**Clinical Meaning of the Stroke-Adapted Sickness Impact Profile–30 and the Sickness Impact Profile–136**

A. van Straten, PhD; R.J. de Haan, RN, PhD; M. Limburg, MD, PhD; G.A.M. van den Bos, PhD

**Background and Purpose**—Handicap or health-related quality of life (HRQL) measures are seldom used in stroke trials, although the importance of these measures has been stressed frequently. We studied the clinical meaning of the Stroke-Adapted Sickness Impact Profile–30 (SA-SIP30) and the original SIP136 for use in stroke research.

**Methods**—We included 418 patients who had had a stroke 6 months earlier. We studied the associations between the SA-SIP30 and SIP136 scores versus other frequently used outcome measures from the International Classification of Impairments, Disabilities, and Handicaps (ICIDH) (Barthel Index, Rankin Scale) and the HRQL model (health perception items, Euroqol). To interpret the continuous SA-SIP30 and SIP136 scores, we used receiver operating characteristic curve analysis with the aforementioned measures as external criteria.

**Results**—The psychosocial dimension scores of both SIP versions remained largely unexplained. The physical dimension and total scores of both SIP versions were mainly associated with the disability measures derived from the ICIDH model, as well as with the physical HRQL domains. Most patients with an SA-SIP30 total score $\geq 33$ or an SIP136 total score $\geq 22$ had poor health profiles. There were no major differences between the SA-SIP30 and the SIP136, although the SA-SIP30 scores were less skewed toward the healthier outcomes than the SIP136.

**Conclusions**—Our study showed that (1) both SIP total scores primarily represent aspects of physical functioning and not HRQL; (2) both SIP versions provide more clinical information than the frequently used disability measures; and (3) the SA-SIP30 should be preferred over the SIP136. *(Stroke. 2000;31:2610-2615.)*

**Key Words:** health | psychometrics | quality of life | stroke | validity

**S**troke may have a devastating impact on patients’ lives. In stroke research, this impact should be accurately assessed. Since many patients die after a stroke, patient outcomes should always be expressed in terms of mortality. The health of the survivors can be hierarchically described according to the International Classification of Impairments, Disabilities, and Handicaps (ICIDH) or in terms of Health-Related Quality of Life (HRQL). The ICIDH classification defines impairments as disturbances at the level of the organ (eg, paresis, eye deviation, or dysarthria) and disabilities as the consequences of the impairments on a personal level. In general, disabilities are defined as limitations in (Instrumental) Activities of Daily Living ([I]ADL). Handicap is defined as a disadvantage for a given individual, resulting from impairments or disabilities, that limits or prevents the fulfillment of a role that is normal for that individual. Although no formal HRQL definition exists, there is consensus that it should encompass physical, psychological, and social aspects of health. Moreover, HRQL is usually considered a personal and subjective evaluation of a patient’s health status. HRQL can be seen as an umbrella concept that embodies measures of impairment, disability, and handicap.

In general, disabilities, handicap, and HRQL are the most meaningful for patients. Nevertheless, a recent review on stroke outcome measures, which included all acute stroke trials performed between 1955 and 1995, showed that death was recorded in 76%, impairment in 76%, disability in 42%, and handicap or HRQL in only 2% of the trials. There was a (statistically nonsignificant) trend for more trials to record disability over time. One of the reasons that HRQL is seldom assessed may be the lack of a comprehensive yet feasible instrument.

Quality of life measures can be differentiated into generic scales and disease-specific scales. Generic scales are not specifically developed for a specific target population and may be suitable for use with many patient populations. An advantage of generic scales is that they allow comparison of HRQL results across patient populations. Generic scales, however, are not always sufficiently focused on the specific problems of any given patient population. For this reason, disease-specific HRQL scales have been developed, which are often more sensitive to the HRQL issues that are particularly relevant to a specific patient group. At present, no stroke-specific HRQL scales exist.

Received January 24, 2000; final revision received June 26, 2000; accepted June 26, 2000.

From the Departments of Social Medicine (A. van S., G.A.M. van den B.), Clinical Epidemiology and Biostatistics (R.J. de H.), and Clinical Informatics (M.L.), Academic Medical Center, University of Amsterdam; Netherlands Heart Foundation (M.L.); and Department for Health Services Research, National Institute of Public Health and the Environment, Bilthoven (G.A.M. van den B.), Netherlands.

Correspondence to A. van Straten, PhD. Trimbos Institute, PO Box 725, 3500 AS Utrecht, Netherlands. E-mail astraten@trimbos.nl

*Stroke* is available at [http://www.strokeaha.org](http://www.strokeaha.org)
Currently, the Sickness Impact Profile (SIP) is frequently used as a generic HRQL measure in stroke research. It has been demonstrated that the SIP is reliable and valid in many patient populations. Its validity has also been demonstrated in stroke patients, although data on the reliability in this patient group are limited. However, the SIP has a major disadvantage: its length. In stroke populations, it usually takes 30 minutes to complete the 136 items. Therefore, we designed a Stroke-Adapted Version of the Sickness Impact Profile, the SA-SIP30. The 12 subscales and the 136 items of the original SIP were reduced to 8 subscales with 30 items for the SA-SIP30. We demonstrated that the considerable reduction of items is associated with a relatively small loss of clinical information. The SA-SIP30 scores could explain almost 90% of the variation in scores of the original SIP. Furthermore, the reliability and the validity of the SA-SIP30 were comparable to the psychometric properties of the original SIP 136-item version.

Although the SA-SIP30 addresses the feasibility problem of assessing HRQL in stroke patients, 2 other issues remain. First, since the original SIP (and consequently the SA-SIP30) measures observable behavior instead of more subjective health perceptions, both SIP versions can be regarded as a measure of disability and not as a measure of HRQL. The second issue concerns the clinical interpretation of continuous scale scores. Since clinical trials often make use of dichotomous end points (eg, poor and good health outcome), more knowledge is needed regarding the clinical meaning of continuous scale scores of the SIP. Therefore, in this study we determined the clinical meaning of the SA-SIP30 and SIP136. First, we compared the SA-SIP30 and SIP136 with other frequently used stroke outcome measures derived from both the ICIDH model and the HRQL model. Second, we interpreted the continuous SA-SIP30 and SIP136 scale scores.

### Subjects and Methods

#### Patients

The study comprised 418 patients who had had a stroke 6 months earlier. They were the survivors of an original cohort of 760 consecutively admitted stroke patients. All 760 patients were admitted within 1 week after stroke onset to 1 of 23 randomly selected Dutch hospitals. Patients were traced by administrative records shortly after hospital discharge. Patients were considered to have had a stroke if there was a focal neurological deficit of sudden onset that lasted at least 24 hours with no known alternative to a vascular cause.

Both first and recurrent strokes were included. Patients with a transient ischemic attack, a subarachnoid hemorrhage, or nonstroke pathology were excluded. Six months after stroke, all surviving patients were contacted for an interview at their place of residence.

Of all 760 patients, 258 died within the first 6 months after stroke. Of the 502 survivors, 17 patients refused to participate in the 6-month follow-up interview; 31 were interviewed by telephone because they declined a home visit, and SIP data were not collected. Finally, for 36 of the remaining 454 patients, either SIP data were not collected during the follow-up interview (because of the length of the interview) or the obtained SIP data were insufficient for analyses (arbitrarily defined as >10% of all SIP136 items missing or >50% of items missing on 1 subscale). Of the 418 studied patients, 99 were not communicative because of severe speech, language, or cognitive disorders. These patients were rated by a proxy respondent, primarily the partner.

The original SIP version consists of 136 dichotomous items that are grouped into 12 subscales: Body Care and Movement, Mobility, Ambulation, Social Interaction, Emotional Behavior, Alertness Behavior, Communication, Households Management, Sleep and Rest, Recreation and Pastimes, Eating, and Work. An aggregate score can be obtained for the total SIP136, as well as for each subscale individually. Additionally, 3 subscale scores can be aggregated into a physical dimension (Body Care and Movement, Mobility, Ambulation) and 4 into a psychosocial dimension score (Social Interaction, Emotional Behavior, Alertness Behavior, Communication). By convention, scores are presented as a percentage of maximal dysfunction, ranging from 0% to 100%. Therefore, higher scores indicate less desirable health outcomes. Since the majority of the patients (77%) did not work before their stroke occurred, the subscale “Work” was removed from all analyses.

The SA-SIP30 consists of 8 subscales: Body Care and Movement, Mobility, Ambulation, Social Interaction, Emotional Behavior, Alertness Behavior, Communication, and Household Management. The scoring of items, subscales, dimensions, and total score is the same as for the original SIP136 version. The scores are also presented as a percentage of maximal dysfunction, ranging from 0% to 100%.

### Comparison of the SA-SIP30 and SIP136 With Other ICIDH and HRQL Measures

We first determined the association between the various SIP subscale scores on the one side and the total SIP scores and the dimension SIP scores on the other side by multiple linear regression analyses (forward selection procedure). Thereafter, we studied the associations between the SIP and the measures of disability, global functioning, health perceptions, and HRQL. Disabilities and global functioning were assessed with the Barthel Index and the modified Rankin Scale, respectively. The health perceptions were based on 2 single items: “How would you rate your present health?” and “How satisfied are you with your present life?” Both questions were rated on a 5-point scale (from [very] healthy/satisfied to moderately healthy/satisfied to [very] unhealthy/dissatisfied). HRQL was assessed with the Euroqol, encompassing 5 items: mobility, self-care, main activity, pain/discomfort, and mood (anxiety and/or depression). Each item was rated on a 3-point scale: no problems, some problems, very severe problems. Furthermore, on the basis of the 5 Euroqol dimensions, we computed a single index score. This index score incorporates patients’ values for the different health states. The associations were examined by correlation analyses. We used Pearson’s correlation coefficients since they were almost equivalent to the nonparametric Spearman’s correlation coefficients. Univariate associations were expressed as the percentage of variance of the SIP scores that could be explained by the score on the other outcomes.

#### Interpretation of Continuous SIP Scores

To interpret SIP continuous scores, we used receiver operating characteristic (ROC) curve analysis. The ROC curve shows the ability of the SIP to detect poor and good health outcomes by using...
an external criterion. The curve depicts the true-positive rate (sensitivity) and false-positive rate (1 minus specificity). The area under the ROC curve represents the probability that a random pair of patients will be correctly classified as having a good or a poor outcome with the SIP. A value of 0.50 is obtained if the SIP does not perform better than chance, and a value of 1.0 indicates perfect accuracy. For this analysis, we dichotomized the criterion scores for disability, global functioning, health perceptions, and HRQL as follows: ADL dependent (Barthel Index \( \geq 20 \)); unable to live independently (Rankin classification 3 to 5); feeling moderately to (very) unhealthy (item score 3 to 5); feeling moderately to (very) dissatisfied (item score 3 to 5); and poor HRQL (score 2 to 3 for each item of the Euroqol; score below median \( \leq 0.34 \) for the single index score).

**Results**

**Subject Characteristics and Scores**

The mean age of the 418 patients was 70 years (SD 13 years), and 55% were men. The type of stroke was unspecified in 24 of the 418 patients (6%). Of the remaining 394 patients, 48 (12%) suffered intracerebral hemorrhages and 346 patients (88%) suffered ischemic strokes; 57 of these were infratentorial strokes, 79 lacunar infarctions, and 210 (sub)cortical infarctions.

The mean SA-SIP30 total score for all patients was 32.0 (SD 20.7), and the mean SIP136 score was 22.5 (SD 14.2). The distributions of both total scores were skewed toward the healthier outcomes but to a greater extent for the original SIP136 than for the SA-SIP30 (median scores, 29.8 and 20.8 for the SA-SIP30 and SIP136, respectively) (Figure). This skewed distribution could also be observed for the physical dimensions and especially for the psychosocial dimensions (median physical dimension scores, 35.6 and 20.0 for the SA-SIP30 and the SIP136, respectively; median psychosocial dimension scores, 20.0 and 15.4 for the SA-SIP30 and SIP136, respectively).

**Comparison of the SA-SIP30 and SIP136 With Other ICIDH and HRQL Measures**

Linear stepwise regression analyses showed that the SIP136 and SA-SIP30 physical dimension scores were primarily explained by scores on the subscale Body Care and Movement, whereas the psychosocial dimension scores were mainly explained by the subscale Social Interaction (Table 1). The total scores of both SIP versions were more strongly associated with the physical subscales than with the psychosocial scales.

The total scores of the 2 SIP versions, as well as the physical dimension scores, were substantially associated with other physical functioning measures: disabilities, global functioning, and (I)ADL domains of the Euroqol (Table 2). Furthermore, SIP scores were substantially correlated with the Euroqol index score (the valuation of the Euroqol health states). No clear associations could be demonstrated between both SIP versions and the patients’ health perceptions or with the Euroqol dimensions of pain/discomfort and mood. The psychosocial dimensions of the SIP versions were partially correlated with physical functioning, health ratings, and mood (Table 2).

**Interpretation of Continuous SIP Scores**

With the ROC curves, we identified SIP cutoff scores for poor health outcomes. In general, patients with an SA-SIP30 total score of \( \geq 33 \) or an SIP136 total score of \( \geq 22 \) were ADL disabled; unable to live independently; experienced at least some problems in mobility, self-care, and in performing their main activity; and valued their HRQL as poor (Table 3). The same profile was observed in patients with an SA-SIP30 physical dimension score of \( \geq 40 \) or an SIP136 physical dimension score of \( \geq 23 \). In regard to the SIP psychosocial dimension scores, we could not demonstrate clear cutoff scores for poor health outcomes.

**Discussion**

We studied the clinical meaning of the SA-SIP30 and the original SIP136 by comparing them with other frequently used stroke outcome measures from the ICIDH model and
the HRQL model. Furthermore, we also tried to determine which SIP scores might be interpreted as a poor health outcome.

In our study the psychosocial dimension scores of both SIP versions remained largely unexplained. The physical dimension scores of both SIP versions, however, were substantially associated with the disability measures derived from the ICIDH model, the physical domains of the Euroqol, and the Euroqol single index score. The results of the SIP total scores closely resemble those of the physical dimension scores. This is not surprising because the total scores of both SIP versions have a strong correlation with the SIP physical dimension scores and less of a correlation with the psychosocial scores.

Most patients with an SA-SIP30 physical dimension score >40 or an SIP136 physical dimension score >23 were ADL disabled, were unable to live independently, and experienced at least some problems with mobility and self-care. The same profile could be demonstrated for patients with an SA-SIP30 total score >33 or an SIP136 total score >22. No cutoff values for poor psychosocial functioning could be demonstrated.

### Comparison of the SA-SIP30 and SIP136 With Other ICIDH and HRQL Measures

In most stroke research, measures of impairment are inadequate for describing health outcomes of surviving patients. The Barthel Index has been suggested, and accepted, as the standard measure of disability. The Rankin Scale, originally presented as a handicap measure, is currently considered a global measure of disability. However, limitations in the higher levels of physical functioning are not fully explained.

### TABLE 2. Univariate Associations Between Disabilities, Global Functioning, Health Perceptions, and HRQL vs SA-SIP30 and SIP136 Scores: Total Explained Variance (R²*)

<table>
<thead>
<tr>
<th>Level of Functioning</th>
<th>Instrument</th>
<th>Total Score</th>
<th>Physical Dimension</th>
<th>Psychosocial Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SA-SIP30</td>
<td>SIP136</td>
<td>SA-SIP30</td>
</tr>
<tr>
<td>Disabilties</td>
<td>Barthel Index</td>
<td>36%</td>
<td>52%</td>
<td>39%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53%</td>
<td>59%</td>
<td>51%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51%</td>
<td>59%</td>
<td>51%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39%</td>
<td>45%</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24%</td>
<td>29%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13%</td>
<td>12%</td>
<td>13%</td>
</tr>
</tbody>
</table>

*R² is percentage of total variance in the dependent variable score (SIP) explained by the unadjusted independent variable; R² < 10% not shown.

### TABLE 3. SIP Scores That Correspond to a Poor Health Outcome

<table>
<thead>
<tr>
<th>Level of Functioning</th>
<th>Definition of Poor Outcome</th>
<th>Prevalence of Poor Outcome</th>
<th>Total Score</th>
<th>Physical Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>SA-SIP30</td>
<td>SIP136</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SA-SIP30</td>
<td>SIP136</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SA-SIP30</td>
<td>SIP136</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SA-SIP30</td>
<td>SIP136</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SA-SIP30</td>
<td>SIP136</td>
</tr>
</tbody>
</table>

Values in parentheses are area under the curve and percentage of patients correctly classified. Percentages < 75% not shown; psychosocial dimension not shown because the percentage of patients correctly classified was < 75% on all ICIDH and HRQL measures.
covered by these measures. The Barthel Index merely focuses on very basic daily activities, such as transfers and dressing. The Rankin Scale provides a crude rating score only.

As stated in the introduction, it is generally acknowledged that HRQL measures should encompass physical, emotional, and social aspects of functioning. Although the SA-SIP30 and SIP136 aim to measure all 3 aspects, the psychosocial dimension scores could not be explained by other ICIDH measures or by any of the HRQL dimensions. Since we did not include measures of social functioning as an external criterion, we do not know to what extent the SA-SIP30 and SIP136 total scores measure social aspects of functioning. Further research is needed to reveal what is measured with the psychosocial dimension scores. Interestingly, the total scores of both SIP versions were mainly explained by the physical dimension of the SIP and to a lesser extent by the psychosocial dimension.

HRQL should also be a subjective evaluation of a patient’s health status. In this report we demonstrated that the SA-SIP30 and SIP136 total scores were barely associated with other psychological HRQL domains or with health perceptions as measured with the health and satisfaction rating. However, the total scores of both SIP versions were associated with the subjective evaluation of patients’ health states (single Euroqol index score). We conclude that both SIP total scores are largely based on physical disabilities, but apparently these physical disabilities are of major importance in explaining the patient’s valuation of HRQL. Since the SA-SIP30 and SIP136 describe the (physical) functioning of the patients in more detail than the Barthel Index and the Rankin Scale, we recommend the (additional) use of the short version of the SIP in stroke outcome research.

Interpretation of Continuous SIP Scores
To estimate an effect size in a clinical trial, many stroke researchers prefer a dichotomous primary end point (eg, poor versus good health outcome). This approach gives the investigator the opportunity to compare 2 percentages of response rates, using statistics such as relative risk, relative or absolute risk reduction, or number needed to treat. We demonstrated which cutoff scores might be used for the SA-SIP30 and SIP136. Therefore, by using the SA-SIP30 or SIP136 it is possible not only to demonstrate a statistical difference in the risks for poor outcome between 2 groups but also to designate the involved area of functioning.

SA-SIP30 Versus SIP136
There were 2 differences between the results of the SA-SIP30 and those of the original version of the SIP. The first is concerned with the association between the dimension scores and the subscale scores. The SIP136 dimension scores could be largely explained by 1 subscale each (Body Care and Movement, Social Interaction). The SA-SIP30 dimension scores, however, were substantially related to all relevant subscales. Second, the score distributions of the SA-SIP30 scores were less skewed than the SIP136 scores. The skewed distribution of the SIP136 toward good health outcomes implies that the long version of the SIP is less able to discriminate between patients in relatively good functional health. The skewed distribution of the SIP136 scores was also demonstrated in other studies. Therefore, we conclude that the feasible SA-SIP30 is to be preferred over the original version of the SIP.

In summary, our analyses showed that (1) the original generic SIP and the stroke-adapted 30-item version of the SIP mainly measure aspects of disability instead of HRQL; (2) both SIP versions provide more clinical information than the frequently used disability measures; and (3) in stroke outcome research, the SA-SIP30 should be preferred over the SIP136.

Acknowledgments
This study was supported by the Netherlands Organization of Scientific Research (NWO/KWAZO 900-571-032). Additional support was provided by the National Committee on Investigative Medicine of the Health Insurance Executive Board of the Netherlands (OG91-037) and the Netherlands Heart Foundation (NHS 40.004).

References
Clinical Meaning of the Stroke-Adapted Sickness Impact Profile–30 and the Sickness Impact Profile–136

A. van Straten, R.J. de Haan, M. Limburg and G.A.M. van den Bos

Stroke. 2000;31:2610-2615
doi: 10.1161/01.STR.31.11.2610

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2000 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

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/31/11/2610