Replicability of SF-36 Summary Scores by the SF-12 in Stroke Patients

A. Simon Pickard, BScPharm; Jeffrey A. Johnson, PhD; Andrew Penn, MD; Francis Lau, PhD; Tom Noseworthy, MD

Background and Purpose—The replicability of the physical and mental component summary scores of the Short Form (SF)-36 has been established using the SF-12 in selected patient populations but has yet to be assessed in stroke patients. If the summary scores of the SF-12 are highly correlated with those of the SF-36, the benefits of using a shorter health-status measure may be realized without substantial loss of information or precision. Both self-reported and proxy assessments were evaluated for replicability.

Methods—Intraclass correlation coefficients (ICCs) and linear regression were used to assess the ability of the SF-12 physical component summary (PCS-12) scores to predict PCS-36 scores and the SF-12 mental component summary (MCS-12) scores to predict MCS-36 scores. Multivariate regression was used to explore the relationship between SF-12 and SF-36 scores.

Results—The MCS-12 and PCS-12 scores were strongly correlated with the corresponding SF-36 summary scores for surveys completed by proxy or self-report (ICCs ranged from 0.954 to 0.973). Regression analysis of the proxy assessments indicated that patient age was an important effect modifier in the relationship between MCS-12 and MCS-36 scores.

Conclusions—The SF-12 reproduced SF-36 summary scores without substantial loss of information in stroke patients. Accordingly, the SF-12 can be used at the summary score level as a substitute for the SF-36 in stroke survivors capable of self-report. However, the mental health summary scores of proxy assessments are influenced by patient age, thereby limiting the replicability of the SF-36 by the SF-12 under these conditions. (Stroke. 1999;30:1213-1217.)

Key Words: health status ■ quality of life ■ stroke outcome

Measurement of health status or health-related quality of life (HRQL) in stroke patients can potentially provide clinicians and researchers with information on the recovery process, identify predictors of patient outcome, and assist in the evaluation of medical interventions. Among health status instruments used to assess general patient outcomes, probably the most widely used is the SF-36. The SF-36 is a generic measure of health status that has been validated in patients with stroke.1 Recently, an abbreviated version of the SF-36, the SF-12,2,3 has been developed. The SF-12 generates the physical and mental component summary (PCS and MCS, respectively) scores of the SF-36 with considerable accuracy, while imposing less burden on the respondents. Evidence of the replicability of SF-36 PCS and MCS scores by the SF-12 has been demonstrated in samples from the general US population4 and by UK researchers for patient subgroups with Parkinson’s disease, congestive heart failure, sleep apnea, and benign prostatic hypertrophy.5

The importance of using the SF-12 in place of the SF-36 is of particular consequence for the research evaluation of stroke patients. Time and cost savings may be realized through the use of a shorter battery of questions to be included into a longitudinal questionnaire series, while providing essentially the same prognostic information as the longer form. The shorter SF-12 questionnaire can substantially reduce the time spent by respondent and interviewer in an administered survey. Decreased respondent burden through use of the SF-12 in stroke patients may result in findings comparable with those of Dorman et al 6 with the EuroQol questionnaire. Shorter instruments with less missing data increase the efficiency of the study and reduce the resources required. Furthermore, by enabling responses from patients with poorer outcomes, a shorter, simpler instrument may provide more power to detect differences between groups because larger sample sizes will counter small losses in precision. The total time to complete the SF-12 questionnaire is less than 2 minutes for the majority of individuals,4 while the corresponding time to respond to the SF-36 is 10 to 12 minutes. These completion times are likely to be greater for both instruments in stroke patients.
Compared with the SF-36, the disadvantages of using the SF-12 include less-precise estimate of individual health and an inability to calculate summary scores when 1 item is left unanswered. This contrasts with the ability to impute missing data on the SF-36 because of multiple item domains. The SF-12 also does not appear to accurately reproduce the 8 domain scores of the SF-36. As a consequence, disaggregated summary scores may be less informative to users of the SF-12. Furthermore, use of the SF-12 in place of the SF-36 as a screening instrument to detect health problems may compromise the sensitivity and specificity offered by the more extensive 36-item survey.

Proxy assessment of the health status of stroke survivors is sometimes necessary because of cognitive impairment of the patient. This assessment is often performed by the patients’ caregivers. Thus, it is also important to assess the replicability of SF-36 summary scores obtained via proxy assessment.

The purpose of this analysis was twofold: (1) to determine the degree to which the summary scores of the SF-12 replicate the MCS and PCS scores of the SF-36 in stroke patients and (2) to characterize identifiable differences in the relationship between the SF-36 and SF-12 summary scores. We adopted the convention suggested by the developers of the instruments, referring to the summary scores calculated from the SF-36 as the PCS-36 and MCS-36, and scores from the SF-12 as the PCS-12 and MCS-12.

Hypothesized for objective 1 was that PCS-36 and MCS-36 are highly correlated (ICC > 0.90) with the scores of the PCS-12 and MCS-12, respectively. The hypothesized relationship between the SF-36 and SF-12 summary scores was also tested using simple linear regression, with a 2-tailed test to determine whether the SF-12 summary score differed significantly from a slope of 1.0 or intercept of 0 in relation to the respective SF-36 summary score. Objective 2 was hypothesis generating and was accomplished through exploratory correlational and regression analysis.

**Subjects and Methods**

**Patient Sample**

This analysis is part of an ongoing study to evaluate the implementation, use, and effects of an evidence-based disease guidance system on managing stroke (also known as the stroke guidance system, or SGS). Ethics approval was granted by the Health Research Ethics Board at the University of Alberta. The patient group belonged to the baseline control group before implementation of the SGS. The group consisted of 207 ischemic stroke patients (International Classification of Diseases, Ninth Revision [ICD-9] codes 433, 434, 436, and 437) admitted to the emergency room at the University of Alberta Hospital in Edmonton, Alberta, Canada, between February 1 and August 31, 1997. All of the surveys were conducted by telephone at 6 months after discharge by the research coordinator on 162 of these patients who consented and were alive at the time. The remaining 45 patients were either deceased or not locatable. Up to 3 attempts were made to contact patients who could not be located initially. At the time of analysis, 1 survey was incomplete, leaving 161 survey responses as the final sample size. The majority of stroke patients in the sample experienced mild functional impairment according to the Barthel Index, with 58% of patients scoring ≥95 and 15% of patients scoring ≤90. Self-report was not feasible for 53 of the 161 patients due to the extent of their disability, so proxy assessment via a family member was used to estimate the health status of these individuals. The case mix of diagnoses was similar for both proxy-assessed and self-reporting patients, with cerebral artery occlusion (ICD 434.91) the diagnosis in 69% of self-reporting patients and 75% of patients assessed by proxy. Acute but ill-defined cerebrovascular disease (ICD 436.00) was diagnosed in 24% of self-reporting patients and 25% of patients assessed by proxy. The remaining 7% of self-assessed patients suffered other stroke-related diagnoses.

**Data Analysis**

Scores for the MCS-36 and PCS-36 of the SF-36 Health Survey were calculated using the SAS scoring program. Based on findings by Ware and colleagues that scores for the MCS-12 and PCS-12 did not differ if the items are presently separately as opposed to being reconstructed from the SF-36, scores were calculated from the subset of SF-12 items embedded within the SF-36.

Proxy or self-reported assessments were described and analyzed independently. Because proxy assessments and self-assessments were not obtained for the same individuals, the direct comparison of these forms of assessment was not possible and was therefore not an objective of this analysis. One-way ANOVA was performed to detect significant differences between proxy- and self-assessed groups. Agreement between SF-12 and SF-36 summary scores were determined using ICCs and simple linear regression for the full sample and within the assessment groups.

Multiple linear regression analysis was used to address the second objective. The SF-36 summary scores were the dependent variables, with the corresponding SF-12 summary scores used as independent variables. Because age is typically related to PCS and gender to MCS, the linear regression models initially included age and gender. A dichotomous variable identified assessments as being completed by proxy or self-report. Interaction terms were also included in the regression model. Differences were considered statistically significant for a value of P < 0.05. All analyses were performed using SPSS for Windows, Release 7.5.1.

**Results**

As noted above, SF-36 data were available for 108 self-assessed and 53 proxy-assessed stroke patients. The mean (SD) age of stroke patients in this sample was 72.1 (11.6) years. Stroke patients assessed by proxy (73.7 ± 9.5 years) were slightly older than self-reporting respondents (71.3 ± 12.6 years), but this difference was not statistically significant (P = 0.223). The full sample was predominantly male (58.4%). There was no significant gender difference (P = 0.298) between the patient groups assessed by proxy (64.2% male) and by self-report (56.6% male). The mean difference in Barthel Index scores between self-reporting patients and those requiring proxy respondents was statistically significant using a 2-tailed t test for independent samples (P < 0.001), with mean scores of 92 (SD 14.1) and 64 (SD 34.8), respectively.

**Physical Health**

The mean PCS-36 scores for proxy-assessed stroke patients were significantly below those of self-reporting stroke patients (Table 1). PCS-12 scores were not statistically different between proxy and self-report (P = 0.06). PCS-12 scores were in high agreement with PCS-36 scores for both self-assessments (ICC = 0.959) and proxy assessments (ICC = 0.973). The slopes of PCS-12 in the simple regression models were significantly different from 1.0, as evidenced by the boundaries of the confidence interval (Table 1). However, this is not an important difference in terms of the interpretation of scores. There was no significant deviation from 0 for the intercepts on any of the regression models for PCS scores.
TABLE 1. SF-12 and SF-36 Summary Scores for Stroke Patients

<table>
<thead>
<tr>
<th>Variable</th>
<th>Proxy Model</th>
<th>Self-Report Model</th>
<th>Combined Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Beta (SE)</td>
<td>95% CI</td>
<td>Beta (SE)</td>
</tr>
<tr>
<td>Constant</td>
<td>12.46 (10.99)</td>
<td>-9.61, 34.54</td>
<td>6.23 (9.55)</td>
</tr>
<tr>
<td>PCS-12</td>
<td>0.80 (0.25)*</td>
<td>0.30, 1.30</td>
<td>0.91 (0.20)*</td>
</tr>
<tr>
<td>Age</td>
<td>-0.23 (0.15)</td>
<td>-0.52, 0.070</td>
<td>-0.11 (0.41)</td>
</tr>
<tr>
<td>Proxy</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>PCS-12/age</td>
<td>0.004 (0.003)</td>
<td>-0.003, 0.011</td>
<td>0.002 (0.003)</td>
</tr>
<tr>
<td>PCS-12/proxy</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Age/proxy</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>PCS-12/proxy/age</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>F</td>
<td>391.05</td>
<td></td>
<td>405.69</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.957</td>
<td></td>
<td>0.920</td>
</tr>
</tbody>
</table>

SE indicates standard error.
*Significant, $P<0.05$.

In univariate correlations with SF-36 summary scores, age was negatively correlated with PCS-36 for both self-reporting ($r = -0.308$) and proxy-assessed ($r = -0.346$) stroke patients. A weak to medium-strength relationship (point biserial correlation = 0.305) was observed between patient gender and proxy-assessed PCS-36 score, but not for self-reporting patients. All of the linear regression models predicting PCS-36 scores showed PCS-12 to be the only significant independent variable, and this relationship was not modified by age, gender, or proxy assessment of health status (Table 2).

Mental Health

The mean scores obtained by proxy on both the MCS-12 and MCS-36 were significantly lower than for assessments by self-report (Table 1). The absolute difference in health status between proxy-assessed and self-reported mental summary scores was greater than for the physical summary scores. MCS-12 and MCS-36 scores were significantly correlated for self-report (ICC = 0.954) and proxy assessments (ICC = 0.973). The slopes of MCS-12 in the simple regression models were not significantly different from 1.0, but the intercept for proxy assessments was significantly different from 0 (Table 1).

Neither age nor gender independently correlated with MCS-36 score. In fact, MCS-12 score was the only significant predictor variable for self-reported MCS-36 scores using linear regression, giving an adjusted $R^2 = 0.918$ (Table 3). However, the regression model predicting proxy-assessed MCS-36 scores differed from the model based on patient self-report. Age and the interaction term between age and MCS-12 score were significant, while the MCS-12 score itself was not a significant predictor. In the combined sample model, which identified an assessment as being completed by proxy or self-report, significant independent variables included MCS-12 score, proxy assessment status, and interaction terms between age and proxy; MCS-12 and proxy; and MCS-12, proxy, and age (Table 3).

Discussion

As a measure of health status in stroke patients, the shorter, simpler SF-12 offers some potential advantages over the SF-36, particularly in terms of reduced respondent burden. The substitutability of the SF-12 for the SF-36 is fundamentally dependent on the extent to which SF-12 summary scores predict scores for the SF-36. This is not to say, however, that the SF-36 is the “gold standard” for validating HRQL instruments, but rather the criterion set for evaluating the SF-12. The strong agreement between the respective physical and mental component summary scores for the SF-12 and SF-36 in stroke patients affirms the replicability of SF-36 summary scores using the SF-12. This implies that clinicians can administer the simpler, briefer instrument with lower respondent burden with minimal loss of information and precision when summary scores are sought.

The importance of incorporating generic health status measures in stroke outcome measure was recently emphasized by Duncan et al., who discussed the inadequacy of using measures such as the Barthel Index to capture the full impact of stroke-related disability. These authors indicated that standardized assessment of individuals with stroke must evaluate across the entire continuum of health-related functions, and they recommended that measures such as the
MOS-36 (SF-36) be used in addition to the Barthel Index, which has a ceiling effect and captures only physical functionality.10

However, the use of generic health status measures in stroke patients can be similarly compromised by ceiling effects, floor effects, and insensitivity to change. A review by Williams12 cites potential problems with the content validity of both the domains and the items comprising the domains of the SF-36. For instance, the SF-36 does not assess language or cognition. A floor effect is likely to be encountered on some items, such as those regarding mobility, and the limited number of response options on some of the SF-36 items may impair the ability of the SF-36 to detect improvements in health status. As the SF-12 presents only one third of the domains and the items comprising the domains of the MCS-36 scores, it may be even less responsive to change. For these reasons, the SF-12 may be better suited to discriminate between patients rather than to evaluate change within individuals over time.

Due to the inability of some stroke survivors to self-complete health status questionnaires, the use of proxy assessments has been studied in several generic health status instruments, including the Health Utilities Index (HUI),13 the EuroQol,14 the Sickness Impact Profile (SIP),15 and the Health Status Questionnaire (formerly the SF-36 of the Medical Outcomes Study).16 The conclusions of these investigations have been mixed.

Mathias et al13 reported moderate to high agreement in interrater reliability between stroke patients and proxies on the HUI, suggesting that family caregivers can complete the HUI reliably when patients are unable to do so. Dorman et al14 concluded that the HRQL information obtained by proxy on the more observable domains of the EuroQol may be sufficiently valid and unbiased to be used in most types of trials and surveys, but found poor agreement for the domain that assessed psychological function. Rothman et al15 studied the validity of proxy assessments using the SIP and also found that proxy responses for psychosocial attributes were less predictive of patient responses than proxy responses to observable attributes.

Segal and Schall16 indicated that proxy agreement for the HSQ (SF-36) scales was poor, with a median ICC of 0.32 for the 8 dimensions. Agreement was highest on the physical functioning dimension (ICC=0.67) but was otherwise poor for the other dimensions that largely consisted of more subjective items. The authors postulated that poorly educated respondents had more difficulty with comprehension of the HSQ items, further detracting from interrater reliability.

Stroke patients assessed by a proxy respondent in this study had significantly lower MCS-36 and PCS-36 scores. This was predictable and is precisely the reason that proxy assessments may have been necessary. Large differences in MCS-12 and MCS-36 scores were detected between assessments by proxy and self-report. However, the PCS-12 did not demonstrate a significant difference between proxy and self-report, which might suggest that the SF-36 is a more sensitive than the SF-12 as a discriminative measure.

The finding that age was an effect modifier in the relationship between MCS-36 scores and the MCS-12 in proxy assessments was of interest. Several possible explanations have been considered, including limitations in the study design, which did not randomly assign patients to assessment by proxy or self-report. This finding may have resulted from poorer health status typical of patients requiring proxy assessments as opposed to an association with surrogate assessment. Another plausible explanation is that proxy assessments of health status are less informative for domains that are more difficult to observe, such as the domains comprising the MCS summary score.

Age appears to be a clinically important effect modifier for proxy assessments when examined in the context of the multivariate regression models. Table 3 conveys the dramatic changes in the intercept and slope coefficients when proxy assessments are separated from self-reporting patients. Patient age modified the relationship between the MCS-12 and MCS-36 scores generated by proxy assessment, which may imply that proxy respondents took the age of the patient into account when assessing the patient’s mental health. Elaborating on the proxy assessment model from Table 3, younger stroke patients with lower MCS-12 scores had poorer predict-

### TABLE 3. Variables in Regression Model Predicting MCS-36 Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Proxy Model</th>
<th></th>
<th>Self-Report Model</th>
<th></th>
<th>Combined Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta (SE)</td>
<td>95% CI</td>
<td>Beta (SE)</td>
<td>95% CI</td>
<td>Beta (SE)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Constant</td>
<td>34.45 (12.52)*</td>
<td>9.30, 59.61</td>
<td>-2.20 (9.61)</td>
<td>-21.25, 16.86</td>
<td>-2.20 (9.96)</td>
<td>-21.87, 17.47</td>
</tr>
<tr>
<td>MCS-12</td>
<td>0.40 (0.26)</td>
<td>-0.13, 0.92</td>
<td>1.01 (0.181)*</td>
<td>0.65, 1.37</td>
<td>1.01 (0.19)*</td>
<td>0.64, 1.38</td>
</tr>
<tr>
<td>Age</td>
<td>-0.47 (0.17)*</td>
<td>-0.80, -0.13</td>
<td>0.057 (0.14)</td>
<td>-0.22, 0.34</td>
<td>0.057 (0.15)</td>
<td>-0.23, 0.35</td>
</tr>
<tr>
<td>Proxy</td>
<td></td>
<td></td>
<td>36.65 (15.36)*</td>
<td>6.29, 67.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCS-12/proxy/age</td>
<td></td>
<td></td>
<td>-0.0085 (0.003)*</td>
<td>0.001, 0.016</td>
<td>-0.006, 0.005</td>
<td>-0.006, 0.005</td>
</tr>
<tr>
<td>Age/proxy</td>
<td></td>
<td></td>
<td>-0.0005 (0.003)</td>
<td>-0.006, 0.005</td>
<td>-0.03 (0.21)*</td>
<td>-0.95, -0.10</td>
</tr>
<tr>
<td>MCS-12/proxy/age</td>
<td></td>
<td></td>
<td>0.0089 (0.004)*</td>
<td>0.00, 0.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>346.50</td>
<td>381.69</td>
<td>361.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.952</td>
<td>0.918</td>
<td>0.941</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant, $P<0.05$. SE indicates standard error.
tions of MCS-36 scores compared with older proxy-assessed stroke patients. When MCS-12 scores were higher, however, there was better prediction of MCS-36 scores in younger patients assessed by proxy compared with older patients. The intercept of the simple regression model for proxy respondents in Table 1 would also seem to indicate that the scaling of MCS-12 scores is not equivalent to the MCS-36 scores. Age did not modify the relationship between MCS-36 and MCS-12 scores for self-reporting patients.

The discrepancy discussed here between self-report and proxy assessments of health status was similarly observed in the health assessments of elderly men from the Finnish cohorts of the Seven Countries Study. The authors noted that age was not related to self-perceived health, whereas a significant association was detected between patient age and proxy (physician) ratings. Depression and other symptoms that explained self-ratings were not related to proxy assessments. These observations are supportive of our explanation for the interaction between MCS-12 summary scores and age being attributed to the use of surrogate assessments.

While the exploratory analysis of proxy assessments of stroke patients requires further research, previous studies have discussed the limitations of using both SF-36 and SF-12 as a means of generating information about the more subjective mental health and functional aspects of health status. A previous validation study of the SF-36 in stroke survivors reported that the SF-36 did not appear to characterize social functioning well. In addition, the poor agreement between proxy and self-completion responses reported by Segal and Schall and Rothman et al cast doubt on the validity of proxy respondents, particularly for the more subjective items. Similar to recommendations regarding the SF-36, the SF-12 needs to be supplemented by other measures for a comprehensive assessment of health in stroke survivors.

The SF-36 has the advantage of producing scores for the 8 subscales of the instrument. Although subscale scores have been produced for the SF-12, agreement with the SF-36 subscale scores was poor. Currently, use of only the physical and mental health summary scales are recommended for the SF-12. If greater detail on patient status and outcome is required, we suggest that the SF-36 should be chosen over the SF-12.

Conclusions
The SF-12 replicated SF-36 summary scores in this sample of stroke survivors without substantial loss of information. Assessment of the physical functionality of a stroke patient using the PCS-12 appears to adequately replicate PCS-36 scores for both proxy and self-report. MCS-12 also appears to replicate the MCS-36 scores for stroke survivors capable of self-report. However, the relationship between MCS-12 and MCS-36 scores by proxy respondents was modified by age in this sample, a finding that requires further investigation. We recommend that the SF-12 is an appropriate substitute for the SF-36 in stroke survivors capable of self-report and possibly in proxy respondents when subscale scores are not sought, but it should be supplemented by other measures.

Acknowledgments
This project has been partly funded by HEALNet, a Network of Centres of Excellence for Health Research, which is funded by the Medical Research Council and the Social Sciences and Humanities Research Council of Canada. Mr Pickard is supported by a research assistantship at the Institute of Pharmaco-Economics. Dr Johnson is a population health investigator with the Alberta Heritage Foundation for Medical Research. The authors would like to acknowledge Deborah Wilson and Douglas Vincent for contributions to the collection of the data for this analysis, Dr David Feeny for reviewing an earlier draft of this manuscript, and the journal reviewers for their constructive comments.

References
Replicability of SF-36 Summary Scores by the SF-12 in Stroke Patients
A. Simon Pickard, Jeffrey A. Johnson, Andrew Penn, Francis Lau and Tom Noseworthy

Stroke. 1999;30:1213-1217
doi: 10.1161/01.STR.30.6.1213

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/30/6/1213

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/