Two Simple Questions to Assess Outcome After Stroke

A European Study

Christopher McKevitt, PhD; Ruth Dundas, MSc; Charles Wolfe, MD; on behalf of the European BIOMED II Study of Stroke Care Group

Background and Purpose—The “2 simple questions” were designed as an efficient way of measuring outcome after stroke. We assessed the sensitivity and specificity of this tool, adapted for use in 8 European centers, and used it to compare outcomes across centers.

Methods—Data were taken from the Biomed II prospective study of stroke care and outcomes. Three-month poststroke data from 8 European centers were analyzed. Sensitivity and specificity were assessed by comparing responses to the 2 simple questions with Barthel Index and modified Rankin scale scores. Adjusting for case mix, logistic regression was used to compare patients in each center with “good” outcome (not dependent and fully recovered) at 3 months.

Results—Data for 793 patients were analyzed. For the total sample, the dependency question had a sensitivity of 88% and a specificity of 77%; the recovery question had a sensitivity of 78% and a specificity of 90%. Dependency data from Riga had much lower sensitivity. There was variation in good outcome between centers (P=0.0015). Compared with the reference center (Kaunas), patients in Dijon, Florence, and Menorca were more likely to have good outcome, after adjusting for case mix.

Conclusions—Dependency and recovery questions showed generally high sensitivity and specificity. There were significant differences across centers in outcome, but reasons for these are unclear. Such differences raise particular questions about how patients interpreted and answered the simple questions and the extent to which expectations of recovery and perceived needs for assistance vary cross-culturally. (Stroke. 2001;32:681-686.)

Key Words: disability ■ outcome assessment ■ recovery

Assessment of patient outcomes is essential for quality assurance of care, for investigations of the efficacy and effectiveness of new methods of treating or caring for patients, and for investigation of the efficiency of resource use. Mortality is the most straightforward outcome to measure, but this provides limited information about the efficacy of an intervention or the quality of a service. Outcomes such as disability or handicap provide more information for evaluators, but even commonly used measures may be interpreted in different ways.1

There is a drive toward comparing outcomes cross-nationally. This can be related to several factors, including the World Health Organization’s goal of “Health for All.”2 The globalization of markets targeted by multinational pharmaceutical companies, and, in the European context, moves toward harmonization within the region. This drive has led to a need for indicators which are equally meaningful across countries so that questions may be expected to elicit the same types of responses.

Many measures are lengthy, making them unsuitable for large-scale studies. Two simple questions were developed by Lindley et al3 to meet the need for a simple, inexpensive, and quick way of assessing functional outcome in large numbers of stroke patients recruited to an open, randomized, controlled trial. It was suggested that the tool would also be useful to monitor patient outcome in routine practice and clinical audit.4,5 The questions were developed to assess dependency and recovery.

The 2 simple questions were developed in the United Kingdom and were used in the pilot and subsequent International Stroke Trial (IST).6 The IST authors accepted the validity and reliability of the 2 simple questions, although in fact studies testing the measures were conducted only in the United Kingdom, whereas the IST involved centers in many non-English speaking countries. Because IST data are not presented in disaggregated form, we do not have information about any possible intercountry differences in responses to the 2 simple questions. In this study we aimed to investigate the sensitivity and specificity of the 2 simple questions used in a pan-European prospective study of stroke care, resource use, and outcomes. If possible, we then aimed to compare outcomes across centers using this tool.

Subjects and Methods
Data were collected in the European Union BIOMED II stroke study, which was undertaken to investigate the relationships between
Testing the Validity of the 2 Simple Questions

To investigate how successfully the 2 simple questions captured dependency and recovery, sensitivity and specificity were calculated. Sensitivity is defined as how accurately a question identifies positive cases. Specificity is defined as how well a question identifies negative cases. The standard way of calculating sensitivity and specificity of a new measure is to compare responses to the new measure with responses to a gold standard or usual measure. Following discussion at an early planning meeting of participants to introduce the questions and clarify the concepts, the questions were then translated into the relevant language by each local study coordinator.

For the present study, data from 8 centers were analyzed (ie, all centers at which the 2 simple questions were asked at 3-month follow-up). Five centers are excluded from this analysis because they did not collect either the 2-simple-questions data or 3-month follow-up data.

Testing the Validity of the 2 Simple Questions

To investigate how successfully the 2 simple questions captured dependency and recovery, sensitivity and specificity were calculated. Sensitivity is defined as how accurately a question identifies positive cases. Specificity is defined as how well a question identifies negative cases. The standard way of calculating sensitivity and specificity of a new measure is to compare responses to the new measure with responses to a gold standard or usual measure. Following discussion at an early planning meeting of participants to introduce the questions and clarify the concepts, the questions were then translated into the relevant language by each local study coordinator.

For the present study, data from 8 centers were analyzed (ie, all centers at which the 2 simple questions were asked at 3-month follow-up). Five centers are excluded from this analysis because they did not collect either the 2-simple-questions data or 3-month follow-up data.

Testing the Validity of the 2 Simple Questions

To investigate how successfully the 2 simple questions captured dependency and recovery, sensitivity and specificity were calculated. Sensitivity is defined as how accurately a question identifies positive cases. Specificity is defined as how well a question identifies negative cases. The standard way of calculating sensitivity and specificity of a new measure is to compare responses to the new measure with responses to a gold standard or usual measure. Following discussion at an early planning meeting of participants to introduce the questions and clarify the concepts, the questions were then translated into the relevant language by each local study coordinator.

For the present study, data from 8 centers were analyzed (ie, all centers at which the 2 simple questions were asked at 3-month follow-up). Five centers are excluded from this analysis because they did not collect either the 2-simple-questions data or 3-month follow-up data.

Testing the Validity of the 2 Simple Questions

To investigate how successfully the 2 simple questions captured dependency and recovery, sensitivity and specificity were calculated. Sensitivity is defined as how accurately a question identifies positive cases. Specificity is defined as how well a question identifies negative cases. The standard way of calculating sensitivity and specificity of a new measure is to compare responses to the new measure with responses to a gold standard or usual measure. Following discussion at an early planning meeting of participants to introduce the questions and clarify the concepts, the questions were then translated into the relevant language by each local study coordinator.

For the present study, data from 8 centers were analyzed (ie, all centers at which the 2 simple questions were asked at 3-month follow-up). Five centers are excluded from this analysis because they did not collect either the 2-simple-questions data or 3-month follow-up data.

Testing the Validity of the 2 Simple Questions

To investigate how successfully the 2 simple questions captured dependency and recovery, sensitivity and specificity were calculated. Sensitivity is defined as how accurately a question identifies positive cases. Specificity is defined as how well a question identifies negative cases. The standard way of calculating sensitivity and specificity of a new measure is to compare responses to the new measure with responses to a gold standard or usual measure. Following discussion at an early planning meeting of participants to introduce the questions and clarify the concepts, the questions were then translated into the relevant language by each local study coordinator.

For the present study, data from 8 centers were analyzed (ie, all centers at which the 2 simple questions were asked at 3-month follow-up). Five centers are excluded from this analysis because they did not collect either the 2-simple-questions data or 3-month follow-up data.

Testing the Validity of the 2 Simple Questions

To investigate how successfully the 2 simple questions captured dependency and recovery, sensitivity and specificity were calculated. Sensitivity is defined as how accurately a question identifies positive cases. Specificity is defined as how well a question identifies negative cases. The standard way of calculating sensitivity and specificity of a new measure is to compare responses to the new measure with responses to a gold standard or usual measure. Following discussion at an early planning meeting of participants to introduce the questions and clarify the concepts, the questions were then translated into the relevant language by each local study coordinator.

For the present study, data from 8 centers were analyzed (ie, all centers at which the 2 simple questions were asked at 3-month follow-up). Five centers are excluded from this analysis because they did not collect either the 2-simple-questions data or 3-month follow-up data.
larger proportion of respondents as having good outcome than the 2 simple questions, although the upper limit of the confidence interval is only 9.4%. In Almada, Riga, and Warsaw, there was no significant difference between the 2 methods for identifying good outcome. For the remaining centers, the modified RS classified a larger proportion of respondents as having good outcome than the 2 simple questions.

Table 4 reports the distributions of the BI for those who reported dependency using the 2 simple questions and also for those who did not report dependency. With the exception of respondents in Riga, high proportions of patients who were not dependent reported a BI score of 20. In all centers, only negligible numbers who were not dependent had BI score ≤9. In Riga, a large proportion (43%) of respondents who reported dependency had BI = 20. In all centers, the proportion of respondents who reported that they had not made a complete recovery but were scored RS = 0 was small (3%). However, there were high proportions of respondents who reported having made a complete recovery while scoring RS = 1 or 2, especially in Dijon (71%) and Kaunas (79%).

Table 5 reports the unadjusted proportions classified into each outcome category (bad, indifferent, good) by using the 2 simple questions. After adjusting for case mix, the centers with the least likelihood of good outcome were in Eastern

### TABLE 1. Initial Case Mix and Survival and Outcome at 3 Months, by Center

<table>
<thead>
<tr>
<th>Center</th>
<th>Overall</th>
<th>Almada (n=106)</th>
<th>Dijon (n=133)</th>
<th>Florence (n=144)</th>
<th>Kaunas (n=320)</th>
<th>London (n=107)</th>
<th>Menorca (n=50)</th>
<th>Riga (n=287)</th>
<th>Warsaw (n=153)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>At time of maximum impairment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In coma</td>
<td>32 (31.0)</td>
<td>30 (22.6)</td>
<td>33 (22.9)</td>
<td>20 (6.2)</td>
<td>22 (20.6)</td>
<td>6 (12.0)</td>
<td>94 (32.8)</td>
<td>21 (13.7)</td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Urinary incontinence</td>
<td>49 (46.2)</td>
<td>42 (31.6)</td>
<td>70 (48.6)</td>
<td>114 (35.6)</td>
<td>50 (46.7)</td>
<td>25 (50.0)</td>
<td>79 (27.5)</td>
<td>60 (39.2)</td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Limb weakness</td>
<td>90 (84.9)</td>
<td>103 (74.4)</td>
<td>89 (83.2)</td>
<td>261 (84.0)</td>
<td>118 (41.1)</td>
<td>235 (81.9)</td>
<td>144 (94.1)</td>
<td></td>
<td></td>
<td>0.011</td>
</tr>
<tr>
<td>At 3 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>72 (67.9)</td>
<td>101 (76.9)</td>
<td>99 (68.7)</td>
<td>227 (70.9)</td>
<td>59 (55.1)</td>
<td>27 (54.0)</td>
<td>118 (41.1)</td>
<td>90 (58.8)</td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>BI=20</td>
<td>42 (58.3)</td>
<td>41 (40.2)</td>
<td>41 (41.4)</td>
<td>109 (48.0)</td>
<td>21 (35.6)</td>
<td>12 (44.4)</td>
<td>69 (58.5)</td>
<td>43 (47.8)</td>
<td></td>
<td>0.023</td>
</tr>
<tr>
<td>Rankin=0</td>
<td>11 (15.3)</td>
<td>7 (6.9)</td>
<td>17 (16.7)</td>
<td>28 (12.3)</td>
<td>7 (25.9)</td>
<td>18 (15.4)</td>
<td>13 (9.7)</td>
<td>14 (15.6)</td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Home alone</td>
<td>7 (9.7)</td>
<td>17 (16.7)</td>
<td>8 (8.1)</td>
<td>28 (12.3)</td>
<td>14 (14.8)</td>
<td>9 (7.7)</td>
<td>14 (15.6)</td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Home with carer</td>
<td>59 (81.9)</td>
<td>62 (60.8)</td>
<td>81 (81.8)</td>
<td>196 (86.3)</td>
<td>30 (50.8)</td>
<td>23 (85.2)</td>
<td>107 (91.5)</td>
<td>73 (81.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In institution</td>
<td>6 (8.3)</td>
<td>23 (22.5)</td>
<td>10 (10.1)</td>
<td>3 (1.3)</td>
<td>14 (23.7)</td>
<td>0</td>
<td>1 (0.8)</td>
<td>3 (3.3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values in parentheses are percent, unless otherwise stated. Ellipses indicate data not collected.
Europe. The odds ratios, compared with Kaunas, were highest in Menorca and Florence. London and Dijon also had higher odds of a good outcome compared with Kaunas.

Discussion
Simple outcome measures that can be used across centers and countries are required for large-scale studies and trials, but their strengths and weaknesses need to be understood when deciding which should be adopted. The 2 simple questions were used as outcome measures in the IST, one of the largest trials of stroke care ever undertaken. This study investigated the use of the 2 simple questions as part of a larger study of stroke outcomes across selected European studies. We compared responses to the 2 simple questions with widely used measures of disability and handicap. The overall sensitivity of the dependency question was high (88%), although sensitivity found in the data from Riga is considerably lower than in all other centers. Nevertheless, the sensitivity of the dependency question in all centers including Riga is higher than reported by Lindley et al,3 who found a sensitivity of 61% in their original study conducted in Scotland. The sensitivity of the recovery question found in the present study (78%) was similar to that reported by Lindley et al3 (83%).

TABLE 4. Distribution of BI Scores for Dependency Question and RS Scores for Recovery Question

<table>
<thead>
<tr>
<th>Center</th>
<th>Not Dependent</th>
<th>Dependent</th>
<th>Not Completely Recovered</th>
<th>Completely Recovered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0–9</td>
<td>10–19</td>
<td>20</td>
<td>0–9</td>
</tr>
<tr>
<td>Almada</td>
<td>2 (5)</td>
<td>2 (5)</td>
<td>37 (90)</td>
<td>9 (29)</td>
</tr>
<tr>
<td>Dijon</td>
<td>1 (2)</td>
<td>10 (22)</td>
<td>34 (76)</td>
<td>16 (28)</td>
</tr>
<tr>
<td>Florence</td>
<td>0 (0)</td>
<td>11 (24)</td>
<td>35 (76)</td>
<td>20 (38)</td>
</tr>
<tr>
<td>Kaunas</td>
<td>1 (1)</td>
<td>33 (23)</td>
<td>108 (77)</td>
<td>19 (22)</td>
</tr>
<tr>
<td>London</td>
<td>0 (0)</td>
<td>17 (18)</td>
<td>38 (43)</td>
<td>11 (30)</td>
</tr>
<tr>
<td>Menorca</td>
<td>0 (0)</td>
<td>2 (5)</td>
<td>12 (71)</td>
<td>8 (50)</td>
</tr>
<tr>
<td>Riga</td>
<td>2 (3)</td>
<td>23 (30)</td>
<td>51 (67)</td>
<td>9 (21)</td>
</tr>
<tr>
<td>Warsaw</td>
<td>0 (0)</td>
<td>3 (8)</td>
<td>36 (92)</td>
<td>9 (18)</td>
</tr>
</tbody>
</table>

Overall 6 (1) 90 (21) 331 (76) 101 (28) 219 (60) 47 (13) 58 (47) 61 (49) 5 (4) 16 (3) 316 (52) 275 (45)

Values in parentheses are percent unless stated otherwise.

Overall, there was a significant difference between the proportion classified as having a bad outcome using the definition of BI<20 and the proportion classified as bad using the single question. Considering the 95% confidence interval, the definition of bad outcome of BI<20 will classify at most 9.2% more as bad outcome defined by the 2-simple-questions method. However, the wider confidence interval in the estimate might represent more of problem in choosing between 2 methods when the sample size is small, as in the case of Menorca in this study, or for other unknown reasons, as in the case of Kaunas (upper limit of the 95% CI, 19.2%).

There was a similar pattern for the proportions classified as having a good outcome. Overall, the upper limit of the 95% CI for the difference was 9.4%, although the upper limits of the 95% CI for the differences in Dijon and Florence were high: 21.9% and 23.1%, respectively. This means that the modified RS score of 0 may classify considerably more respondents as having good outcome than the 2-simple-questions method. Nevertheless, the patterns of outcome identified in this way are broadly similar to previous cross-national studies of stroke10,11 in which poorer outcomes have been reported in Eastern European countries, and in the United Kingdom, compared with other Western European countries.

There are a number of potential limitations to the study, which must be taken into consideration when interpreting the outcomes presented here. First, variations in data collection methods adopted across centers for 3-month follow-up data (face-to-face interviews, telephone interviews, and postal survey) might constitute a methodological limitation. On the one hand, data collection of BI scores by telephone interview has been shown to have high validity compared with face-to-face assessment.16 On the other hand, reporting of subjective health status using the Short Form (SF)-36 is affected by the data collection method. A study comparing postal and telephone administration of the SF-36 found that health ratings were poorer and chronic illness more frequently reported in postal responses.17 Thus, we acknowledge the possibility that different data collection methods might produce different
results, with more subjective assessments (general health status rather than ability to perform specific task) perhaps more liable to collection method influence. Allowing centers to use their own preferred method of data collection in this study was not ideal but was necessary to encourage the participation of all centers, including some centers in eastern Europe with limited resources. It is also a pragmatic approach to data collection that will be required should such outcome measures be used in routine practice in the future. A further limitation, as acknowledged above, relates to the small sample size of 1 center in particular, Menorca, which resulted in wide confidence intervals.

Other factors should be considered when differences in outcome across centers are interpreted. Because variations in case mix have an impact on outcomes and their interpretation, statistical correction for confounding variables is essential. In this study, outcome as measured by the 2 simple questions was adjusted for by using variables appropriate to the clinical condition being investigated. Nevertheless, some case-mix variation might remain unmeasured, which could account for different outcomes across centers. Social class data were not collected in the study, because this information is difficult to standardize across countries, especially in the elderly. However, social class might be associated with expectations of recovery and uptake of support services.

Another caveat relates to the quality of the questionnaire translations from English into the local languages. In the field of cross-national quality-of-life measurement, it has been proposed that, at the very least, correct instrument translation requires forward-backward translation, as well as a test of psychometric criteria on appropriate subjects. Such development work was beyond the resources of this project, as well as those of the IST. The 2 simple questions were therefore translated by local study coordinators following discussions at initial project meetings attended by all participants. We were unable to monitor the quality of the translation used in each center, although it was assumed that the questions were conceptually and semantically unproblematic. This assumption may well be unfounded: the meaning of “complete recovery” in particular may be open to different interpretations. This should also be taken into consideration when interpreting differences between the 2-simple-questions scores and modified RS scores, as well as differences in 2-simple-questions outcome across centers.

The importance of our study, however, lies not so much in the outcome data themselves but in the questions raised about the kind of outcome the 2 simple questions capture. They were devised as functional outcomes. Unlike the modified RS, which was developed for assessment by an observer, the 2-simple-questions method asks the patient (or proxy) for information and therefore requires the patient to interpret the question and to decide what information to divulge. One question might be assumed to ask only for factual information: “Did you require help from another person for everyday activities?” However, even this raised problems in the original study, with some subjects being unclear about the question’s meaning or intention. Researchers participating in this study agreed to a definition of everyday activities that encompassed basic functional tasks (i.e., feeding, dressing, personal hygiene). Nevertheless, the question is open to differing interpretations, because the concept of “everyday” activities may well vary from one respondent to another, as might expectations about activities for which it is legitimate to receive assistance from another person. Factors that might influence views of what constitutes an everyday activity and views of legitimate assistance include age, prestroke activities, family situation, and local culture. A striking example is provided by Ali and Mulley in their study of the use of the BI in rural Pakistan, in which they concluded that the measure was not appropriate, given local customs and lifestyle.

The second question asks subjects to comment on their own feeling of recovery and therefore invites an entirely subjective assessment. Consequently, the issue of how respondents interpret the question posed is likewise of crucial importance. To interpret better the responses elicited across different age groups (younger stroke patients compared with older stroke patients) and across cultures (a stroke patient in post-Soviet Latvia compared with an Italian living in Florence or a first-generation African-Caribbean living in inner-city London), we need to know more about the meanings attached to “recovery.” These will surely be influenced by different expectations of recovery, disability, and well-being in different age groups, cultures, and perhaps also social class/wealth. The investigators of an American qualitative study of 102 stroke patients have reported that none of the patients interviewed, “even those who had visibly and tangibly regained lost function,” considered themselves to have recovered. However, it is difficult to generalize about stroke patients’ concepts of recovery, because few studies have investigated their expectations, much less how these may vary in different contexts.

In conclusion, this study has tested the feasibility of using simple questions rather than longer outcome instruments in a cross-national investigation. Both simple questions showed high sensitivity and specificity, demonstrating their theoretical validity. There were significant differences across centers in outcome, but reasons for these differences remain unclear. They may reflect inadequate translation of the questions, residual case-mix variation, or different cultural expectations of recovery and needs for assistance. The 2 simple questions were proposed as a resource-efficient method of collecting data on physical outcome after stroke. The need to collect data efficiently in large-scale studies can lead to a trade-off between sophistication and feasibility. Because the 2 simple questions have been used in a well-known international study that recruited over 19,000 subjects, their feasibility has been demonstrated. However, our study suggests that this method of outcome assessment should be adopted with caution. Although designed to assess functional outcome, respondents may interpret the questions with reference to a number of criteria. Some of these, such as culturally specific views of what constitutes recovery, have not yet been investigated. However, their importance in interpreting responses to questions such as those considered here should not be discounted.

Appendix

Study Participants
Prof A. Czlonkowska and Dr D. Ryglewicz (Institute of Psychiatry and Neurology, Warsaw, Poland); Dr J. Aleixo Dias (Division of...
Epidemiology and Biostatistics, Direccao General Da Saude, Lisbon Portugal); Prof M.O. Carrageta, Dr I. Remidios, and Dr J. Namora (Hospital Garcia de Orta, Almada, Portugal); Prof G. Enina and Dr I. Purina (Latvian Neuroangiological Centre, Riga, Latvia); Prof M. Giroud and Dr M. Menassa (Service de Neurologie, Hôpital General, Dijon, France); Prof D. Inzitari, Dr P. Vanni, Dr A. Di Carlo, Dr M. Lamassa, and Dr S. Rossi (Dipartimento di Scienze Neurologiche & Psichiatriche, Ospedale Careggi, Florence, Italy); Dr S. Spolveri and Dr M.C. Baruffi (Torregalli Hospital, Florence, Italy); Dr D. Rastenye (Institute of Cardiology, Medical Academy, Kaunas, Lithuania); Dr M. Torrent (Health Care Research Unit, Mao, Menorca, Spain); Dr A.G. Rudd (Department of Care of the Elderly, Guy’s and St Thomas’ Hospital, London, England); Dr R. Beech (Centre for Health Planning and Management, Keele University, England).

Acknowledgments

The study was funded by the European Union Biomed II programme. We are grateful to all patients and their family members who participated in the study.

References

Two Simple Questions to Assess Outcome After Stroke: A European Study
Christopher McKeivitt, Ruth Dundas and Charles Wolfe
on behalf of the European BIOMED II Study of Stroke Care Group

Stroke. 2001;32:681-686
doi: 10.1161/01.STR.32.3.681
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2001 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/32/3/681

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/