How Valid Are Family Proxy Assessments of Stroke Patients’ Health-Related Quality of Life?

Linda S. Williams, MD; Tamilyn Bakas, RN, DNS; Edward Brizendine, MS; Laurie Plue, MA; Wanzhu Tu, PhD; Hugh Hendrie, MB, ChB, DSc; Kurt Kroenke, MD

**Background and Purpose**—Proxy respondents are often needed to report outcomes in stroke survivors, but they typically systematically rate impairments worse than patients themselves. The magnitude of this difference, the degree of agreement between patients and proxies, and the factors influencing agreement are not well known.

**Methods**—We compared patient and family proxy health-related quality of life (HRQL) responses in 225 patient-proxy pairs enrolled in a clinical trial for poststroke depression. We used paired t-tests and the intraclass correlation (ICC) statistic to evaluate the agreement between patient and proxy domain scores and the overall Stroke-specific Quality of Life (SS-QOL) score. We used multivariate linear regression to model patient- and proxy-reported SS-QOL scores.

**Results**—Patients were older (63 versus 55 years) and less often female (48% versus 74%) than proxies. Proxies rated all domains of SS-QOL slightly worse than patients. The Mood, Energy, and Thinking domains had the greatest disparity with mean patient-proxy differences of 0.45, 0.37, and 0.37 points, respectively. The ICC for each domain ranged from 0.30 (role function) to 0.59 (physical function). Proxy overall SS-QOL score was also lower (worse) than patient score (3.7 versus 3.4, \( P < 0.001 \)) with ICC of 0.41. Agreement was higher among patient-proxy pairs with higher patient depression scores and with lower proxy report of caregiving burden.

**Conclusions**—Proxies systematically report more dysfunction in multiple aspects of HRQL than stroke patients themselves. Agreement between patient and proxy HRQL domain scores is modest at best and is affected by patient depression and proxy perception of burden. These differences may be large enough to impact the outcome assessment in stroke clinical trials. (Stroke. 2006;37:2081-2085.)

Key Words: caregiver ■ outcomes ■ quality of life ■ scales ■ stroke

**Self-report measures** are often used to assess stroke outcome. However, as many as 25% of stroke survivors may be unable to self-report their status as a result of language or other cognitive effects of stroke or as a result of preexisting conditions. In this instance, assessment of outcome from a secondary informant, often a family proxy, is typically substituted for the patient’s own report.

Although obtaining outcome information from a secondary informant is preferable to obtaining no outcome information at all, literature in many other conditions has shown that proxies often systematically rate patient outcomes worse than the patients rate their own outcomes. For example, family proxies of persons with stroke, brain cancer, dementia, epilepsy, and other disabilities, as well as proxies of general medical patients, have been found to report worse health-related quality of life (HRQL) outcomes than patients themselves. This observation is especially prominent for the more subjective domains of HRQL like emotional well-being or fatigue and is remarkably consistent across various HRQL scales.

The use of a proxy rating in place of a patient self-rating has practical implications for stroke trials, in which the status of the informant is not often recorded and even less often adjusted for in analyses of treatment effect. Inclusion of unbalanced numbers of family proxy respondents in different treatment groups may obscure the effect of an intervention and at the very least will lead to increased noise in outcome assessment, thus necessitating larger sample sizes to observe a given intervention effect size. Furthermore, patient-proxy disagreement would have an even greater impact on analysis of within-person change.

The objective of our analysis was to evaluate the agreement between patient and family proxy assessment of HRQL after ischemic stroke and, if scores were significantly different, to investigate the factors related to this difference. A secondary objective was to evaluate the factors that were associated with patient and with proxy HRQL ratings. Our primary hypothesis was that family proxy HRQL scores would be systematically worse than patient scores, and that
depressive symptoms in the patient and the family proxy would be associated with this difference.

Methods

Subjects were patients enrolled in the National Institute for Neurologic Disorders and Stroke-funded AIM ( Activate, Initiate treatment, Monitor) poststroke depression study. The AIM study consists of a randomized clinical trial of case-management intervention versus usual care in depressed subjects nested within a longitudinal cohort study that includes nondepressed subjects. Nondepressed subjects were matched 1:1 by site of enrollment to depressed subjects. Between 1 and 2 months posts ischemic stroke, eligible patients at four Indianapolis hospitals were screened for poststroke depression using the Patient Health Questionnaire (PHQ)-9 scale,11,12 which has been validated for use in stroke survivors.13 Patients with more than moderate aphasia (National Institutes of Health Stroke Scale [NIHSS] language item score >1)14 or cognitive impairment (modified six-item Mini-Mental Status score <3) were excluded.15 The local human subjects review board approved this study.

One family caregiver for each enrolled stroke survivor was also eligible to enroll in the study, but the presence of a caregiver was not a requirement for study entry. Family caregivers were eligible to enroll if they were in physical contact with the stroke survivor at least 3 days per week and were providing at least two caregiving tasks. Baseline patient data collection included assessment of demographic data, stroke-specific quality of life (SS-QOL),16 stroke impairments at the time of admission, using a validated retrospective measure of the NIHSS,14 and depression. The SS-QOL assesses stroke-specific QOL in seven domains and provides an overall estimate of HRQL. Scores range from 1.0 to 5.0 with higher scores indicating better HRQL. Patient depression was determined by screening with the PHQ-9 and depression diagnostic assessment with the Structured Clinical Interview for Depression (SCID), considered a criterion standard for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition depressive disorder diagnoses in clinical research.17,18 Baseline family caregiver data collection included demographic data, caregiver burden with the Oberst Caregiving Burden Scale (OCBS),19 global caregiver outcomes (Bakas Caregiving Outcomes Scale),20 and depression (PHQ-9).11 The 15-item OCBS measures perceived difficulty with caregiver tasks such as providing personal care, assisting with medications, managing emotions and behaviors, and dealing with finances, among others. In addition, family caregivers rated the patient’s stroke-specific quality of life outcomes using a proxy version of the SS-QOL.

Statistical Methods

Paired t tests were used to determine whether there were systematic differences between patient and proxy SS-QOL domain and overall scores. Intraclass correlation coefficients (ICCs) were estimated to assess the strength of agreement between patients and proxies on domain and overall SS-QOL scores.21 As an aid in interpreting the ICCs, we used the metric defined by Landis and Koch (0 to 0.2 indicating slight agreement, 0.21 to 0.4 fair agreement, 0.41 to 0.6 moderate agreement, 0.61 to 0.8 substantial agreement, and 0.81 to 1.0 almost perfect agreement).22 We further assessed the patient-proxy agreement by performing a Bland-Altman agreement analysis and visually inspecting the resulting Bland-Altman plots.23 Multiple linear regression was used to model the difference in patient and proxy dyad SS-QOL scores controlling for NIHSS at stroke onset, patient depression symptoms, caregiver depression symptoms, and the difficulty subscore of the OCBS. To determine variables independently related to patient and proxy SS-QOL ratings, we again used linear regression to construct separate models of patient SS-QOL scores and proxy SS-QOL scores controlling for demographic variables, NIHSS at the baseline evaluation, depression, and caregiver burden (in the proxy-reported SS-QOL models only). SAS version 8.2 (SAS Institute) was used for all analyses.

Results

Of the total enrolled patient sample of 392, 227 (58%) had a family caregiver enroll at baseline. For these analyses, the 225 patient-caregiver dyads with complete baseline data are included. Demographic data for the 225 patient-proxy pairs are shown in Table 1. As expected, patients were older than their family caregiver, and more caregivers than patients were female. The mean OCBS task difficulty score was 21 in the caregivers, indicating overall slight difficulty with caregiving tasks. Mean PHQ-9 scores were 9 in the patients and 5 in the proxies. Given the study design with nondepressed subjects enrolled and matched to depressed subjects, 42% of the patients were depressed; 21% of the caregivers also screened positive for depression. Caregiver burden and depression were modestly correlated (Pearson r=0.53, P<0.001 for OCBS difficulty score and proxy PHQ-9 score).

Mean patient scores in each SS-QOL domain and the overall SS-QOL score were higher than mean proxy scores (Figure 1). Differences ranged from 0.22 in the Role Function...
domain to 0.45 in the Mood domain; all differences were statistically significant at \( P < 0.001 \). Agreement between individual patient–proxy scores was only fair to moderate (Figure 2) with the best agreement in Physical Functioning and the lowest agreement in Thinking and Role Function.

Modeling the difference in patient and proxy scores, we found that increased depression symptoms in the patient and decreased perception of burden by the caregiver were associated with greater patient–proxy agreement (Table 2, \( r^2 = 0.15 \)). Stroke impairments (NIHSS score), proxy depression, and patient/proxy demographic characteristics (age, gender, race, or caregiving relationship) were not independently associated with agreement. The Bland-Altman plot of the SS-QOL scores (Figure 3) showed that although patient SS-QOL scores were on average slightly higher than the proxy scores, the within-dyad differences of the overall SS-QOL did not appear to vary systematically with the average overall patient-reported SS-QOL. As expected, the vast majority of SS-QOL difference scores were between \( \pm 2 \) standard deviation limits (–1.1 and 1.7), suggesting a well-behaved bell shape distribution for the difference scores. Individual multivariable models of patient- and proxy-reported overall SS-QOL score demonstrated that higher patient-reported SS-QOL scores were associated with less stroke impairment and lower patient depression scores, whereas higher proxy-reported SS-QOL scores were associated with less stroke impairment, less caregiver depression, and less caregiver burden (Table 2). Patient and proxy demographic variables were not independently associated with SS-QOL scores in any of these models.

**Discussion**

These data show that stroke survivors and their family proxies have only relatively modest agreement on assessment of functioning and quality of life of the stroke survivor at 1 to 2 months poststroke. In all functional domains, the family proxy rated quality of life worse than the stroke survivor themselves rated that domain with agreement being slightly higher in more objective domains like physical functioning and lower in more subjective domains like mood. Although this tendency has been shown in other studies of patient–proxy assessments, a few studies specifically report the level of agreement poststroke.

An important finding from our study is that agreement varies by patient depression status and proxy perception of

---

**Figure 1.** Mean patient–proxy difference in overall and domain SS-QOL scores. Horizontal bars represent 95% CI around the estimate of the score difference.

**Figure 2.** Intraclass correlations for SS-QOL overall and domain scores. Horizontal bars represent 95% CI around the estimate of the ICC.

**Figure 3.** A Bland-Altman plot of the difference in the SS-QOL versus the average patient overall SS-QOL score.

**Table 2. Multivariate Models of Patient and Proxy SS-QOL Scores**

<table>
<thead>
<tr>
<th>Model of patient–proxy overall SS-QOL score difference ( (r^2 = 0.15) )</th>
<th>( t ) Value</th>
<th>( P )</th>
<th>Standardized Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient depression (PHQ-9 score)</td>
<td>-5.01</td>
<td>&lt;0.001</td>
<td>-0.32</td>
</tr>
<tr>
<td>Caregiver burden (OCBS score)</td>
<td>4.26</td>
<td>&lt;0.001</td>
<td>0.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model of patient-reported SS-QOL overall score ( (r^2 = 0.54) )</th>
<th>( t ) Value</th>
<th>( P )</th>
<th>Standardized Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke impairments (NIHSS score)</td>
<td>-3.79</td>
<td>&lt;0.001</td>
<td>-0.17</td>
</tr>
<tr>
<td>Patient depression (PHQ-9 score)</td>
<td>-15.07</td>
<td>&lt;0.001</td>
<td>-0.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model of proxy-reported SS-QOL overall score ( (r^2 = 0.29) )</th>
<th>( t ) Value</th>
<th>( P )</th>
<th>Standardized Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke impairments (NIHSS score)</td>
<td>-4.06</td>
<td>&lt;0.001</td>
<td>-0.24</td>
</tr>
<tr>
<td>Proxy depression (PHQ-9 score)</td>
<td>-2.35</td>
<td>0.020</td>
<td>-0.16</td>
</tr>
<tr>
<td>Caregiver burden (OCBS score)</td>
<td>-5.06</td>
<td>&lt;0.001</td>
<td>-0.35</td>
</tr>
</tbody>
</table>
burden. Prior studies that have investigated patient and/or proxy variables that impact agreement on patient outcome assessment have found that increased patient and proxy educational levels, family relationship (versus other type of proxy), lower levels of disability, and less patient depression were associated with greater patient–proxy agreement. In our study, we found that depressed patients had greater agreement with their proxy’s rating of patient HRQL, consistent with that of Sands et al., but whether this represents “accurate” reporting from the patient, “overestimation” of dysfunction by the proxy, or both is not clear. Our finding that increased proxy burden is associated with greater disagreement also corresponds to data from dementia caregivers and underscores prior reports that caregivers with greater stress and more time investment in caregiving report more negative assessment of patient HRQL. These data suggest that caregiver task difficulty is an important aspect affecting caregiver outcomes and likely patient–caregiver interactions poststroke; troublesome tasks such as managing finances, managing patient behaviors, and providing emotional support may be especially important areas to target in caregiver support interventions. In our analysis, caregiver depression was associated with proxy HRQL ratings but not with agreement between patient and proxy HRQL. The relatively modest prevalence of depression among caregivers (18%) limits our ability to assess the impact of proxy depressive symptoms on agreement.

What is difficult to determine is whether the patient or the proxy is closer to the “truth.” For example, depressed patients report lower HRQL and these scores are more consistent with proxy HRQL ratings, but whether the patient’s report of lower HRQL is accurate or whether it is lower than their “actual” HRQL resulting from the overlying depression is not clear. Although this notion of truth may be elusive, it is usually best to view the patient’s self-report as the more accurate, but to also take into consideration family comments regarding the stroke survivor’s depressive symptoms and quality of life in the clinical setting. It is also critical to acknowledge that the patient and the family proxy have different factors influencing their assessments of the impact of stroke.

Although to some extent it may thus be best to agree to disagree, it is particularly important to evaluate the magnitude and direction of agreement in patient–proxy outcomes in stroke and other conditions associated with neurologic dysfunction. Several studies that report patient–proxy differences have found that these differences are small and of questionable clinical meaning, suggesting that the benefit of using proxy responses, albeit with the potential for bias, outweighs the risk of not including outcome data for persons unable to respond. Although our data also show small mean differences of questionable clinical relevance in SS-QOL domain scores, the finding that agreement is only modest would suggest caution in direct substitution of proxy scores, especially for within-patient comparisons. In the clinical setting, the impact of differences in outcome reporting between patients and proxies may be minimal, but these differences can potentially have a large impact on clinical trials. For example, if outcome measures used to evaluate the intervention are completed either by the patient or the proxy, and if treatment groups are unbalanced in the proportion of outcome assessments completed by proxies, there may be less power to detect an effect of the intervention. One potential method to take these differences into account is to adjust the models of intervention affect by respondent status (patient versus proxy). However, if agreement between patients and their family proxies is only modest at best, simple model adjustment is not likely to be effective or meaningful. Other authors have suggested propensity score matching as a possible strategy, although this approach may reduce but not entirely eliminate differences in patient and proxy scores.

Evaluation of HRQL reports by proxy respondents presumes that both parties (the patient and the proxy) can report equally reliably on the patient’s outcome. As such, the results from such studies of patient–proxy pairs must always be extrapolated to patients who are completely unable to respond and thus have only proxy HRQL ratings. Although our enrollment criteria eliminated stroke survivors with severe language and cognitive effects of stroke, we did include patients with at least mild language and cognitive effects of stroke in an effort to be as inclusive of the spectrum of stroke impairments as possible. We also studied patients who consented to participate in a clinical trial, so agreement may be different than in a less selected cohort of stroke survivors. Nonetheless, these data demonstrate that there are important differences in patient and family proxy assessments of function and HRQL poststroke and that these differences can be large enough to impact the assessments of interventions in clinical trials. Future studies should not only assess the status of the respondent reporting outcome data, but should ensure that family proxy respondents are balanced between treatment groups. Clinicians should be alert to the impact of patient depression and caregiver burden on HRQL to provide appropriate interventions to improve patient and caregiver outcomes poststroke.

Acknowledgments

We thank Connie Dagon, Carrie Dixon, Monta Gazvoda, Carol Kempf, Gloria Nicholas, Kim Moran, and Jennifer Stuart for interviewing stroke survivors and their family proxies.

Sources of Funding

This work was supported by a grant from the National Institutes of Health, National Institutes of Neurological Disorders and Stroke R01 NS3 9571. Dr Williams was supported by an Advanced Career Development Award from the Department of Veterans’ Affairs, Health Services Research and Development.

Disclosures

None.

References


How Valid Are Family Proxy Assessments of Stroke Patients' Health-Related Quality of Life?
Linda S. Williams, Tamilyn Bakas, Edward Brizendine, Laurie Plue, Wanzhu Tu, Hugh Hendrie and Kurt Kroenke

Stroke. 2006;37:2081-2085; originally published online June 29, 2006; doi: 10.1161/01.STR.0000230583.10311.9f

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/37/8/2081

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/