Evaluation of Patient and Proxy Responses on the Activity Measure for Postacute Care

Alan M. Jette, PhD, PT; Pengsheng Ni, MD, MPH; Elizabeth K. Rasch, PT, PhD; Jed Appelman, PhD; M. Elizabeth Sandel, MD; Joseph Terdiman, MD, PhD; Leighton Chan, MD, MPH

Background and Purpose—Our objective was to examine the agreement between adult patients with stroke and family member or clinician proxies in activity measure for postacute care (AM-PAC) summary scores for daily activity, basic mobility, and applied cognitive function.

Methods—This study involved 67 patients with stroke admitted to a hospital within the Kaiser Permanente of Northern California system and were participants in a parent study on stroke outcomes. Each participant and proxy respondent completed the AM-PAC by personal or telephone interview at the point of hospital discharge or during ≥1 transitions to different postacute care settings.

Results—The results suggest that for patients with a stroke proxy, AM-PAC data are robust for family or clinician proxy assessment of basic mobility function and clinician proxy assessment of daily activity function, but less robust for family proxy assessment of daily activity function and for all proxy groups’ assessments of applied cognitive function. The pattern of disagreement between patient and proxy was, on average, relatively small and random. There was little evidence of systematic bias between proxy and patient reports of their functional status. The degree of concordance between patient and proxy was similar for those with moderate to severe strokes compared with mild strokes.

Conclusions—Patient and proxy ratings on the AM-PAC achieved adequate agreement for use in stroke research when using proxy respondents could reduce sample selection bias. The AM-PAC data can be implemented across institutional as well as community care settings while achieving precision and reducing respondent burden. (Stroke. 2012;43:824-829.)

Key Words: disability evaluation ■ rehabilitation ■ stroke assessment ■ stroke outcome

Patient-reported measures of function are important outcomes in stroke research. However, with as many as one-quarter of stroke survivors unable to report their status as a result of language or other cognitive deficits, functional assessments from another individual, often a family member or clinician, are typically substituted for the patient’s perspective.1–6 Whereas the use of proxies may reduce selection bias or clinician, are typically substituted for the patient's perspective assessments from another individual, often a family member or clinician proxies in activity measure for postacute care (AM-PAC) summary scores for daily activity, basic mobility, and applied cognitive function.1–6 Wherever the use of proxies may reduce selection bias in stroke research, there are concerns that disagreement between proxy and patient responses may introduce significant measurement error and bias into study results. Patient–proxy disagreement may have an impact on analysis of within-person functional change, may introduce noise in functional outcome assessment, and may necessitate larger sample sizes in clinical trials to observe a given treatment effect. The inclusion of unbalanced numbers of proxy responses in different treatment groups may bias the evaluation of an intervention if not appropriately adjusted for in the analysis.

The reliability of proxy respondents has been examined in generic measures of function,3,7–10 as well as in stroke-specific instruments.11–13 Previous research in patients with stroke has shown that proxies often systematically rate patient function worse than the patient’s own report,11–14 and that the level of agreement between proxies and patients may differ depending on the type of outcome being assessed. Adequate agreement tends to be reported for the assessment of more concrete and observable domains of function such as physical function, whereas lower levels of agreement are more frequently seen in more subjective domains such as psychosocial functioning.12,15–18 Agreement between proxy and patient reports has also been reported to be lower among more patients with more severe strokes,3,11,19,20 Across different types of proxy respondents, agreement between the patient and proxy rating tends to be better for family members and lower for health care providers.14,21 Good agreement between proxy and patient is a necessary criterion for a functional outcome measure to be useful in stroke research. The activity measure for postacute care (AM-PAC), created to assess patient-reported function across postacute care settings, con-

Received March 8, 2011; accepted November 21, 2011.
From the Boston University Medical Campus (A.J., P.N.), School of Public Health, Health & Disability Research Institute, Boston, MA; National Institutes of Health (E.K.R., L.C.), Mark O. Hatfield Clinical Research Center, Rehabilitation Medicine Department, Bethesda, MD; Kaiser Foundation Rehabilitation Center (J.A.), Vallejo, CA; Physical Medicine and Rehabilitation (M.E.S.), Napa/Solano Service Area, Research and Training, Kaiser Foundation Rehabilitation Center, Vallejo, CA; and Division of Research (J.T.), Kaiser Permanente, Oakland, CA.
Correspondence to Alan M. Jette, PhD, PT, Boston University Medical Campus, School of Public Health, Health & Disability Research Institute, 715 Albany Street, T3W, Boston, MA 02118-2526. Email Ajette@bu.edu
© 2012 American Heart Association, Inc.
Stroke is available at http://stroke.ahajournals.org DOI: 10.1161/STROKEAHA.111.619643

824
sists of 3 domains of function: basic mobility, daily activity, and applied cognitive functioning. In previous research, although acceptable proxy–patient agreement has been reported across all 3 domains of the AM-PAC (intraclass correlation coefficient ICC = 0.68–0.90) in patients with major medical, orthopedic, and neurological conditions who were receiving postacute care, no reports have been published that focus specifically on patients with stroke and that included comparisons across different types of proxy respondents and across different care settings. In this study, therefore, we examined the agreement between family member and clinician proxy reports with those of patient on the 3 AM-PAC functional domains among survivors of stroke who were receiving postacute care in a variety of institutional and community-based settings.

Materials and Methods

AM-PAC Instrument

The AM-PAC instrument was designed to measure functional status in adults across all postacute care settings. Early content and analytic work with the AM-PAC established 3 distinct activity content domains on which 3 AM-PAC scales were constructed: basic mobility, daily activities, and applied cognitive functioning. The AM-PAC instrument, developed on a sample of 1041 post acute care patients, uses a 4-point difficulty scale rating scale (no difficulty, a little difficulty, a lot of difficulty, cannot/unable to do) across 259 functional tasks and activities. In this study, the computer-adaptive test version (CAT) of the AM-PAC was used. In CAT administration of the AM-PAC, an iterative computer program uses information from a subject’s previous responses to tailor item selection, thereby eliminating questions that are too difficult or too easy for a particular patient. In the AM-PAC CAT, a small sample of the overall items in an instrument are administered and all scores are generated on the same metric, regardless of the number of items administered, thus facilitating comparisons across time or across groups with different functional levels while reducing respondent burden. We have conducted a series of simulated validation tests of the AM-PAC CAT by personal interview or by telephone by a trained data collector. Participants and proxy respondents provided responses to questions in all 3 AM-PAC domains at the point of hospital discharge or during transitions to different postacute care settings (inpatient rehabilitation hospital, skilled nursing facility, home care, outpatient therapy). At each assessment period, we collected patient-reported data from the AM-PAC CAT and proxy reports from a physician or nurse or family member using the AM-PAC CAT. If the patient was receiving care in an institution, then the first choice of proxy was a clinician, whereas a family member was the first choice if the patient was living at home. These patient and proxy AM-PAC assessments were completed within 1 to 7 days of each other, with a mean interval of 3 days. Self-report data were always collected before obtaining proxy information. In this study, we report 67 complete pairs of family–proxy AM-PAC data and 63 clinician–proxy pairs.

Statistical Analyses

As an overall measure of patient–proxy agreement, we used the linear mixed model to calculate the ICC, which represented the ratio between subject variance to the total variance. We used the linear mixed model to estimate ICC coefficients because we had multiple assessments for some subjects and this model could account the correlation between assessments for each subject. We included the time between stroke onset and assessment, the rater effect, and the interaction between those 2 variables as fixed effects in the model. The subject effect was treated as the random effect. The standard error of ICC was calculated based on delta method. For family–proxy data, we compared the ICC values in each subgroup, such as first visit versus second visit or more, in person versus by telephone, and in institution versus at home. These subgroup comparisons were not performed for clinician–proxy and clinician–patient pairs because of sample size limitations. For group estimates, reliability was considered high if the ICC was >0.80, substantial if it was between 0.61 and 0.8, moderate between 0.41 and 0.60, and poor to fair if the ICC was <0.4.

In addition to ICC, we constructed a folded cumulative distribution curve for each AM-PAC scale (called mountain plots) and then generated separate plots for each proxy type (clinician or family). To construct a mountain plot, we calculated the difference scores between proxy respondent and patient respondent (proxy minus self-report) for each AM-PAC scale and then sorted these difference scores in ascending order within each domain. Then, the percentile rank for each difference score (the proportion of scores less than and equal to that score) was computed. It should be noted that because we “folded” the empirical cumulative distribution plot at the line y = 50%, percentile ranks for scores ranked on the second half were calculated using 100 minus the actual percentile rank. Finally, the mountain plot was generated by plotting the percentile rank against the difference score for each AM-PAC scale. For within-group (family and clinician proxy) mountain plots, we recalculated the percentile ranks separately for each group.

Results

The 67 participants in this proxy substudy included a wide age range (30–94 years old); 58.2% were female. Twenty-two percent of subjects participating in this study had moderate to severe stroke severity according to the modified National Institutes of Health stroke scale. The majority of subjects (92.5%) had an ischemic stroke (Table 1).

Table 2 presents the overall ICC coefficients by type of proxy. The overall patient–proxy reliability coefficients ranged from 0.50 for the applied cognitive scale, to 0.63 for the daily activity scale, and to 0.72 for the basic mobility scale. Clinician proxies showed a trend toward less agreement in the applied cognitive scale (0.41) than did family member proxies (0.59), whereas clinicians showed more agreement in the daily activity scale (0.78) than did family member proxies (0.57). Agreement was highest and very similar across type of proxy for the basic mobility scale. (Table 2 and Figure 1). The mean difference score for patient-reported versus proxy-reported data across all 3 AM-PAC domains is quite modest (0.15–1.39). The magnitude of patient–proxy agreement across all 3 AM-PAC domains was not different for those...
Table 1. Baseline Characteristics of Patients in the Kaiser Proxy Substudy

<table>
<thead>
<tr>
<th>Variable</th>
<th>N=67</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average age, y (SD)</td>
<td>67.3 (13.6)</td>
</tr>
<tr>
<td>Age distribution, % (n)</td>
<td></td>
</tr>
<tr>
<td>Younger than 60 y</td>
<td>25.3 (17)</td>
</tr>
<tr>
<td>60–69 y</td>
<td>26.9 (18)</td>
</tr>
<tr>
<td>70–79 y</td>
<td>29.8 (20)</td>
</tr>
<tr>
<td>80 y or older</td>
<td>18.0 (12)</td>
</tr>
<tr>
<td>Gender, % (n)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>58.2 (28)</td>
</tr>
<tr>
<td>Severity of Stroke (NIHSS) % (n)</td>
<td></td>
</tr>
<tr>
<td>Mild (0–3)</td>
<td>77.6 (52)</td>
</tr>
<tr>
<td>Moderate to severe (4–31)</td>
<td>22.4 (15)</td>
</tr>
<tr>
<td>Stroke ype % (n)</td>
<td></td>
</tr>
<tr>
<td>Ischemic</td>
<td>92.5 (62)</td>
</tr>
<tr>
<td>Hemorrhagic</td>
<td>7.5 (5)</td>
</tr>
<tr>
<td>Charlson Index (Deyo modification) (SD)</td>
<td>1.5 (1.7)</td>
</tr>
<tr>
<td>Modified Rankin Scale Score (SD)</td>
<td>.04 (0.97)</td>
</tr>
<tr>
<td>Days from stroke to first AM-PAC assessment (SD)</td>
<td>3.9 (2.9)</td>
</tr>
<tr>
<td>AM-PAC function at acute hospital discharge (SD)</td>
<td></td>
</tr>
<tr>
<td>Basic mobility (range = 4.2–78.6)</td>
<td>52.9 (14.1)</td>
</tr>
<tr>
<td>Daily activity (range = 7.3–70.4)</td>
<td>45.6 (13.9)</td>
</tr>
<tr>
<td>Applied cognitive (range = 28.9–58.8)</td>
<td>43.4 (7.5)</td>
</tr>
</tbody>
</table>

SD indicates standard deviation; AM-PAC, activity measure for postacute care; NIHSS, National Institutes of Health Stroke Scale.

Table 2. Overall Intraclass Correlation Coefficient Correlations and by Type of Proxy Respondent for Activity Measure for Postacute Care Scale Scores

<table>
<thead>
<tr>
<th>Activity Measure for Postacute Care Domain</th>
<th>Family Proxy (N=67)</th>
<th>Clinician Proxy (N=63)</th>
<th>Total Proxies (N=67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC (95% CI)</td>
<td>ICC (95% CI)</td>
<td>ICC (95% CI)</td>
<td></td>
</tr>
<tr>
<td>Basic mobility</td>
<td>0.72 (0.06)</td>
<td>0.80 (0.05)*</td>
<td>0.72 (0.05)</td>
</tr>
<tr>
<td>Daily activity</td>
<td>0.57 (0.07)</td>
<td>0.78 (0.05)</td>
<td>0.63 (0.06)</td>
</tr>
<tr>
<td>Applied cognition</td>
<td>0.59 (0.07)</td>
<td>0.41 (0.12)</td>
<td>0.5 (0.07)</td>
</tr>
</tbody>
</table>

CI indicates confidence interval; ICC, intraclass correlation coefficient.

*N=62.

Discussion

The results of this study suggest that for patients with a stroke proxy data are sufficiently concordant with patient self-report for use in research in which patient-reported outcomes are of interest. The data reveal that concordance is most robust for family or clinician proxy assessment of basic mobility function and clinician proxy assessment of daily activity function, but less concordant for family proxy assessment of daily activity function and all proxies’ assessments of applied cognitive function. In contrast to previous research, proxy respondents did not appear to systematically rate patients as functioning at a lower level in comparison to the patient reported values. It is unclear why this occurred. It could be that there is something inherent in the administration of a CAT instruments that makes them less prone to this bias. Most of the articles documenting this phenomenon used standard paper and pencil instruments. Furthermore, the degree of concordance between patient and proxy was equal to or slightly better for patients with more severe strokes, those who would be most likely to need a proxy because of cognitive or language impairments. In using the AM-PAC among survivors of a stroke, the pattern of disagreement between patient and proxy appears to be, on average, relatively small, random, and not clinically meaningful. The overall agreement between patient-reported and proxy-reported function was within acceptable limits, with little evidence of systematic bias between proxy and patient reports of their functional status. That agreement was lowest in the domain of applied cognitive functioning is not surprising because this domain of functioning is less observable than the areas of mobility and daily activities. Because CAT tools administer different items across patients, it was not possible to analyze which subdomains of applied cognitive functioning led to the lower levels of concordance.

The magnitude of agreement between patient and proxy reports of functioning seen in this study was similar to that reported in previous studies of the AM-PAC in different patient groups and with the literature that compares patient and proxy agreement using other functional outcome instruments. Similar to existing literature, ICC coefficients for the more observable and more concrete domains of the AM-PAC (ie, basic mobility) was substantial to high in contrast to the coefficients for the more subjective and less observable domain (ie, applied cognitive), which was only moderate, although this finding varied across the methods we used to assess concordance. In contrast to other studies, agreement between proxy and patient was the same for those patients with a more severe stroke and those for whom proxy reports are most likely to be needed and proxies did not appear to systematically rate patients as...
functioning at a lower level in comparison to the patient-reported values.11–14,33

We believe the mountain plot approach used in this study to examine the magnitude and direction of discordance between patient and proxy reports is an important complement to the interpretation of standard reliability indices such as the ICC.34 The mountain plot approach was developed for an earlier report of the AM-PAC proxy reliability24 and displays the cumulative distribution of the difference scores between proxy and patient raters. The graph is created by computing a percentile rank for ranked difference between respondents’ scores and folded at the 50th percentile rank (percentile rank is defined as the proportion of cases having lesser or equal value to the score undergoing consideration, i.e., the mountain plot cumulative percentages (y-axis values) against the ranked difference scores (x-axis values). In folding the graph, the percentile ranks for difference scores above the 50th percentile are obtained by subtracting the actual percentile rank from 100. This mountain plot technique allows the reader to locate the median immediately, clearly illustrates degree of symmetry in the patient–proxy differences, and identifies outliers to determine central or tail percentiles. In this sample, the agreement appears to be quite consistent across AM-PAC domains, with no evidence of systematic bias. The absence of a trend in the direction and magnitude of the scoring difference between type of proxy and patient is clearly displayed in Figures 1 and 2 because the plots peak at x=0 and are symmetrical at approximately the line of x=0. If proxy reports were systematically different than those of patients, then the mountain peak would have shifted either to the left or to the right of the line x=0.

Although summary indices such as the ICC are the standard in the literature for examining proxy–patient agreement, we have some concerns in using this approach exclusively

![Figure 1. Intraclass correlation coefficient correlations by type of proxy respondent for each activity measure for postacute care (AM-PAC) computer-adaptive test (CAT) domain.](image1)

![Figure 2. Mountain plot for each activity measure for postacute care (AM-PAC) domain of differences between patient and overall proxy reports.](image2)
because the amount of variability in scores within the sample being evaluated has a strong influence on the magnitude of this statistic. For instance, a very low ICC can be obtained when agreement is low or, alternatively, if the variability in the sample is severely restricted. This inconsistency across approaches can be seen in the AM-PAC data we analyzed in this study. For the AM-PAC, the variability of the applied cognitive scale (SD, 8.5) was substantially smaller than that of the daily activity (SD, 15.6) and basic mobility (SD, 13.9) scales. Consequently, it is difficult to interpret whether the lower ICC value for the applied cognitive scale was attributable to lower agreement or to the ratio of the disagreement relative to the underlying variability in the sample. The mountain plot results revealed that the magnitude of agreement between patient and proxy is similar across all 3 AM-PAC domains. Another limitation with the ICC approach is that the measure of agreement (the ICC) is not in the metric of the functional instrument, which makes practical interpretation more difficult. It is for this reason we presented multiple approaches to examining proxy–patient agreement in this study.

There are several limitations to this study. Because of the limited number of patient–clinician proxy pairs, we were unable to compare agreement and bias within subgroups defined by mode of data collection, location where the assessment was performed, or first versus subsequent assessment. Furthermore, the study sample was limited to patients who were able to provide a self-report on the AM-PAC, and therefore the study sample excludes those who would typically require a proxy report of their functioning. Thus, we have to assume the results found in this self-selected subsample can be generalized to the population needing a proxy. The overall study sample was quite small, with a low representation of subjects with a severe stroke. Another concern in this analysis was the lack of independence of the pairs across settings, but this was adjusted for in our analyses.

Although the basic mobility portion of the AM-PAC is quite robust when considering patient proxy comparisons, further work may be necessary to more confidently validate the AM-PAC for all proxy types across other domains through multiple patient severity rankings and poststroke time points.

There are suggested patterns of change in ICC coefficients illustrated in Figure 1 that may be important for future investigation. One such pattern is the magnitude and direction of change observed in the daily activity function domain. In most cases, the ICC for daily activity agreement shifts ICC category, ie, is in the substantial range for first visit but declines to moderate for second visit; there are similar declines from in-person to telephone and for the in-institution to at-home comparisons. The basic mobility and applied cognitive subscales appear to be more consistent across these categories. Further study is also needed on the potential effects of using telephone versus face-to-face assessment.

In conclusion, patient and proxy ratings on the AM-PAC CAT are sufficiently concordant for use in stroke research in which patient-reported outcomes are of interest and in which the inclusion of proxy reports can reduce selection bias. There
was little evidence of systematic bias between proxy and patient reports of their functional status using the AM-PAC CAT instrument. Use of the AM-PAC CAT using patient-reported or proxy-reported can be implemented in 2 to 3 minutes per domain and can be used across institutional as well as community-based and home care settings, thus achieving precise outcome estimates while reducing respondent burden. Although some degree of error is present when using a proxy in place of the patient, the pattern of disagreement between patient and proxy appears to be, on average, relatively small, random, and not clinically meaningful.

Acknowledgments

The authors acknowledge the work of our data collectors who worked on this project. Lida Zarrabi and Sarah-Jane Kim recruited, enrolled, and followed-up each of the subject-proxy pairs through postacute care. The authors also acknowledge Mary Slavin for training the staff.

Sources of Funding

This research was supported, in part, by the National Institutes of Health Intramural Research Program. It was also supported by an American Recovery and Reinvestment Act grant to Kaiser Permanente Health. The authors acknowledge Mary Slavin for training the staff.

References
