How Do Scores on the EuroQol Relate to Scores on the
SF-36 After Stroke?

Paul J. Dorman, MD, MRCP; Martin Dennis, MD, FRCP; Peter Sandercock, MD, FRCP;
on behalf of the United Kingdom Collaborators in the International Stroke Trial (IST)

Background and Purpose—The EuroQol and Medical Outcome Survey 36-item short-form health survey (SF-36) questionnaires have both been validated for the assessment of health-related quality of life after stroke. However, the relationship between these instruments has not been studied after stroke. We therefore sought to compare the responses of a group of stroke patients to both instruments.

Methods—A total of 2253 patients with stroke entered by United Kingdom hospitals in the International Stroke Trial were randomized to follow-up with either the EuroQol or SF-36 instruments. We randomly selected one third of patients who had responded to the EuroQol for follow-up, again using the SF-36, and two thirds of patients who had responded to the SF-36 for follow-up, again using the EuroQol. We assessed the patients’ mean score for each domain of the SF-36 categorized by their response to the corresponding EuroQol domain and the correlation between the domains of the 2 instruments.

Results—The domains for both instruments, which assessed physical functioning, social functioning, bodily pain, and overall health-related quality of life, correlated closely. The mental health domain of the SF-36 correlated only poorly with the psychological functioning domain of the EuroQol.

Conclusions—Both the EuroQol and SF-36 measure broadly similar domains of health. The weak relationship between the assessments of mental health may reflect a difference in content or more fundamental problems with the validity or reliability of the items in one of the instruments with respect to this domain. This study has provided the first empirical qualitative evidence by which the data on the SF-36 after stroke may be interpreted. (Stroke. 1999;30:2146-2151.)

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

The EuroQol and Medical Outcome Survey 36-item short-form health survey (SF-36) are both valid measures for the assessment of health-related quality of life after stroke. The EuroQol assesses outcome in 6 broad areas (mobility, self-care, activities, pain, psychological functioning, and self-reported overall health-related quality of life) and also provides a utility score for overall health-related quality of life. The SF-36 assesses 8 domains: physical functioning, physical role functioning, social functioning, bodily pain, mental health, psychological role functioning, vitality, and general health. At first it may appear that many of the domains of the SF-36 are similar to those of the EuroQol: for example, the mobility question in the EuroQol appears to relate closely to the physical functioning questions on the SF-36. However, the relationship between responses by an individual to the domains of each instrument has not been well defined. The degree to which any change in health-related quality of life might be reflected differently by the EuroQol and SF-36 is also unknown.

A clearer understanding of the relationship between these instruments might help to improve the interpretation of a change of score with either instrument. It should also be helpful in comparing studies that have used different measures and may even allow the translation of outcomes for meta-analysis. Furthermore, since both instruments aim to measure health-related quality of life, there should be a strong correlation between responses on the 2 instruments. A poor correlation might suggest poor validity of 1 or both of the measures. We therefore administered both the EuroQol and SF-36 to a group of patients after stroke to compare their responses to these instruments.

Subjects and Methods

Patients and Allocation to the EuroQol or SF-36
In a previous study we examined response rates to postal versions of the EuroQol and SF-36; we randomly allocated patients to receive either the EuroQol or the SF-36. We have described in detail elsewhere the methods used to identify patients and the format of the instruments. Briefly, the study included patients with confirmed or suspected ischemic stroke who had been enrolled between March 2, 1993, and May 31, 1995, by any of the United Kingdom hospitals participating in the International Stroke Trial. We included all...
patients who were not known to have died by the time of the survey. We incorporated the EuroQol or SF-36 into booklets that included some additional questions recording the patients’ demographic details, their functional outcome after stroke, and whether the patients completed the booklet by themselves. The only difference in the questionnaire booklets was the nature of the health-related quality of life instrument; they were otherwise identical.

For a study of reproducibility, we then randomly sampled one third of the patients who had responded within approximately 3 weeks to the first questionnaire for repeated testing with the same health-related quality of life instrument (test-retest reliability). Similarly, if the majority of patients score near the top of the measure, it will have little scope to show improvements in health (“ceiling” effects). We initially examined the distribution of scores in each domain to identify the degree of skewness of the distribution, chiefly by assessing, for each domain, the proportion of respondents with a maximum or minimum score for that domain (ie, ceiling and floor effects). We also assessed the levels of missing data for both instruments.

“Ordering” effects are a potentially important source of bias in an unbalanced crossover study. For instance, completing the EuroQol questionnaire first might affect the patients’ subsequent response to the SF-36. We therefore used a simple factorial 1-way ANOVA to investigate whether ordering effects occurred after the administration of either the EuroQol or SF-36. We restricted these analyses to the comparable domains of both instruments (physical functioning [SF-36] versus mobility [EuroQol] and self-care [EuroQol], physical role functioning [SF-36] versus mobility [EuroQol] and self-care [EuroQol], social functioning [SF-36] versus activities [EuroQol], pain [SF-36] versus pain [EuroQol], mental health [SF-36] versus psychological functioning [EuroQol], and psychological role functioning [SF-36] versus psychological functioning [EuroQol]), to avoid the problems that may arise from multiple testing.

We assessed the construct validity of both instruments further by testing the relationship between the EuroQol and SF-36 domains. Thus, the relationship between comparable domains on the EuroQol and SF-36 (such as physical functioning on the SF-36 and mobility on the EuroQol) should be higher than between less comparable domains (such as physical functioning on the SF-36 and psychological functioning on the EuroQol). In contrast, the domains that examine more general aspects of health (such as overall health-related quality of life on the EuroQol) should be moderately correlated with all the other domains. We examined these relationships in 2 separate ways. We initially calculated patients’ median score, for each separate domain of the SF-36, for patients categorized according to their response to the corresponding EuroQol domain. These analyses were performed to facilitate the interpretation of patients’ scores with the SF-36. We subsequently calculated correlation coefficients between the domains of the EuroQol and each of

![Figure 1. Flow of patients through cross-validation study. IST indicates International Stroke Trial; EQ, EuroQol.](http://stroke.ahajournals.org/)

### Table 1. Number of Patients With Maximum, Minimum, and Missing Scores for Each Domain of the SF-36 (n=688)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Range of Potential Levels*</th>
<th>Mean Score in Domain (SD) (Range, 0–100)</th>
<th>No. With Maximum Score in Domain (Best Possible Health Outcome)</th>
<th>No. With Minimum Score in Domain (Worst Possible Health Outcome)</th>
<th>No. With Missing Data for Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical functioning</td>
<td>20</td>
<td>635 (31)</td>
<td>15 (2)</td>
<td>177 (26)</td>
<td>53 (8)</td>
</tr>
<tr>
<td>Physical role functioning</td>
<td>4</td>
<td>593 (36)</td>
<td>83 (12)</td>
<td>416 (61)</td>
<td>95 (14)</td>
</tr>
<tr>
<td>Social functioning</td>
<td>9</td>
<td>678 (33)</td>
<td>112 (16)</td>
<td>83 (12)</td>
<td>10 (2)</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>10</td>
<td>669 (31)</td>
<td>164 (24)</td>
<td>22 (3)</td>
<td>19 (3)</td>
</tr>
<tr>
<td>Mental health</td>
<td>22</td>
<td>648 (12)</td>
<td>3 (&lt;1)</td>
<td>0 (0)</td>
<td>40 (6)</td>
</tr>
<tr>
<td>Psychological role functioning</td>
<td>3</td>
<td>581 (45)</td>
<td>178 (26)</td>
<td>327 (48)</td>
<td>107 (16)</td>
</tr>
<tr>
<td>Vitality</td>
<td>20</td>
<td>647 (22)</td>
<td>3 (&lt;1)</td>
<td>35 (5)</td>
<td>41 (6)</td>
</tr>
<tr>
<td>General health</td>
<td>19</td>
<td>601 (24)</td>
<td>3 (&lt;1)</td>
<td>14 (2)</td>
<td>87 (13)</td>
</tr>
</tbody>
</table>

Values in parentheses are percentages unless otherwise indicated.

*Number of potential levels for raw scores before transformation onto scale of 0 to 100.

†Number of patients with no missing data for the domain.
the domains of the SF-36. All analyses were performed with the statistical software package SPSS for Windows (release 6.1).

Results
Of the 905 respondents to the initial EuroQol questionnaire, 272 (approximately one third of respondents) were selected at random to receive a subsequent SF-36. A separate 505 patients (approximately two thirds of respondents) were selected at random from the respondents to the initial SF-36 to receive a EuroQol questionnaire (Figure 1). Four hundred fifty-eight (91%) of those allocated to the EuroQol questionnaire responded. A slightly lower proportion (85%) of the patients allocated to the SF-36 questionnaire responded (Figure 1). The mean delay between the completion of the initial EuroQol and subsequent SF-36 was 28 days (SD, 12 days). For patients allocated to the SF-36 initially, the mean delay between completion of the questionnaires was 29 days (SD, 13 days).

We performed a simple 1-way factorial ANOVA to assess the effect of the questionnaire ordering (ie, “EuroQol then SF-36” or “SF-36 then EuroQol”) on the relationship between EuroQol and SF-36 scores for comparable domains. The “ordering term” was not a significant determinant of the relationship between the EuroQol and SF-36 scores in the 8 analyses performed. We therefore combined all the data for the remaining analyses.

The distribution of scores for the SF-36 is described in Table 1. There was substantial variation in the proportion of responses with missing data between the different domains, which ranged from 2% to 16% (social functioning and psychological role functioning domains, respectively). The distribution of scores was highly skewed for some domains. A large proportion of respondents scored the minimum possible score (0 of a possible 100, ie, the floor of the scale, and the worst possible outcome) for the domains of physical role functioning and emotional role functioning. Approximately one quarter of patients scored the maximum score for the bodily pain and psychological functioning domains of the SF-36.

The distribution of patients’ responses to the categorical domains of the EuroQol are described in Table 2. The proportion of missing data (approximately 3%) was very similar for each of these 5 domains (Table 2). However, since each of the domains had only 3 potential levels of response, a substantial proportion of patients scored the maximum possible score (no problems) for each domain. Examination of the distribution of overall estimates of health-related quality of life with the EuroQol visual analog scale or the EuroQol utility scores did not suggest problems with ceiling or floor effects (Figures 2 and 3).

The relationships between patients’ responses to the EuroQol and SF-36 questionnaires are presented in Tables 3 and 4.

### Table 2. Number of Patients in Each Response Category for Each Domain of EuroQol (n=688)

<table>
<thead>
<tr>
<th>EuroQol Domain</th>
<th>Range of Potential Levels</th>
<th>n*</th>
<th>No of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>No Problems</td>
</tr>
<tr>
<td>Mobility</td>
<td>3</td>
<td>666</td>
<td>147 (21)</td>
</tr>
<tr>
<td>Self-care</td>
<td>3</td>
<td>669</td>
<td>267 (39)</td>
</tr>
<tr>
<td>Activities</td>
<td>3</td>
<td>674</td>
<td>111 (16)</td>
</tr>
<tr>
<td>Pain</td>
<td>3</td>
<td>669</td>
<td>226 (33)</td>
</tr>
<tr>
<td>Psychological functioning</td>
<td>3</td>
<td>664</td>
<td>257 (37)</td>
</tr>
</tbody>
</table>

Values in parentheses are percentages.

*Number of patients with no missing data for the domain.

![Figure 2. Distribution of estimates of overall health-related quality of life with the EuroQol visual analog scale (n=636).](image)
Table 3 presents the median scores for the relevant SF-36 domains for patients categorized according to their response to the comparable EuroQol domain. For almost all of the domains, the median scores were ordered appropriately and were significantly different between the groups. Indeed, physical functioning, social functioning, and pain measured with the SF-36 were particularly closely related to the corresponding domains on the EuroQol. However, there was no difference in the median scores for the physical role functioning and psychological role functioning domains between patients reporting “some” or “severe” problems with the EuroQol. Furthermore, there was only a weak relationship between the mental health domain (SF-36) and psychological functioning domain (EuroQol) (Table 3).

Table 4 reports the correlation between each of the domains of the EuroQol and those of the SF-36. The physical functioning domain on the SF-36 correlated most closely with the mobility, self-care, and activities domain of the EuroQol; it correlated less closely with the pain and psychological domains of the EuroQol. Social functioning on the SF-36 was moderately correlated with all the domains of the EuroQol. Bodily pain was most closely correlated with the pain domain of the EuroQol. In contrast, mental health correlated only poorly with psychological functioning measured with the EuroQol. The vitality and general health domains of the SF-36 correlated particularly strongly with the overall health-related quality of life domain of the EuroQol, but also moderately with the other domains of the EuroQol.

**Discussion**

**Relationship Between EuroQol and SF-36**

The EuroQol and SF-36 questionnaires have a different background, structure, content, and length and also ask the

**Table 3. Relationship Between EuroQol and SF-36 After Stroke**

<table>
<thead>
<tr>
<th>SF-36 Domain</th>
<th>EuroQol Domain</th>
<th>No Problems (Median, IQR)</th>
<th>Some Problems (Median, IQR)</th>
<th>Severe Problems (Median, IQR)</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical functioning</td>
<td>Mobility</td>
<td>70 (43–85)</td>
<td>15 (0–35)</td>
<td>0 (0–0)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>Self-care</td>
<td>50 (30–75)</td>
<td>10 (0–25)</td>
<td>0 (0–0)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Physical role functioning</td>
<td>Mobility</td>
<td>25 (0–100)</td>
<td>0 (0–0)</td>
<td>0 (0–0)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Physical role functioning</td>
<td>Self-care</td>
<td>0 (0–100)</td>
<td>0 (0–0)</td>
<td>0 (0–0)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Social functioning</td>
<td>Activities</td>
<td>100 (75–100)</td>
<td>50 (25–75)</td>
<td>25 (0–50)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>Pain</td>
<td>100 (74–100)</td>
<td>41 (31–62)</td>
<td>12 (0–22)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Mental health</td>
<td>Psychological functioning</td>
<td>64 (60–68)</td>
<td>60 (52–68)</td>
<td>52 (44–60)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Psychological role functioning</td>
<td>Psychological functioning</td>
<td>100 (0–100)</td>
<td>0 (0–25)</td>
<td>0 (0–0)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Vitality</td>
<td>Psychological functioning</td>
<td>50 (35–65)</td>
<td>33 (20–45)</td>
<td>14 (0–30)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>General health</td>
<td>Psychological functioning</td>
<td>60 (40–80)</td>
<td>37 (25–55)</td>
<td>20 (7–35)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Values are median scores for SF-36 domain for patients classified according to responses to EuroQol domain. Values in parentheses are interquartile range.

*Significance between groups by Kruskal-Wallis test to compare distributions.
patient to consider different time periods. Nonetheless, we observed a close relationship between the domains that assessed physical functioning, social functioning, bodily pain, and overall health-related quality of life. Our results suggest that the 2 instruments are generally sampling similar areas of health. This finding supports the notion that there are several key dimensions that constitute health-related quality of life, as well as providing further support for the construct validity of the assessments of these domains with either instrument.

The correlation between patients’ responses to the mental health domain of the SF-36 and the psychological functioning domain of the EuroQol was poor. There are several possible explanations for this. First, it is possible that these domains, although superficially similar, are measuring different constructs. This is supported by the fact that the EuroQol item focuses on anxiety and depression, whereas the SF-36 mental health scale includes positive emotions as well (eg, feeling calm and peaceful). The psychological role functioning domain of the SF-36, which emphasizes anxiety and depression, correlated much better with the EuroQol psychological functioning domain than did the mental health domain of the SF-36 (Spearman rank correlation coefficient, 0.43 versus 0.21). However, an alternative explanation is that 1 or both of these domains, although superficially similar, are measuring different constructs. This is supported by the fact that the EuroQol item focuses on anxiety and depression, whereas the SF-36 mental health scale includes positive emotions as well (eg, feeling calm and peaceful). The psychological role functioning domain of the SF-36, which emphasizes anxiety and depression, correlated much better with the EuroQol psychological functioning domain than did the mental health domain of the SF-36 (Spearman rank correlation coefficient, 0.43 versus 0.21). However, an alternative explanation is that 1 or both of these domains have poor measurement properties in patients with stroke. There are several indications of this. First, the assessments of mental health with the SF-36 were clustered around the middle of the scale (mean score, 61; SD, 12) and therefore did not appear to take full advantage of the potential breadth of the scale. Second, the reproducibility of the mental health assessments with the SF-36 was also particularly poor (intraclass correlation coefficient = 0.28). Finally, approximately half of these questionnaires were completed with the help of proxies, and the validity of these proxy assessments is particularly questionable for the domain of psychological functioning.

It has been difficult to establish the validity of the numerical assessments of overall health-related quality of life with the EuroQol because this domain is difficult to define and is highly subjective. However, the general health domain of the SF-36 appears to examine a similar construct. It aims to assess an individual’s general health perceptions and satisfaction and, as with the EuroQol, these general health perceptions appear to provide an approach by which different components of health such as disease, functioning, symptoms, and feelings can be integrated. The strong correlation (Pearson correlation coefficient = 0.66) between patients’ responses to these domains supports the view that both these domains are measuring the same underlying trait. The validity of these assessments is further supported by the moderate correlation of the assessments of overall health-related quality of life with the other domains of the SF-36.

### Interpretability

Studies of interventions must show that the observed changes in patients that are due to the intervention are important and substantial enough to warrant further consideration in medical practice and policy planning. One approach to the definition of clinical meaningfulness is the use of anchor-based interpretations. These definitions represent instances in which the changes in quality of life measures were compared, or anchored, to other clinical changes or results. The descriptive nature of the categorical levels of the EuroQol questionnaire could be considered potential anchors. In the present study, a change of 55 points in the physical functioning scale of the SF-36 appeared to be equivalent to the difference between “no problems” and “some problems” in the categorical mobility domain of the EuroQol. However, several factors limit the usefulness of this approach. First, clinicians may be unsure about the significance of such a change in the EuroQol and what the term “some problems” means in practice. Second, the amount of change judged significant may differ with the population and the type of treatment under study. Third, most scales are not linear, ie, not an interval scale, and therefore a change of 10 units at the top of the scale may not be the same as a similar-sized change at the bottom of the scale. An additional factor that limits the interpretability of Likert scaled scores (such as the SF-36) is that the total score will never give clinical information about...
the exact responses to individual items. For example, a total score of 50 on the 10-item physical functioning scale can be achieved in different ways.

The difficulties in defining a clinically significant change in these health-related quality of life measures reflect in part the newness of these measures and our lack of experience with them. Therefore, presenting these correlations should improve researchers’ familiarity with them and may help to develop an intuitive feeling about the relevance of any change.

Distribution of Scores
The large number of patients scoring the minimum score (worst outcome) in the physical and emotional role functioning domains of the SF-36 suggests that floor effects may be present in these domains. The observation that the median scores for the physical role functioning domain did not distinguish between patients classified as having moderate or severe problems by the mobility or self-care domains of the EuroQol confirms this suspicion. The role functioning domains may therefore not measure the consequences of more severe disabilities, and this might reduce responsiveness in these domains. This confirms the findings of other investigators in studies of the SF-36 in patients with stroke, the elderly, and groups of patients with other diagnoses. Indeed, because of these problems the SF-36 has recently been revised.

Methodological Issues
The lower frequency of response and lower levels of data completeness in patients followed up with the SF-36 compared with the EuroQol are consistent with the result of the direct randomized comparison of their feasibility after stroke. We used an interpolation procedure to reduce the proportion of missing data (missing items were substituted with the mean response to other items), and therefore these results underestimate the underlying level of missing data for the SF-36. Brazier and colleagues have expressed concerns over the validity of these interpolation procedures. They suggest that when patients omit items because they do not appear relevant to them, this may indicate that the respondent is in fact unable to perform that particular activity or function, and therefore the average response to the other items could be misleading if interpolation is used for missing values.

The crossover design used in this study seems to have been valid since there did not appear to be any significant carryover or other ordering effects. Furthermore, the study of test-retest reliability demonstrated that the patients did not change significantly in any of the domains of health-related quality of life between test and retest. These findings justified the combined analysis of all the data irrespective of the order of questionnaire administration.

Conclusions
In summary, despite fundamental differences in their background, design, and format, the domains of the EuroQol and SF-36 measured broadly similar aspects of health-related quality of life. The weak relationship between the assessment of mental health with the SF-36 and psychological functioning with the EuroQol may reflect a difference in content or more fundamental problems with the validity or reliability of the items in either of these domains. Unfortunately, it is difficult to resolve which of these explanations applies since no reference instruments were administered concurrently. This study has provided the first empirical qualitative evidence by which data on the SF-36 after stroke may be interpreted.

Acknowledgments
This study was supported by a grant from Glaxo-Wellcome plc. Dr Dorman was supported by a United Kingdom Medical Research Council Training Fellowship. Dr Sandercock was also supported by grants from the United Kingdom Medical Research Council. We would like to thank the patients, their families, and their caregivers for their keen participation. We are also grateful to Stephanie Lewis for statistical support.

References
How Do Scores on the EuroQol Relate to Scores on the SF-36 After Stroke?
Paul J. Dorman, Martin Dennis and Peter Sandercock
on behalf of the United Kingdom Collaborators in the International Stroke Trial (IST)

Stroke. 1999;30:2146-2151
doi: 10.1161/01.STR.30.10.2146

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/30/10/2146