the working groups for the NINDS CDE project and the European Stroke Organisation Outcomes Workshop have identified robust outcome measures for use in stroke RCTs, resulting in the recommendation of the...
mRS for use in acute stroke trials. This has been reinforced by intensive investigations of the reliability, validity, availability of training, and inter-rater variability of this scale. Our study builds on this previous research by describing the QoL information contained within the mRS.

Our study builds on previous work that examined the relationships between the EQ-5D, BI, and NIHSS. However, this analysis was conducted in a much smaller sample (n=67) and the correlations observed were weaker than in our study (EQ-5D and NIHSS [r=−0.40]; EQ-5D and modified BI [r=0.51]). Lai et al examined the relationships between the Short Form-36 (SF-36), SIS-16, and levels of the mRS, concluding that the SIS-16 captured more information on participation and physical functioning when compared with the SF-36. We did not examine the correlations between the SF-36 and primary outcome measures because lack of data on the SF-36. However, both the EQ-5D and SF-36 measure similar domains of health. Of the QoL measures investigated in our study, the visual analog scale of the EQ-5D had weakest correlation with each of the primary outcome measures, possibly attributable to the non-disease-specific nature of the measure.

Our study has several strengths. We selected outcome measures that were commonly used in acute stroke trials through examination of the literature. Our data were derived from VISTA that includes government and commercially sponsored trials. These trials involved rigorous monitoring and source data verification to meet worldwide regulatory standards. Training and certification for use of the mRS and NIHSS were also widely available at the time of trial enrollment, thereby minimizing interobserver variability. The patients enrolled and the outcome measures used within VISTA are representative of those available in acute stroke RCTs. Data on QoL were collected alongside the mRS, BI, and NIHSS. Therefore, measures of functional outcome, activities of daily living, neurological impairment, and QoL were available from the same patients at the same time points. At least 75% of patients in our data set assessed their own QoL and recovery, thereby providing us with robust data detailing the patients’ perspectives. The QoL measures we examined were reliable, validated, and recommended. Both the EQ-5D and SIS were recommended as measures of participation for data collection by the NINDS CDE project. Development of the SIS included the perspectives of patients, carers, and healthcare professionals. Furthermore, the SIS accurately assesses recovery after stroke and is sensitive to differences across the various levels of the mRS. The SIS-16 does not exhibit ceiling effects and can discriminate across the spectrum of stroke severity. Our definitions of recovery and QoL using these scales were, therefore, deemed to be valid and reliable.

Our study had some limitations. We analyzed data on only the EQ-5D and SIS; other QoL measures such as the SF-36 and the Stroke-Specific Quality of Life Measure were not available for inclusion in our analysis. Examination of these and other QoL measures may have strengthened our results. The properties of proxy-derived data may differ from direct patient assessments. Our data were a mix of proxy and patient assessments and so we prespecified subgroup analyses looking specifically at the proxy and direct patient assessment results. The QoL correlations for proxy-derived outcomes differed from the direct patient assessments and seemed to favor BI. Interpreting these data are problematic, particularly as the validity of many proxy-derived stroke assessments is not proven. For SIS, proxies score patients as more severely impaired in strength and activities of daily living. In the case of the EQ-5D, previous investigations have found moderate agreement between responses from patients and their proxies, with maximum agreement occurring when assessments were conducted at 6 months after stroke. As our study assessed outcomes at 3 months, the degree of agreement between patient and proxy ratings in our study may have been poorer than in previous studies. Despite the potential bias introduced by proxy rating of QoL, the inclusion of proxy responses are preferable to the alternative in those patients who are unable to report their own QoL. A larger, direct patient assessment data set is more compelling, and based on our data, we would favor use of mRS over BI.

Conclusions

For any outcome scale, there is a trade-off between ease of use and the richness of qualitative data that the scale can capture. Of the traditional acute stroke outcome measures examined, we observed a stronger association between almost all patient-assessed measures of QoL and the mRS at 3 months. The mRS also explained a higher proportion of the variation in QoL than either NIHSS or BI. We observed more mismatches between good primary outcome and poor QoL, than for the inverse combination. This may indicate that QoL assessments capture aspects of stroke recovery over and above the information gathered by standard outcome measures. Nevertheless, our findings suggest that mRS provides a useful indicator of a patient’s overall QoL at 3 months after stroke. However, if time and resource allows, qualifying mRS (particularly good mRS) with additional QoL data may be a useful exercise.

Appendix

VISTA Steering Committees

and C. Weimar; VISTA-Prevention: H. -C. Diener (Chair), G. Hankey, K. R. Lees, B. Ovbiagele, and C. Weir.

Sources of Funding
Drs Ali and Brady and the Nursing, Midwifery and Allied Health Professions Research Unit are funded by the Chief Scientist Office (CSO), Scottish Government’s Health Directorate, Scotland. This work was funded as part of a CSO Postdoctoral Fellowship Award. The views expressed here are those of the authors and not necessarily those of the CSO. VISTA is a not-for-profit collaboration of researchers from academia and commercial organizations. The VISTA Steering Committee members have each contributed to the organization of contributing trials, and where these have involved industry support, they have acknowledged that within the original publications.

Disclosures
Dr Quinn has created investigator training materials for commonly used stroke outcomes (mRS and BI) and has received honoraria from Training Campus for this work. The other authors have no conflicts to report.

References
How Well Do Standard Stroke Outcome Measures Reflect Quality of Life?: A Retrospective Analysis of Clinical Trial Data
Myzoon Ali, Rachael Fulton, Terry Quinn and Marian Brady
on behalf of the VISTA Collaboration*

*Stroke. 2013;44:3161-3165; originally published online September 19, 2013;
doi: 10.1161/STROKEAHA.113.001126

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/44/11/3161

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/