Quality of hospital care and health outcomes after stroke
van Straten, A.

Citation for published version (APA):
van Straten, A. (2000). Quality of hospital care and health outcomes after stroke

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
A Stroke Adapted 30-Item Version of the Sickness Impact Profile to Assess Quality of Life (Sa-Sip30)
ABSTRACT

Background and Purpose
In view of the growing therapeutic options in stroke, measurement of quality of life has become increasingly relevant as an outcome parameter. The Sickness Impact Profile (SIP) is one of the most widely used measures to assess quality of life. To overcome the major disadvantage of the SIP, its length, we constructed a short stroke adapted 30-item SIP version.

Methods
Data on the original SIP version were collected for 319 communicative patients at 6 months after stroke. The 12 subscales and the 136 items of the original SIP were reduced to 8 subscales with 30 items in a 3 step procedure, on the basis of relevancy and homogeneity. Reliability of the SA-SIP30 was evaluated by means of an analysis of homogeneity (Cronbach's α coefficient). Different types of validity were assessed: construct, clinical, and external validities.

Results
Homogeneity of the SA-SIP30 was demonstrated by a high Cronbach's α (0.85). Principal component analyses revealed the same two dimensions as in the original SIP (a physical and a psychosocial dimension). The SA-SIP30 could explain 91% of the variation in scores of the original SIP in the same cohort of patients, and 89% in a different cohort. Furthermore, the SA-SIP30 was related to other functional health measures similar to how the original SIP was. We could demonstrate that the SA-SIP30 was able to distinguish patients with lacunar infarctions from patients with cortical or subcortical lesions.

Conclusions
We conclude that the SA-SIP30 is a feasible and clinimetrically sound measure to assess quality of life after stroke.

Selected Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>QoL</td>
<td>quality of life</td>
</tr>
<tr>
<td>SIP</td>
<td>Sickness Impact Profile</td>
</tr>
<tr>
<td>SA-SIP30</td>
<td>stroke-adapted version of the SIP</td>
</tr>
<tr>
<td>SIP136</td>
<td>original SIP</td>
</tr>
</tbody>
</table>
4.1 Introduction

The impact of stroke on daily living may be disastrous. As much as a quarter of all patients die during hospitalization, and recovery for surviving patients is often incomplete. The functional outcomes of stroke may range from dependence in all aspects of daily living to independence, often with residual neurological deficits. Assessment of functional outcomes is useful in determining the health care needs and in evaluating treatment benefits in clinical trials.\textsuperscript{1,2} They are often measured with instruments such as the Barthel Index or the Rankin Scale.\textsuperscript{3} However, these instruments mainly focus on physical disabilities and do not include the social and emotional aspects of functioning, which are also highly relevant stroke outcomes.\textsuperscript{4} These three aspects of functioning (physical, social and emotional) are generally referred to as "quality of life" (QoL).

The SIP is a well-known QoL measure that is often used in various patient populations (i.e., the instrument is generic).\textsuperscript{5-7} The instrument's widespread acceptance originates mainly from its clinical comprehensiveness and its emphasis on observable behavior instead of the more subjective health perceptions, as in other QoL measures such as the Nottingham Health Profile or the SF-36.\textsuperscript{8-10} Furthermore, evaluations of the metric properties of the SIP in patients with different diseases, including stroke, have shown that the instrument is reliable and valid.\textsuperscript{11-13} The SIP has been translated and validated in many languages.

A major disadvantage of the SIP is its length. In stroke populations, it usually takes 30 minutes or more to complete the 136 items. Therefore, using the SIP may impose a considerable burden on the patients. Its length also poses practical problems in large-scale clinical trials. In view of the need of a comprehensive, yet shorter QoL measure for use in stroke outcome research, we decided to develop a short, more feasible, version of the SIP. The metric qualities of this stroke-adapted version of the SIP should approach those of the original scale. Furthermore, we wanted to preserve as much as possible the original subscale structure of the SIP, thereby facilitating comparisons between the scores on the new stroke-adapted SIP and the original version. For the sake of clarity we will refer to the original SIP as the SIP136 and to the new Stroke-Adapted SIP version as the SA-SIP30 in the remainder of this article.
4.2 Patients and methods

The study sample consisted of 319 communicative patients who had had a stroke 6 months earlier. These patients were part of an original cohort of 760 consecutively admitted stroke patients who participated in a multicenter study on quality of care in the Netherlands. All 760 patients were admitted within 1 week after stroke onset to one of the 23 participating hospitals. Patients were considered to have had a stroke if there was a focal neurological deficit of sudden onset that lasted at least 24 hours, or led to death, with no known alternative to a vascular cause. Data about the types of stroke and their locations were abstracted from the medical charts by trained research assistants. A hemorrhage was considered to be present if CT showed evidence of a recent intracerebral hemorrhage. The diagnosis of lacunar stroke was made if there was a clinical picture of one of the lacunar syndromes and if the CT scan was compatible with that diagnosis.

The study was approved by the medical ethical committees of the participating centers. Informed consent was given by all patients.

Six months poststroke all 502 surviving patients (66% of the original cohort) were contacted for an interview by research assistants. Of these 502 surviving patients, 17 patients declined to participate in the follow-up interview, and 31 patients were interviewed by telephone because they refused a home visit. No SIP136 data were collected during the telephone interviews. Of the remaining 454 patients, 84 were noncommunicative because of speech, language or cognitive disorders. Of the 370 communicative patients, 41 were unable to complete the SIP136 because the burden associated with the length of the interview was unacceptably high. Finally, for 10 of the remaining 329 patients the collected SIP136 data were not used for analysis because of too many missing values (arbitrarily defined as > 10% of all the SIP136 items missing or > 50% of the items missing on one subscale).

During the interview, level of disability and global functional health were assessed with the Barthel Index and the modified Rankin scale, respectively.

The original SIP consists of 136 statements, each weighted item describes a possible impact of the disease on some aspect of daily
functioning. The 136 dichotomous items are grouped into 12 subscales: Sleep and Rest, Emotional Behavior, Body Care and Movement, Household Management, Ambulation, Social Interaction, Mobility, Alertness behavior, Communication, Work, Recreation and Pastimes, and Eating. An aggregated score can be obtained for each of these subscales individually, as well as for the total SIP136. Additionally, a physical and a psychosocial dimension score can be constructed by aggregating various subscale scores. The physical dimension consists of three subscales (Body care and Movement, Ambulation and Mobility), and the psychosocial dimension is composed of four subscales (Emotional Behavior, Social Interaction, Alertness Behavior and Communication). By convention, scores are presented as a percentage of maximal dysfunction ranging from 0% to 100%. Therefore, higher scores indicate less desirable outcomes.

Development of the SA-SIP30: Elimination of the Items
Three steps were used to develop the SA-SIP30 (Table 1).

**Step 1. Excluding the Least Relevant Items**
First, the items that were judged to be least relevant for stroke patients within each subscale were excluded. Items with a very skewed response pattern were dropped since those items are applicable to only a very small number of patients. Items applying to no more than 10% of all patients were removed.

The relevance of the remaining items within each subscale was subsequently assessed statistically with linear regression analysis with a forward selection strategy, using the F statistic with $P = 0.05$ at the criteria level for selection. For each subscale the item selection was stopped when the items included in the regression model explained 80% of the score variation of the concerning original total subscale.

**Step 2. Excluding the Least Relevant Subscales**
The second step involved the elimination of the least relevant (shortened) subscales. Here again, a stepwise linear regression procedure with forward inclusion was performed in order to explain the variation of the total original SIP136 score with the (shortened) subscales. The selection of relevant subscales was stopped when adding another subscale into the
model did not result in an increase in the percentage of explained variance of more than 1%.

**Step 3. Excluding Unreliable Items**
The third and last step focused on the exclusion of items that did not contribute to the homogeneity of the subscales, provided that three items remained in each subscale. Homogeneity is a form of reliability that refers to the statistical coherence of the scale items.

**Table 1. Procedures followed in the elimination of items of the original 136-item SIP**

<table>
<thead>
<tr>
<th>Step</th>
<th>Procedure</th>
<th>Decision rule for elimination</th>
<th>No. of items (and subscales) removed</th>
<th>No. of remaining items (and subscales)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Excluding the least relevant items from each subscale</td>
<td>a. Items with a skewed response pattern</td>
<td>a. Items that applied to &lt; 10% of the patients</td>
<td>46 (0)</td>
<td>90 (12)</td>
</tr>
<tr>
<td></td>
<td>b. Linear regression analysis</td>
<td>b. For each subscale the item selection was stopped when the items included in the model explained 80% of the variation of the concerned total subscale score</td>
<td>42 (0)</td>
<td>48 (12)</td>
</tr>
<tr>
<td>2. Excluding the least relevant shortened subscales</td>
<td>Linear regression analysis</td>
<td>The selection of relevant subscales was stopped when adding subscales did not increase the percentage of explained variance by more than 1 percent*</td>
<td>12 (4)</td>
<td>36 (8)</td>
</tr>
<tr>
<td>3. Excluding unreliable items</td>
<td>Cronbach's α</td>
<td>Items were deleted when deletion raised the α of the concerned subscale</td>
<td>6 (0)</td>
<td>30 (8)</td>
</tr>
</tbody>
</table>

* the subscale Work was removed beforehand (not included in the linear regression model)
Performance of the SA-SIP30: homogeneity, validity and comparability

**Homogeneity**
Homogeneity of the SA-SIP30 was assessed with mean inter-item correlations and Cronbach's α coefficients. A Cronbach's α coefficient is based on the (weighted) average correlation of items within a scale. They were compared with the mean inter-item correlations and α coefficients of the original SIP136. In general, homogeneity is considered to be good if α ranges from 0.70 to 0.90. Since the α coefficient is related to the number of items included in a scale, we expected some decrease in the α coefficients for the SA-SIP30. We used the Spearman-Brown "prophecy" formula to estimate theoretically how reliable our developed scale would be if it were extended to the original number of items.

**Validity**
Validity reflects the degree to which a scale measures what it is intended to measure. Several types of validity can be distinguished. In this study construct, clinical and external validities of the SA-SIP30 were tested.

Construct validity, the extent to which the SA-SIP30 fits the theoretical concept of QoL, was assessed with a principal component analysis. The underlying dimensions (or factors) of the new scale were identified. Convergent validity, which is a form of construct validity, measures the extent to which the results of the SA-SIP30 are in concordance with other scales that are purported to measure the same concept, and was evaluated in three ways. First, we compared the SA-SIP30 with the SIP136 by calculating the percentage score variance of the SIP136 that could be explained by the SA-SIP30. We additionally computed the Spearman correlation coefficient between the SA-SIP30 and the SIP136. The Spearman correlation coefficient is based on the rank-ordering of the patients. Second, we visually examined whether agreement varied across the range of SIP136 scores by means of a scatter plot. That is, for each patient the difference between the score on the SIP136 and the SA-SIP30 was plotted against the SIP136 score. When depicted graphically, using the y axis to show difference scores and the x axis to show the original SIP136 scores, perfect correspondence would be
represented by a horizontal line through an ordinate of zero. Differences between the two SIP scores may vary across the range of the original SIP136 scores. Last, we examined the correlation patterns between the SIP136 and the SA-SIP30 on the one hand with the Barthel Index and the modified Rankin Scale on the other.

Clinical validity, the extent to which the SA-SIP30 is able to distinguish between patients with different clinical profiles, was assessed by comparing the SA-SIP30 scores in relation to various stroke types by means of two-group *t*-tests. We have shown earlier that patients with infratentorial lesions reported better functioning than patients with supratentorial lesions on the overall SIP136 score, and patients with lacunar infarcts reported better functioning than patients with cortical or subcortical strokes.\(^{22}\) For the SA-SIP30 to be valid, we hypothesized that the same patterns should emerge.

We examined external validity by comparing the scores on the SA-SIP30 and those on the SIP136 in an independent population. This community based cohort consisted of 185 stroke patients identified by their general practitioners.\(^{23}\) At six months poststroke all patients were followed-up, and SIP136 data were collected for 88 of the 99 surviving patients. For these 88 patients, we compared the SA-SIP30 scores with the SIP136 scores by calculating the percentage of explained score variance.

**Comparability to Reference Data**

Linear regression weights were calculated by which the original SIP136 scores can be estimated from the SA-SIP30 scores. A linear regression equation was calculated for each subscale, for both dimensions and for the total score. For each equation the original SIP136 score was entered as the dependent variable and the new SA-SIP30 score as the independent one.

### 4.3 Results

**Patients**

The mean age of the patients in our study sample was 69 years (SD, 12.6 years); 55% were male. In 36 of the 319 studied patients (11%) no CT scan had been performed. Of the remaining 283 patients, 52 (18%) had suffered an infratentorial lesion and 231 (82%) a supratentorial lesion. Of all 231
patients with a supratentorial lesion, 160 (69%) had suffered a cortical or subcortical infarction or hemorrhage and 71 (31%) a lacunar infarction.

Development of the SA-SIP30: Elimination of the Items

**Step 1. Excluding the Least Relevant Items**
Of all 136 items, 46 items were removed because they applied to less than 10% of our patients (Table 1). The mean item weight of these 46 excluded items was 0.90, whereas the mean item weight of the remaining 90 items was 0.68.

Another 42 items were excluded because 80% of the total subscale scores could be explained by the remaining items in the regression model (Table 1).

**Step 2. Excluding the Least Relevant Subscales**
The aim of this step was to select the most relevant shortened subscales. Since the subscale Work comprised only one item ("I am not working at all") after removing the least relevant items during the first step, and because the majority of our patients (77%) did not work at all before their strokes occurred, we decided to drop the subscale Work beforehand.

A forward linear regression analysis was performed with the 11 remaining shortened subscales as the independent variables and the total original SIP136 score as the dependent one. This analysis showed that when the first 8 subscales were entered, none of the 3 remaining subscales (Sleep and Rest, Recreation and Pastimes, and Eating) could improve the model with more than 1%, and hence they were excluded (Table 2).
### Table 2. The variance of the total original SIP136 score explained by the 11 shortened subscales*

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Original no. of items</th>
<th>No. of items after step 1 of the elimination strategy (n=47)</th>
<th>% Variance explained (cumulative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body care and movement</td>
<td>23</td>
<td>5</td>
<td>54.5</td>
</tr>
<tr>
<td>Social interaction</td>
<td>20</td>
<td>8</td>
<td>75.6</td>
</tr>
<tr>
<td>Mobility</td>
<td>10</td>
<td>4</td>
<td>82.5</td>
</tr>
<tr>
<td>Communication</td>
<td>9</td>
<td>4</td>
<td>86.3</td>
</tr>
<tr>
<td>Emotional behavior</td>
<td>9</td>
<td>4</td>
<td>88.8</td>
</tr>
<tr>
<td>Household management</td>
<td>10</td>
<td>4</td>
<td>90.4</td>
</tr>
<tr>
<td>Alertness behavior</td>
<td>10</td>
<td>3</td>
<td>92.5</td>
</tr>
<tr>
<td>Ambulation</td>
<td>12</td>
<td>4</td>
<td>93.5*</td>
</tr>
<tr>
<td>Sleep and rest</td>
<td>7</td>
<td>4</td>
<td>94.4</td>
</tr>
<tr>
<td>Recreation and pastimes</td>
<td>8</td>
<td>4</td>
<td>95.0</td>
</tr>
<tr>
<td>Eating</td>
<td>9</td>
<td>3</td>
<td>95.5</td>
</tr>
</tbody>
</table>

* the subscale Work was removed beforehand and therefore not included in this linear regression model; 47 items (48 minus 1 work item) remained for use in this analysis

** selection of the subscales was stopped at this point because adding the next subscale did not improve the model with one 1% or more

### Step 3. Excluding Unreliable Items

Finally, in the third step, for each of the remaining shortened subscales Cronbach’s α coefficients were calculated. These analyses showed that the reliability of four shortened subscales could be improved by removing six items. Removing three items from the subscale Social Interaction resulted in an increase of Cronbach’s α of 0.07. For each of the three other subscales only one item was removed, which resulted in an increase of Cronbach’s α of 0.01 (Mobility), 0.03 (Ambulation) and 0.14 (Communication).

After excluding those six items, 8 subscales with 30 items remained. The elimination of items was stopped at this point (see Appendix for the contents of the SA-SIP30).
Performance of the SA-SIP30: homogeneity, validity and comparability

Homogeneity
Cronbach's $\alpha$ was high for the total SA-SIP30 (0.85), the psychosocial dimension (0.78) and the physical dimension (0.82). On a subscale level the $\alpha$ coefficients were sufficient with exception of the subscales Emotional behavior (0.57) and Ambulation (0.54) (Table 3).

Table 3. Percentage of variance of the original SIP136 explained by the short version (for both study populations), (estimated) homogeneity of both SIP versions

<table>
<thead>
<tr>
<th>Short subscale (no. of items)</th>
<th>% explained variance (our study, n=319)</th>
<th>% Explained variance (other study, n=88)</th>
<th>Cronbach's $\alpha$ SA-SIP30 (mean inter-item Correlation)</th>
<th>Spearman-Brown Coefficient*</th>
<th>Cronbach's $\alpha$ SIP136 (mean inter-item correlation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional behavior (4)</td>
<td>84</td>
<td>67</td>
<td>0.57 (0.25)</td>
<td>0.75</td>
<td>0.67 (0.18)</td>
</tr>
<tr>
<td>Body Care and movement (5)</td>
<td>82</td>
<td>78</td>
<td>0.67 (0.29)</td>
<td>0.90</td>
<td>0.84 (0.18)</td>
</tr>
<tr>
<td>Household</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>management (4)</td>
<td>83</td>
<td>89</td>
<td>0.67 (0.33)</td>
<td>0.84</td>
<td>0.72 (0.20)</td>
</tr>
<tr>
<td>Mobility (3)</td>
<td>77</td>
<td>73</td>
<td>0.71 (0.47)</td>
<td>0.89</td>
<td>0.75 (0.22)</td>
</tr>
<tr>
<td>Social interaction (5)</td>
<td>69</td>
<td>66</td>
<td>0.63 (0.26)</td>
<td>0.87</td>
<td>0.74 (0.14)</td>
</tr>
<tr>
<td>Ambulation (3)</td>
<td>72</td>
<td>68</td>
<td>0.54 (0.31)</td>
<td>0.82</td>
<td>0.63 (0.13)</td>
</tr>
<tr>
<td>Alertness behavior (3)</td>
<td>83</td>
<td>75</td>
<td>0.71 (0.47)</td>
<td>0.89</td>
<td>0.84 (0.36)</td>
</tr>
<tr>
<td>Communication (3)</td>
<td>78</td>
<td>81</td>
<td>0.63 (0.35)</td>
<td>0.84</td>
<td>0.62 (0.20)</td>
</tr>
<tr>
<td>Physical (11)</td>
<td>87</td>
<td>84</td>
<td>0.82 (0.30)</td>
<td>0.93</td>
<td>0.90 (0.16)</td>
</tr>
<tr>
<td>Psychosocial (15)</td>
<td>88</td>
<td>83</td>
<td>0.78 (0.20)</td>
<td>0.92</td>
<td>0.88 (0.13)</td>
</tr>
<tr>
<td>Total (30)</td>
<td>91†‡</td>
<td>89</td>
<td>0.85 (0.16)</td>
<td>0.96</td>
<td>0.94 (0.10)</td>
</tr>
</tbody>
</table>

* $T_{\text{Spearman-Brown}} = \frac{k r}{1 + (k-1)r}$ with $k$ = the factor by which the scale is to be increased; $r$ = the original correlation.

† Analysis of residuals did not show violations of necessary assumptions in multiple regression in terms of linearity, equality of variance, independence of error, and normality. No significant outliers could be located.
Compared with the homogeneity of the original SIP136 ($\alpha = 0.94$), the SA-SIP30 performed slightly worse. However, from Table 3 it can be seen that the theoretically estimated Spearman-Brown coefficients are higher than the $\alpha$-s found in the original SIP136. Comparison of the mean inter-item correlations showed that these were rather low for the original SIP136, but were substantially higher for the new SA-SIP30.

**Validity**

Construct validity of the SA-SIP30 was evaluated through a principal component analysis. A two-factor solution (eigenvalue > 3) showed that the factors could be interpreted as a physical dimension on the one hand (including the subscales Body care and Movement, Ambulation, Household Management, and Mobility) and a psychosocial dimension on the other hand (including the subscales Alertness Behavior, Communication, Social Interaction, and Emotional Behavior). This supports the original dimension structure of the SIP136, although the subscale Household Management does not belong to the physical dimension in the original SIP136 version. Twenty percent of the SA-SIP30-explained score variance could be ascribed to the first factor, the physical dimension, and 11% to the second factor, the psychosocial dimension.

Convergent validity was first evaluated by comparing the scores of the new SA-SIP30 with the original scores on the 136-item version. The SA-SIP30 total score could explain 91% of the variation on the original SIP136 total scores. Furthermore, 87% of the original physical dimension scores could be explained by the SA-SIP30, and 88% of the psychosocial dimension scores (Table 3). For the different subscales the percentages of explained variance ranged from 69% (Social Interaction) to 84% (Emotional Behavior). The Spearman rank correlation coefficient between the SA-SIP30 and the SIP136 total scores was 0.96 ($P < 0.01$).

Second, convergent validity was evaluated by means of a scatterplot, which illustrates the differences between the SIP136 and the SA-SIP30 total scores (Figure). The higher the SIP136 score (i.e., worse functioning) the larger the difference between the scores on the SIP136 and the SA-SIP30.
Figure. The differences between the SA-SIP30 and the SIP136 scores plotted against the range of original SIP136 scores.*

* The horizontal line through the ordinate of zero represents a perfect correspondence between the SIP136 and SA-SIP30 score.

Finally, convergent validity was assessed by comparing the correlation patterns of both SIP versions with the Barthel Index and the Rankin Scale. As expected, the SIP versions correlated moderately with the disability score on the Barthel Index (0.50 for the SA-SIP30 and 0.60 for the SIP136) and higher with the global functional health score on the Rankin scale (0.68 for the SA-SIP30 and 0.70 for the SIP136).

Clinical validity of the SA-SIP30 was evaluated by comparing the mean percentage scores for patients with different types of stroke. Against our expectations, there were no statistically significant differences \( P = 0.67 \) between patients with supratentorial strokes and patients with infratentorial strokes. Patients with lacunar infarcts reported better functional health than patients with cortical or subcortical lesions on the total SA-SIP30 \( P < 0.01 \), its psychosocial dimension \( P < 0.01 \), and all
subscales with the exception of Emotional Behavior \( (P = 0.49) \) and Mobility \( (P = 0.07) \).

External validity was estimated by comparing the scores of the new SA-SIP30 with the original scores on the 136-item version for 88 patients from a different stroke study population. Within this cohort of patients, the SA-SIP30 total score could explain 89% of the variation on the original SIP136 total scores, 84% of the original physical dimension, and 83% of the psychosocial dimension scores (Table 3). In this study population, the Spearman rank correlation coefficient between the SA-SIP30 and the SIP136 total scores was .96 \( (P < .01) \).

**Comparability with Reference Data**

To allow estimations of the original SIP136 scores from the SA-SIP30 scores, linear regression equations and their constants were calculated for each subscale, for both dimensions as well as for the total score (Table 4).

**Table 4. Regression weights (β) to estimate SIP136 scores from the SA-SIP30 scores**

<table>
<thead>
<tr>
<th>subscale</th>
<th>regression weights (β)</th>
<th>constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body care and movement</td>
<td>0.52</td>
<td>2.65</td>
</tr>
<tr>
<td>Social interaction</td>
<td>0.41</td>
<td>4.96</td>
</tr>
<tr>
<td>Mobility</td>
<td>0.44</td>
<td>2.37</td>
</tr>
<tr>
<td>Communication</td>
<td>0.47</td>
<td>4.68</td>
</tr>
<tr>
<td>Emotional behavior</td>
<td>0.59</td>
<td>2.69</td>
</tr>
<tr>
<td>Household management</td>
<td>0.69</td>
<td>11.43</td>
</tr>
<tr>
<td>Alertness behavior</td>
<td>0.74</td>
<td>5.77</td>
</tr>
<tr>
<td>Ambulation</td>
<td>0.42</td>
<td>8.27</td>
</tr>
<tr>
<td>Physical dimension</td>
<td>0.52</td>
<td>1.78</td>
</tr>
<tr>
<td>Psychosocial dimension</td>
<td>0.60</td>
<td>3.16</td>
</tr>
<tr>
<td>Total</td>
<td>0.60</td>
<td>2.58*</td>
</tr>
</tbody>
</table>

* For example: a SA-SIP30 total score of 20 will lead to an estimated SIP136 score of 14.58 \[ (\text{SA-SIP30 score} \times 0.60) + 2.58 \].

56
4.4 Discussion

In view of the need for a comprehensive and short QoL measure for use in stroke outcome research we constructed a short stroke-adapted SIP version. With a three-step procedure, which was based on statistical relevancy and homogeneity of the items and subscales, we reduced the original 12 subscales and 136 items to 8 subscales with 30 items. The resulting SA-SIP30 was tested for reliability and validity.

In the past, short SIP versions have been developed for other specific patient populations, for patients with low-back pain, for patients with rheumatoid arthritis, and for nursing home patients. Recently, a generic SIP version consisting of 68 items was constructed. Each of these short SIP versions was achieved through a different procedure. Our methods are comparable to the statistical procedures used to construct the SIP for nursing home patients.

Although the first results from the generic SIP68 seem to be promising, it has two important drawbacks for use in stroke research. First, a scale consisting of 68 items is still quite lengthy, as regards both patient burden and feasibility for large-scale clinical trials. Second, the original subscale structure of the SIP136 was not maintained in the SIP68, but rather new subscales were constructed that contained various items from the original 12 ones. Besides the fact that this approach ignores the widespread familiarity of the SIP136 and its subscale structure, it also hampers comparisons with reference data obtained with the original SIP136. To overcome these problems, we have specifically chosen to preserve the original subscale structure of the SIP136 as much as possible. Although we are aware that it is impossible to calculate the scores on the SIP136 exactly on the basis of the SA-SIP30, the use of the calculated regression weights will at least give an impression on how stroke patients perform in comparison to other groups of patients as measured with the original 136-item version.

The homogeneity coefficients of the original SIP136 in this study are comparable to those found in other studies. Although, the α-coefficients for the SA-SIP30 were lower than for the SIP136, we judged the homogeneity of the SA-SIP30 as good. First, Cronbach's α for the total score of the SA-SIP30 was sufficient, as well as for its dimensions. Second,
the decrease of the α-coefficients appeared to be due solely to the reduction in the number of items and not to a lesser degree of coherence between the remaining items. This was depicted by both the high inter-item correlations and the Spearman-Brown coefficients for the SA-SIP30.

The validity of the SA-SIP30 was supported because the same two dimensions (a physical and a psychosocial one) as in the original SIP136 were found, because the SA-SIP30 was meaningfully associated with other functional health measures as was the SIP136, and because only a relatively small amount of information was lost (the SA-SIP30 could explain 91% of the variation in scores of the original SIP136). The loss of clinical information was most apparent in severely ill patients. In these patients the agreement between the original SIP136 score and the SA-SIP30 was less than in healthier patients. Moreover, as was indicated by the high item weights, the removed skewed items described the most severe health states. This implies that the SA-SIP30 is less effective than the SIP136 in the distinction of patients with seriously impeded health.

We could not demonstrate a difference in SA-SIP30 scores between patients with supratentorial lesions and patients with infratentorial lesions, as was possible with the SIP136. This might be explained by differences in patient populations between the two studies. In this study we excluded the noncommunicative patients, mostly patients with supratentorial strokes, whereas in the study that demonstrated differences, all patients were included. On the other hand, this may indicate a reduced discriminative power of the SA-SIP30. In patients with supratentorial lesions, the SA-SIP30 could distinguish between patients with lacunar lesions and patients with cortical or subcortical lesions, as could the SIP136.

We have constructed the SA-SIP30 and evaluated its metric qualities using one and the same data set. This may have resulted in an overestimation of the metric properties of the SA-SIP30. However, the results obtained from an independent cohort of stroke patients are promising. Especially when considering that the studied patients were selected differently, as they were enrolled by general practitioners and not through hospitals as in our study. These findings support the generalization of our results. Still, further research on different stroke
populations and different time intervals after stroke is required to ascertain the reliability and validity of the new SA-SIP30.

Apart from being reliable and valid, the SA-SIP30 should also be responsive when used in clinical trials. Responsiveness can be defined as the ability of an instrument to detect clinically relevant health changes over time within a patient. Studies on the responsiveness of the SIP136 are still scarce, and the results are contradictory.\textsuperscript{31,32} The inconsistencies may be attributed to the lack of consensus regarding the most efficient responsiveness statistics\textsuperscript{33} as well as the lack of consensus about the definition of a clinically relevant change.\textsuperscript{11,31} Therefore, definite conclusions are not yet possible. However, even in case the responsiveness of the SIP136 turns out to be poor, this might be different for the SA-SIP30. In one study in which the SIP136 was compared with a short 24-item SIP version, the latter turned out to be more responsive. It was argued that this is an advantage of a shorter disease-specific instrument above a longer less-focused scale.\textsuperscript{34} At this moment we are collecting new data in order to study the responsiveness of the SIP136 and the SA-SIP30 in more detail.

It should be noticed that our results are based on communicative patients only. We decided to exclude the noncommunicative patients because of the many conceptual problems surrounding data which are derived through proxy respondents.\textsuperscript{35,36} Nevertheless, we think that the SIP, with its emphasis on observable behavior, is one of the most suitable QoL instruments available for use in proxy respondents.

We conclude that the SA-SIP30 is a feasible and clinimetrically sound measure to assess quality of life after stroke.

4.5 References

1 European Carotid Surgery Trialists' Collaborative Group. MRC European Carotid Surgery Trial: interim results for symptomatic patients with severe (70-99%) or with mild (0-29%) carotid stenosis. \textit{Lancet}. 1991;337:1235-1243.


35 Spranglers MAG, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *J Clin Epidemiol.* 1992;45(7):743-760.